Amir Hussain

Editor

Electronics, Communications and Networks V

Proceedings of the 5th International Conference on Electronics, Communications and Networks (CECNet 2015)

Springer

Lecture Notes in Electrical Engineering

Volume 382

About this Series

"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

Amir Hussain

Editor

Electronics, Communications and Networks V

Proceedings of the 5th International Conference on Electronics, Communications and Networks (CECNet 2015)

Springer

Editor
Amir Hussain
University of Stirling
Scotland
UK

Contents

A Battery Equalizing Scheme Using Flyback Converter and PhotoMOS Switch

Wen-Hui Li, Feng Ran, Yuan Ji, Jia-Qi Qin and Hao Xu

Abstract This article presents the concept of a series battery balance circuit using flyback converter to implement multiple filling valley balance. The optoelectronic switches are used as the control components for increased safety. Each battery cell has a Photo Metal-Oxide-Semiconductor Field-Effect Transistor (PhotoMOS) switch. The switch is used to determine the time at which the battery cell should be charged. The equalization current can reach 2A, hence the equalization speed is faster. The simulation and experimental results verify the operating characteristics of this proposed topology and method, and show that the circuit is straightforward and has been a low cost implementation.

Keywords Lithium battery · Equalization · Photomos · Flyback transformer

1 Introduction

In recent years, lithium battery packs are widely used in power tools. The lithium battery has many advantageous characteristics, including a stable discharge voltage, low self-discharge rate, wide range of operating temperatures, and light in weight and volume to list a few. Large capacity batteries must connect multiple batteries

W.-H. Li · F. Ran · Y. Ji (✉) · J.-Q. Qin · H. Xu
Microelectronic Research and Development Center, Shanghai University,
Shanghai 200072, China
e-mail: jiyuan@shu.edu.cn

W.-H. Li
e-mail: 18321771578@163.com

F. Ran
e-mail: ranfeng@shu.edu.cn

J.-Q. Qin
e-mail: 249109456@qq.com

H. Xu
e-mail: xuhaoee@hotmail.com

using series connections. Whilst performance difference between cells will cause a decline in the entire battery life and discharge time, a battery overcharge will cause an explosion [1]. Many balance topologies have been proposed to solve this problem [2–5].

Balance topologies can be divided into two methods: dissipative [3] and non-dissipative [4–11]. Dissipative method is the energy consumption of the battery that has more battery power. Non-dissipative method is the use of a transformer or storage element to achieve balance. As the dissipative method takes more power dissipation, most researchers are recently focusing on non-dissipative methods. The isolated DC-DC converter is used to shunt the other battery's currents to the lower voltage cells [4, 5]. The switched-capacitor was proposed to reduce switching loss by switching the buck-boost and cuk converters [7]. However, most of the schemes mentioned above can only resolve the issue of voltage inconsistencies between adjacent batteries, and their costs are high due to each battery needing two or more switches.

This article will use non-energy consumption balance technology based on the flyback converter [8], and can achieve the effect of supplementing the energy of several battery cells by using the battery pack energy at the same time. Each battery cell has a PhotoMOS switch. The switch is used to decide whether the lower voltage battery cell should be balanced. This circuit has the following advantages:

- It can allow low voltage signals control of the high voltage circuit using a PhotoMOS switch.
- It can achieve the equalization of several battery cells at the same time.
- One switch is required for each battery cell in this circuit, so the control method is very simple.

2   Proposed Equalization Circuit

2.1   Circuit Configuration

The equalization circuit topology of single battery is shown in Fig. 1. Part I is the Metal-Oxide-Semiconductor Field-Effect Transistor(MOSFET) M1 driver circuit for enhancing Pulse-Width Modulation (PWM) wave signal to speed MOSFET conduction velocity. Part II is flyback converter, it is used to transfer the battery pack energy to a battery. Each battery is required to use such circuit, so the system can balance among several battery cells by controlling the PhotoMOS switches.

Battery equalization circuit working principle: when a battery needs energy supplement, PhotoMOS switch is turned on and the flyback converter primary side is inputted the PWM wave. When M1 is turned on, the primary coil is equivalent to a pure inductance, it is charged by battery pack during the conduction time. The primary side current straights up and the energy is stored, the secondary side can't be turned on due to the diode. When M1 is turned off, all winding voltage is

Fig. 1 Battery equalization circuit
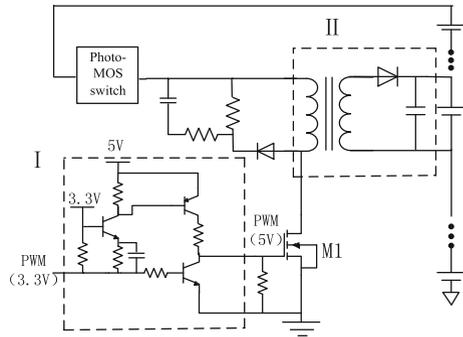
reversed and diode in the secondary side is turned on, the energy stored in the primary side start charging the secondary side of the capacitor and the battery. During the off time of M1, to ensure that the secondary side current is linearly decreased to zero finally. By PWM control, M1 periodically turned on and off, to give the low energy of battery energy supplement. The initial value of the PWM input always keep high level to ensure that M1 is turned off to prevent the flyback converter burn out.

PhotoMOS switch is invented by Shanghai University. The biggest difference between PhotoMOS switch and traditional mechanical relay is that contact is not mechanically opened and closed. The principle of PhotoMOS switch is shown in Fig. 2. U1 is composed of light emitting diode and light-battery, SEL is connected with the I/O port of MCU. When the SEL is high level, the MOSFET N1 is turned on. The light emitting diode in U1 is turned on and generates light at this time. The light-battery will generate voltage when the light shines on it. The positive of the light-battery connects to the gate electrode of the MOSFET Q1 and Q2, the negative of it connects to the source of the MOSFET Q1 and Q2. Q1 and Q2 will be turned on when the light-battery has voltage. So PhotoMOS switch can realize low voltage signal control the high voltage circuit.

This PhotoMOS switch has the following advantages:

- It has a high sensitivity, can control a maximum 0.30 A load current with a 5 mA input current.

Fig. 2 PhotoMOS switch schematic
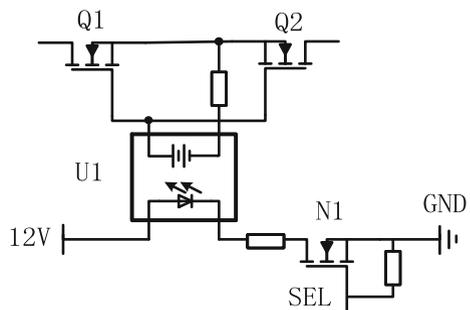
- It has low-level off state leakage current. The PhotoMOS relay has only 100 pA even with the rated load voltage of 60 V.
- Its opening and closing time is particularly fast. Its turn-on time is 0.5 ms and turn-off time is 0.03 ms.

2.2 Circuit Analysis and Calculation

In order to calculate balanced current size, we need to analyze the waveform of critical points in equalization circuit. The simulation waveforms of PWM, primary current, secondary current and drain voltage are shown in Fig. 3. The first waveform is PWM wave which is inputted to the gate of MOSFET M1.

When the MOSFET is turned on, the transformer primary current rises linearly. Before the MOSFET is turned off, the peak current is I_p:

$$I_p = \frac{dI_p}{dt} T_{on} = \frac{U_{in}}{L_p} T_{on} = \frac{U_{in}TD}{L_p}$$ (1)

where U_{in} denotes the battery pack voltage, L_p denotes the primary inductance, T denotes PWM wave cycle and D denotes the duty cycle.

When the MOSFET is turned off, the secondary current peak I_s is:

$$I_s = \sqrt{\frac{L_p}{L_s}} I_p = \frac{n_1 U_{in}TD}{n_2 L_p}$$ (2)

n_1 and n_2 respectively denote the number of turns of the primary and secondary side.

Fig. 3 Simulation waveform of circuit
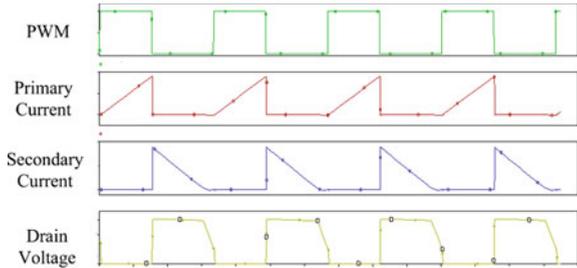
Since working is in discontinuous state, the time of the secondary current by the maximum value to zero is denoted by T_{off}, then:

$$I_s = \frac{(U_0 + U_D)}{L_s} T_{off} \tag{3}$$

$$T_{off} = \frac{I_s L_s}{(U_0 + U_D)} = \frac{n_2 U_{in} TD}{n_1 (U_0 + U_D)} \tag{4}$$

wherein U_0 denotes the battery voltage of secondary side and U_D denotes the voltage of rectifier diode.

The average value of the secondary current is:

$$\bar{I}_0 = \frac{1}{2} I_s \frac{T_{off}}{T} \tag{5}$$

Substitute Eqs. (3) and (4) into Eq. (5) and get equalization current of the flyback equalization circuit:

$$I_{ec} = \frac{1}{2} I_s \frac{T_{off}}{T} = \frac{U_{in}^2 TD^2}{2L_p(U_0 + U_D)} \tag{6}$$

3 Structure and Software Design of the System

3.1 Structure of the System

The structure of the system is shown in Fig. 4. In the process of charging and discharging, the voltage of 13 lithium batteries is detected by battery voltage detection circuit. Then this signal should be send to the MCU through I2C interface. According to the voltage of battery pack, the MCU unit determines its equilibrium strategy and generates PWM control signal to control the equalization unit to balance the battery. MCU communicates with the Upper computer through UART. The Upper computer can show the status of each cell battery.

3.2 Software Design of the System

In the process of battery equalization, current flows into the positive of the battery when a section of the battery energy supplement. It leads to the drop of battery internal resistance voltage. If the battery is sampled voltage at this time, the collected voltage will be error. So the equalization circuit based on these data can't make each section battery energy to reach balance. Therefore, sampling and
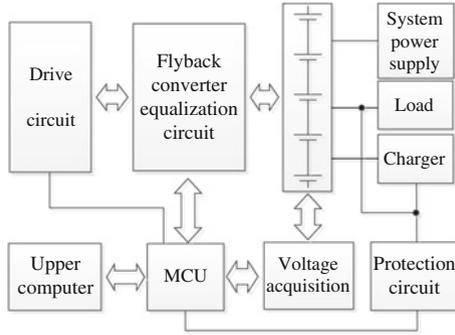
Fig. 4 Structure of the equalization system

equalization is separated in time in order to avoid errors caused by voltage sampling. This system takes 1.5 s for the cycle and takes 0.5 s for voltage sampling and 1 s for energy equilibrium of the battery.

The battery will have the energy loss in the equalization process. A balanced strategy can greatly reduce energy consumption. The system adopts two indicators to determine the need for balance. Taking into account that the voltage sampling accuracy of the system is 4 mV, battery management system begins to balance the battery whose voltage is less than average when the voltage difference between the highest and lowest voltage exceeds 40 mV. The state can be expressed in State = balance. Battery management system finishes balance when the voltage difference between the highest and lowest voltage is less than 20 mV. The state can be expressed in State = unbalance. System flow chart is shown in Fig. 5.
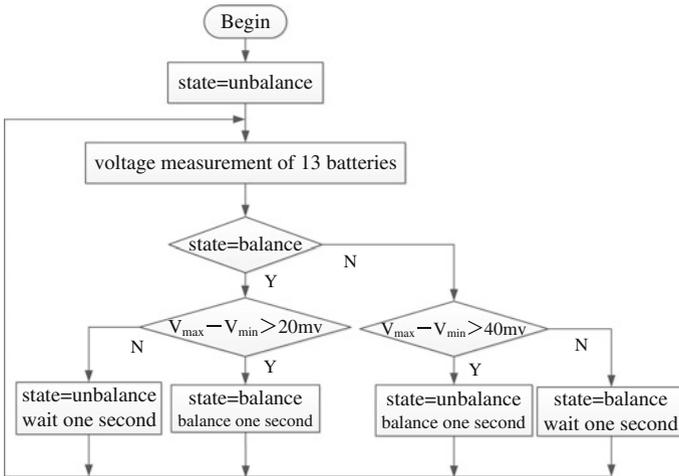
Fig. 5 Equalization algorithm flowchart

4 Experimental Results and Analysis

The circuit board of this battery management system is shown in Fig. 6. It mainly includes battery sampling circuit, the battery equalization circuit, MCU control circuit and protection circuit. We can see PhotoMOS switch in the board and it has not been packaged.

We use PC software to observe the battery voltage waveform. This experiment uses 13 lithium batteries for charging experiment. All initial voltage of batteries are 2.8 V. The unused balance charging waveform is shown in Fig. 7. One of the batteries is first charged when the battery pack is charged 120 min, but other batteries are still not fully charged. In the end, the curve of 13 battery voltages are separated. When the battery performance difference is bigger, the dispersion of voltage curve is more obvious. The charging waveform with using equalization circuit is shown in Fig. 8. All batteries are fully charged when the battery pack is

Fig. 6 Vertical view of hardware circuit
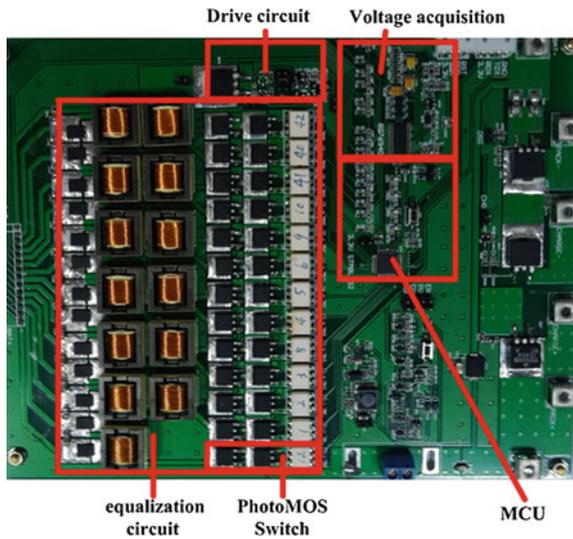
Fig. 7 The unused balance charging waveform
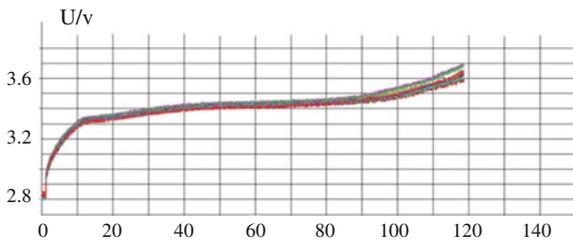
Fig. 8 The charging waveform with using equalization circuit
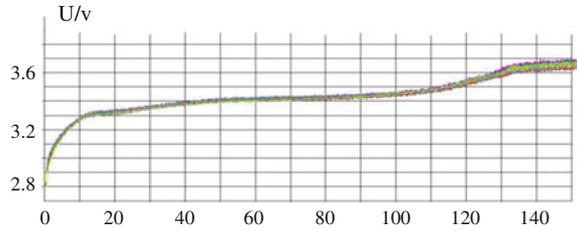
charged 150 min. The voltage difference is very small in the whole charging process. Experimental results show that the battery management system can achieve that all batteries are fully charged through the transfer of battery pack energy.

5 Conclusion

This paper proposes a battery management system (BMS) using the parallel active balancing scheme for the electrical bike application. The balancing circuit topology consists of the PhotoMOS switches and the fly back transformers that share one MOSFET to shift energy from the high energy cells to the low energy cells. The equalization current can reach 2 A, so the equalization speed is faster. The proposed circuit configuration can be expanded to more battery cells with the same circuit topology for practical systems.

References

1. Hasegawa, K., Akagi, H.: A new DC-voltage-balancing circuit including a single coupled inductor for a five-level diode-clamped PWM inverter. IEEE Trans. Indus. Appl. **47**, 841–852. IEEE, Japan (2011)
2. Kuhn, B.T.: Electrical properties and equalization of lithium-ion cells in automotive applications. In: Vehicle Power and Propulsion, 2005 IEEE, pp. 232–239. IEEE, USA (2005)
3. Haq, I.N.T.: Development of battery management system for cell monitoring and protection. In: 2014 international conference on electrical engineering and computer science (ICEECS), pp. 203–208. IEEE, Kuta (2014)
4. Baughman, A.C., Ferdowsi, M.: Double-tiered switched-capacitor battery charge equalization technique. IEEE Trans. Indus. Electron. **55**, 2277–2285, IEEE (2008)
5. Yang, X.J., Jiang, H.R.: Design of a battery management system based on matrix switching network. In: 2015 IEEE international conference on information and automation, pp. 138–141. IEEE, Lijiang (2015)
6. Yang, X.: Design of a battery management system based on matrix switching network. In: Information and Automation, 2015 IEEE International Conference, pp. 138–141. IEEE, Lijiang (2015)

7. Bowkett, M.; Thanapalan, K.; Stockley, T.: Design and implementation of an optimal battery management system for hybrid electric vehicles. In: 2013 19th international conference on automation and computing (ICAC), pp. 1–5. IEEE, London (2013)
8. Ji, W.G., Liu, X., Ji, Y., Tang, Y.B., Ran, F., Peng, F.Z.: Low cost battery equalizer using buck-boost and series LC converter with synchronous phase-shift control. In: 2013 28th annual IEEE applied power electronics conference and exposition (APEC), pp. 1152–1157. IEEE, CA, USA (2013)
9. Daowd, M., Omar, N., Bossche, P., Mierlo, J.: Passive and active battery balancing comparison based on MATLAB Simulation. In: 7th IEEE Vehicle power and propulsion Conference, VPPC'11, pp. 1–7. IEEE, Chicago, IL (2011)
10. Hatami, A.: Power management strategy for hybrid vehicle using a three port bidirectional DC-DC converter. In: Electrical Engineering (ICEE), 2015 23rd Iranian Conference, pp. 1498–1503. IEEE, Tehran (2015)
11. Alvarez, B.L., García, S.V.: Developing an active balancing model and its battery management system platform for lithium ion batteries. In: 2013 IEEE International Symposium on Industrial Electronics (ISIE), pp. 1–5. IEEE, Taipei (2013)

A Chord Curvature Model for Silhouette Curvature Calculation

Bi-Tao Fu

Abstract Silhouettes contain abundant shape information of image objects, which have advantages ranging from invariance to rotation and scale changes, and interference immunity against illumination and noise. Curvature calculation models are applied to detect contour feature points sequentially along silhouette curves. The normal curvature calculation method is often disturbed by coordinate perturbation of adjacent points in a support domain, which is usually caused by rotation transformation and noise interference. We therefore propose a novel curvature calculation method, termed a Chord Curvature Model, which has characteristics consisting of simple calculations, rotation invariance and robustness. Experiments demonstrate that this model can detect contour feature points from complex discrete curves.

Keywords Silhouette curvature calculation · Chord curvature · Contour feature point

1 Introduction

Although contour feature points (CFP) only account for a small percentage of all silhouette points, they can describe silhouette's overall structure accurately. Existing CFP detection methods include polygonal approximation algorithm, corner detection operator and curvature calculation method. According to certain principles, the polygonal approximation algorithm adds feature points successively between two original endpoints to approximate the original curve [1, 2]. The corner detection operators include Moravec, Forstner, Hanna, Harris, Patch-Duplets [3], anisotropic Gaussian kernel [4, 5], etc. Unfortunately, their results are unordered points. While the curvature calculation method can detect points arranged sequentially along

B.-T. Fu (✉)
School of Hydropower and Information Engineering, Huazhong University
of Science and Technology, Wuhan 430074, China
e-mail: fubitao@163.com

silhouette curves, which includes classic curvature algorithm [6–8], L-curvature [9, 10], U chord length curvature [11, 12], visual curvature [13], etc.

The curvature value is apt to be disturbed by coordinate perturbation of adjacent points in support domain, which is caused by scale transformation, rotation transformation or noise interference [9, 11]. To solve it, the author proposed a chord curvature (CCUR) model about CFP detection in this paper.

2 Methodology

2.1 Definition of CCUR

A chord in length of $2L$, midpoint laid on a point p_i on a curve $f(x, y)$ (Fig. 1a), is rotated clockwise at intervals of a small angle $\Delta\theta$. There is a prerequisite that adjacent points p_{i-1} and p_{i+1} must locate on the same side of the chord, in other words, its rotation range locates in the angle θ in Fig. 1b. The gray area enclosed by the curve, chord's perpendicular on two endpoints, and the chord is $A = A_1 + A_2$. It should be noted that the gray area must locate at the same side of the chord. That of the j th rotation is $A_j = A_{j1} + A_{j2}$, $j = 1, 2, \ldots, m$, m is total rotation times. The CCUR is defined as

$$K = \frac{\min(A_j)}{L^2} \tag{1}$$

Suppose an equal-area transformation from $\min(A_j)$ in Fig. 1a to the gray area in Fig. 2. There are two equal included angles α between two symmetrical lines $p'_{i-1}p_i$, $p_ip'_{i+1}$ and chord, so the gray area in Fig. 2 is

$$\min(A_j) = 2 \times L \times L \tan \alpha/2 = L^2 \tan \alpha \tag{2}$$
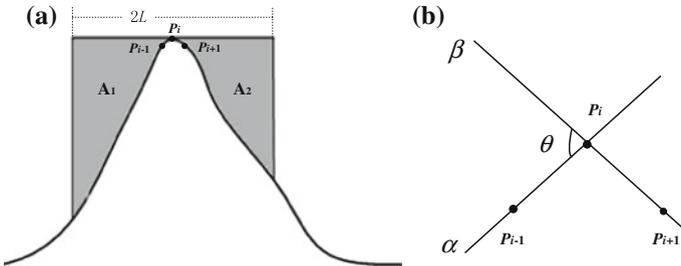
Fig. 1 The diagram of CCUR calculation. **a** CCUR, **b** rotation range of chord

Fig. 2 The diagram of
CCUR's physical meaning
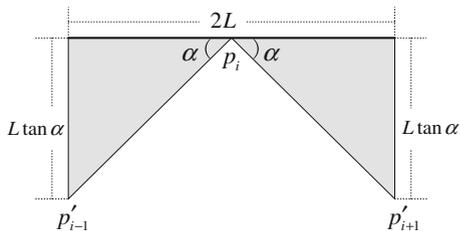
and

$$\kappa = \frac{\min(A_j)}{L^2} = \tan \alpha \qquad (3)$$

So CCUR is the slope of $p'_{i-1}p_i$ for chord as x-axis, in fact it is the synthetic curvature near p_i. Its physical meaning is simple and clear: The more protruding the p_i, the larger α becomes, and the larger κ becomes. When both p_{i-1} and p_{i+1} locate on chord, p_i, p_{i-1} and p_{i+1} are collinear and $\kappa = 0$.

2.2 Calculation Procedures of CCUR

Preprocessing procedures are necessary to simplify silhouette curves and reduce calculation amount:

1. Image preprocessing. It includes image enhancement, image denoising, image filtering, image deblurring, image smoothing, etc.
2. Edge preprocessing. Many existing edge detection operators can be adopted, but before that, the edge preprocessing procedures is necessary. In fact, there are always plenty of disconnected edge, double edge, jagged edge, and massive short fragmental contours.
3. Coordinate transformation. The Cartesian coordinate system is convenient to calculate area and the pedal coordinates of point to the curve, whereas the polar coordinate system or polar coordinates is convenient to process rotation, so the coordinate transformation is necessary. In polar coordinates for the pole with p_i, $\theta + \Delta\theta$ and constant ρ mean the rotation of chord. While in Cartesian coordinates, rotated chord becomes x-axis, and the distance between curve points to chord is y coordinate, so the calculation of area A_j becomes a simple thing. Set $M \approx 5L$, the neighborhood point set of p_i is $p_{i-M}, \ldots, p_{i-1}, p_i, p_{i+1}, \ldots, p_{i+M}$. To speed up above calculation, the author substituted the point set for all curve points in the calculation.
4. Filter preprocessing by included angle. The calculation of included angle is much easier than that of CCUR, so it can decrease calculation time markedly. Simple loop computation is applied to calculate the included angle cluster $\angle p_{i-s}p_ip_{i+t}$, among of which, $1 \leq s \leq 5$, $1 \leq t \leq 5$. If there is an included angle between $0.95\pi \sim 1.05\pi$, p_i is not a CFP certainly and does not need subsequent calculation.
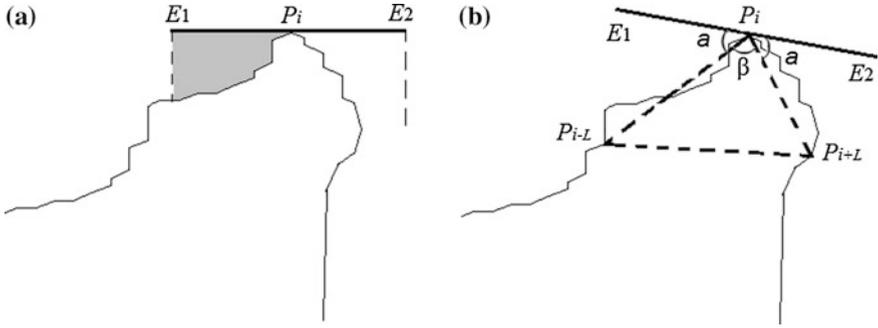
Fig. 3 The diagram of dangle points. **a** dangle point, **b** CCUR of dangle point

2.3 Dangle Endpoints

Perhaps there are dangle endpoints when the chord lay on the sharp turn points as shown in Fig. 3a. The endpoint E2 is a dangle endpoint, that is to say, there isn't an intersection between perpendicular line and the neighborhood point sets of p_i, so the closed area can't be calculated, and the formula 1 is unsuitable for it.

The author designed an alternative simplified method to solve it. As shown in the Fig. 3b, the included angle $\beta = \angle p_{i-L} p_i p_{i+L}$ can be calculated by their coordinates, and $\alpha = (\pi - \beta)/2$, and then $\kappa = \tan\alpha$, which can be seen as the CCUR of p_i.

3 Experiment and Conclusion

3.1 Experiment I—CCUR Values

The experiment data I is a giraffe image (Fig. 4a). The author compiled the corresponding MATLAB program to realize above model and algorithms. When the sampling scale is $L = 5$, the detected points are CFPs.

By a series of statistical analysis experiments, threshold $th = \tan\alpha_0$ is ascertained. When $\kappa > th$, p_i is judged as a CFP. Too small α_0 will result in overmuch CFPs; otherwise too big α_0 will lead to CFP lack. Three corresponding threshold is: when $L = 5$, $th = 0.8$; when $L = 10$, $th = 0.6$; when $L = 20$, $th = 0.45$.

There are too many CFPs on the same section of curve (Fig. 4b), so non-maximum value suppression is necessary. In this paper a simple algorithm is
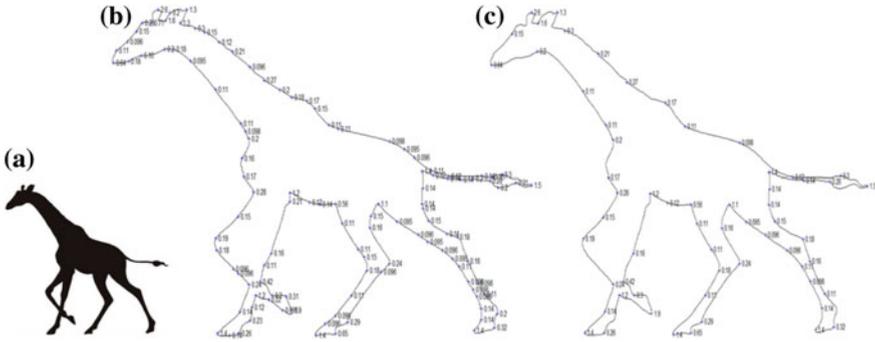
Fig. 4 The CFPs of giraffe image. **a** Original image, **b** *th_dis* = 10, **c** *th_dis* = 20

designed to screen CFPs by their distance: if the distance between two CFPs is less than a given distance threshold *th_dis*, the smaller CFP will be eliminated (Fig. 4c).

3.2 Experiment II—Robustness of CCUR

The experiment II is designed to compare the CCUR and the classic mathematics curvature method, and verify their anti-interference performance against coordinate perturbation of adjacent points. The experiment data is the discrete point set of a unit circle (Fig. 5a).

A comparison method adopted here is a classic mathematics curvature calculation after discrete curve fitting [7]. Obviously the mathematics curvature of a unit circle is $K = 1.0$, exactly the reciprocal of radius, but CCUR is no longer equivalent to it. CCUR values have no longer any actual physical implications, and the only significance is their relative sizes, which reveal the possibility of CFP. CCUR values vary with chord length: when $L = 1.0$, $\kappa = 0.429$; $L = 0.5$, $\kappa = 0.174$; $L = 0.2$, $\kappa = 0.067$; $L = 0.1$, $\kappa = 0.0334$.
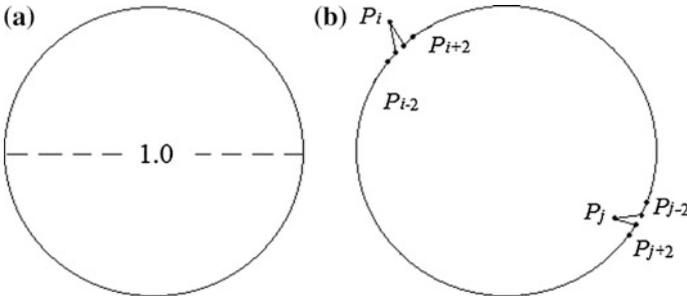
Fig. 5 The experiment data II-unit circle

There are two noise points to verify their anti-interference performance, among them, p_i is a protruding point, and p_j is a concaving point (Fig. 5b). And p_{i-2}, p_{i+2}, p_{j-2}, p_{j+2} are their neighborhood points. In contrast to Fig. 5a, their mathematics curvatures change sharply: $K_{i-2} = 3.07$, $K_{i+2} = 4.16$, $K_{j-2} = 0.82$, $K_{j+2} = 0.74$. Their change rate is 207, 316, −18, −26 %. When $L = 0.2$, the CCUR is $\kappa_{i-2} = \kappa_{i+2} = 0.064$, $\kappa_{j-2} = \kappa_{j+2} = 0.071$, change rate is −4.48, −4.48, 5.97, 5.97 %. Obviously change rate of the CCUR is far less than that of the former. The reason is that the area disturbance is far less than the shape disturbance of fitted curves, so curvature model by calculation of closed area is more robust.

4   Conclusion

CCUR calculation merely requires the minimum area closed by chord and curve, so it costs a small calculation amount. It has the properties of rotation invariance, noise immunity, robustness, and suitability of various discrete curves. It can detect CFPs on various scale silhouette by altering chord length.

References

1. Wang, J.W.: Research on shape contour analysis and matching based on combined global and local information. Ph.D. thesis. Huazhong University of Science and Technology, Wuhan (2012)
2. Carmona-Poyato, A., Madrid-Cuevas, F.J., Medina-Carnicer, R.: Polygonal approximation of digital planar curves through break point suppression. Pattern Recogn. **43**, 14–25 (2010)
3. Johansson, B., Moe, A.: Patch-duplets for object recognition and pose estimation. In: The 2nd Canadian Conference on Computer Vision Proceedings, pp. 444–451. IEEE Computer Society, Piscataway, NJ, USA (2005)
4. Zhang, W.: Image edge and corner detection based on anisotropic Gaussian kernels. Ph.D. thesis. Xidian University, Xian (2012)
5. Akagunduz, E.: Shape recognition using orientational and morphological scale-spaces of curvatures. IET Comput. Vision **9**, 750–757 (2015)
6. Kronenbergera, M., Wirjadib, O., Freitag, J., Hagen, H.: Gaussian curvature using fundamental forms for binary voxel data. Graph. Models **6**, 1–14 (2015)
7. Worring, M., Smeulders, A.: Digital curvature estimation. CVGIP: Image Understanding **58**, 366–382 (1993)
8. Koenders, R., Lindenbergh, R.C., Storms, J.E., Menenti, M.: Multiscale curvatures for identifying channel locations from DEMs. Comput. Geosci. **68**, 11–21 (2014)
9. Zhong, B.J.: Research on algorithms for planar contour processing. PhD. thesis. Nanjing University of Aeronautics and Astronautics, Nanjing (2006)
10. Zheng, R.L., Zhong, B.J., Xu, D.S.: Scale-space shape analysis based on L curvature. J. Nanjing univ. (Nat. Sci.), **48**(2), 172–181 (2012)

11. Guo, J.J., Zhong, B.J.: U-Chord curvature: a computational method of discrete curvature. Pattern Recog. Artif. Intell. **27**(8), 683–691 (2014)
12. Chen, B.B., Deng, X.P., Yang, J.G.: Orientation-invariant generalized Hough transform algorithm based on U-chord curvature. J. Comput. Appl. **35**(9), 2619–2623 (2015)
13. Liu, H.R.: Curvature representation and decomposition of shape. Ph.D. thesis. Huazhong University of Science and Technology, Wuhan (2009)

A Low-Complexity Decoding Algorithm for Quasi-orthogonal Space-Time Block Codes

Xiao-Ping Jin, You-Ming Li, Zheng-Quan Li and Ning Jin

Abstract In this paper we discuss full-rate quasi-orthogonal space-time block code with a low decoding problem. Based on the traditional Bilateral Jacobi transformation, we propose a new decoding algorithm which can reduce the multiplication and root computation complexity at the Multiple Input Multiple Output (MIMO) receivers. Simulation results show that the bit error probability of our scheme is comparable to that of the traditional algorithms but its computation complexity is much lower than the traditional algorithms.

Keywords QO-STBC (quasi-orthogonal space time block codes) · Low complexity decoding algorithm · Matrix jacobi transformation

1 Introduction

Ever since H. Jafarkhani, Choi and Nimmagadda proposed the Quasi-Orthogonal Space-Time Block Codes (QO-STBC) [1–3], there has been considerable interest in codes with low decoding complexity. A variety of methods have been proposed to

X.-P. Jin · Y.-M. Li (✉)
Institute of Communication Technology, Ningbo University,
Ningbo 315211, China
e-mail: liyouming@nbu.edu.cn

X.-P. Jin
e-mail: jxp1023@cjlu.edu.cn

X.-P. Jin · Z.-Q. Li · N. Jin
Department of Information Engineering, China Jiliang University,
Hangzhou 310018, China
e-mail: 10a0303117@cjlu.edu.cn

N. Jin
e-mail: jinning1117@cjlu.edu.cn

reduce decoding complexity [4]. M.Y. Chen constructs a class of QO-STBCs based on symbol-by-symbol maximum likelihood detection by Chen [5], but the complexity of this decoding method is still high. Leuschner stated a kind of QO-STBCs using a sphere decoding algorithm to reduce decoding complexity [6]. However, this detection method sometime has high decoding latency at low SNR (Signal to Noise Ratio), so the complexity still remains high.

However, those algorithms cannot attain full-diversity, full rate transmission. A QO-STBC with full-diversity, full-rate transmission and double-symbol decoding has been proposed for a four transmit antennas system [7, 8]. However, the decoding complexity is high due to the nonlinear characteristic [9–12]. To reduce the complexity, a class of three- and four-antenna QO-STBCs based on a given rotation scheme is proposed, and the decoding complexity is linear [13], because some element of channel matrix are canceled. However, these algorithms need 16 times the multiplication and double the square root operations which are relative to the channel matrix procedure, still maintain high complexity. So a fast decoding algorithm using a Jacobi transform is proposed to reduce the computation complexity at the receiver. Simulation results are provided to compare the bit error probability between the proposed scheme and the conventional method.

2 System Model

Consider a MIMO wireless communication system with N transmits antennas and one receive antenna. The number of transmit time slots is T. The received signal is given by

$$r = Sh + n \tag{1}$$

where $\mathbf{r} = (r_1, r_2, \ldots, r_T)^T$ is a received signal vector. $r_t(t = 1, \ldots, T)$ is a received signal at t time slot, $[.]^T$ represents transpose. $N \times 1$ matrix $\mathbf{h} = (h_1 \; h_2 \ldots h_N)$ represents the channel gain between the N transmitters and the receiver, and $h_m(m = 1, \ldots, N)$ is assumed to be independent and identically distributed (i.i.d) complex Gaussian random sample, whose real and imaginary variance are both 0.5, $\mathbf{n} = (n_1 \; n_2 \ldots n_T)^T$ is the addictive noise matrix, where n_t $(t = 1, 2, \ldots, T)$ denotes the i.i.d zero-mean and unit-variance complex Gaussian random noise variable. According to [1], the 4×4 transmitted code matrix $\mathbf{S}(2)$ denotes a full-rate quasi-orthogonal space-time block codes matrix.

$$\mathbf{S} = \begin{pmatrix} \tilde{s}_1 & 0 & \tilde{s}_3 & 0 \\ -\tilde{s}_3^* & 0 & \tilde{s}_1^* & 0 \\ 0 & \tilde{s}_2 & 0 & \tilde{s}_4 \\ 0 & -\tilde{s}_4^* & 0 & \tilde{s}_2^* \end{pmatrix} \tag{2}$$

where $\tilde{s}_1 = \alpha s_1 + \beta s_3$, $\tilde{s}_2 = \alpha s_2 + \beta s_4$, $\tilde{s}_3 = \beta s_1 - \alpha s_3$, $\tilde{s}_4 = \beta s_2 - \alpha s_4$. Here, $\alpha = \sqrt{(P/2) - \lambda}$ and $\beta = \sqrt{\lambda}, (0 < \lambda < P/2)$ are the power scaling of M-ary constellation s_i (i = 1, 2, 3, 4), P is the total transmit power of the N transmit antennas in each slot, which is normalize t to be 1.$[.]^*$ represents the conjugate.

3 Decoding Algorithm

Conducting an appropriate transformation of the receiving signal $\mathbf{r}$ at the receiver-end, we take conjugate of r_2, r_4, which are vectors of the received signal in the second time slot and fourth time slot, respectively. Then we transform matrix $\mathbf{S}$ with drawing sent signal vector $\mathbf{s} = (s_1 \ s_2 \ s_3 \ s_4)^T$, and channel gain vector $\mathbf{h}$ becomes channel matrix $\mathbf{H}$. The received signal becomes $\mathbf{r}' = \mathbf{H}\mathbf{s} + \mathbf{n}'$. Next both sides of the equation are left multiplied by $\mathbf{H^H}$, $[\cdot]^H$ represents conjugate and transpose, the receiver signal becomes as

$$\hat{\mathbf{r}} = \underbrace{\begin{pmatrix} e\alpha & 0 & e\beta & 0 \\ 0 & f\alpha & 0 & f\beta \\ e\beta & 0 & -e\alpha & 0 \\ 0 & f\beta & 0 & -f\alpha \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{pmatrix} + \begin{pmatrix} h_1^* n_1 + h_3 n_2^* \\ h_2^* n_3 + h_4 n_4^* \\ h_3^* n_1 - h_1 n_2^* \\ h_4^* n_3 - h_2 n_4^* \end{pmatrix} == \mathbf{As} + \hat{\mathbf{n}} \quad (3)$$

where $e = |h_1|^2 + |h_3|^2, f = |h_2|^2 + |h_4|^2$, and

$$\mathbf{H} = \begin{pmatrix} h_1 & 0 & h_3 & 0 \\ h_3^* & 0 & -h_1^* & 0 \\ 0 & h_2 & 0 & h_4 \\ 0 & h_4^* & 0 & -h_2^* \end{pmatrix} \quad (4)$$

Matrix $\mathbf{A}$ in (3) can be expressed by (a_{ij}). It is found that if a_{13} and a_{24} both equal to 0, the self-interference terms can be totally avoided. So we need to transform symmetric matrix $\mathbf{A}$ in formula (3) into a diagonal matrix $\mathbf{A}'$ with bilateral Jacobi methods.

To perform the transformation, first decompose matrix $\mathbf{A}$ to find a diagonal matrix $\mathbf{K} = \text{diag}(k_1, k_2, k_3, k_4)$ satisfying $\mathbf{A} = \mathbf{KBK}$. Because matrix $\mathbf{K}$ is a diagonal matrix, after decomposition of matrix $\mathbf{A}$, matrix $\mathbf{B}$ is a symmetric matrix too. According to the traditional decomposition method, we find a diagonal matrix $\mathbf{K}' = \text{diag}(k_1', k_2', k_3', k_4')$ to decompose unitary matrix $\mathbf{R}$, and make $\mathbf{R} = \mathbf{K}'\mathbf{TK}^{-1}$. Next, multiply the right side of this formula by $\mathbf{KT}^{-1}$. Then we get the relationship between matrix $\mathbf{T}$ and matrix $\mathbf{K}$, $\mathbf{K}' = \mathbf{RKT}^{-1}$, where $\mathbf{T}$ meets the following formula: $\mathbf{T} = \mathbf{I} - (1 - t_{ii})e_i e_i^T - (1 - t_{jj})e_j e_j^T + t_{ij}e_i e_j^T + t_{ji}e_j e_i^T$, and t_{ii} and t_{jj} are

diagonal elements of matrix $\mathbf{T}$. t_{ij} and t_{ji} are non-diagonal elements of matrix $\mathbf{T}$. Which yields:

$$\mathbf{R} = \mathbf{K}'\mathbf{T}\mathbf{K}^{-1} = \begin{pmatrix} \frac{k_1'}{k_1}t_{11} & 0 & \frac{k_1'}{k_3}t_{13} & 0 \\ 0 & \frac{k_2'}{k_2} & 0 & 0 \\ -\frac{k_3'}{k_1}t_{31} & 0 & \frac{k_3'}{k_3}t_{33} & 0 \\ 0 & 0 & 0 & \frac{k_4'}{k_4} \end{pmatrix} = \begin{pmatrix} c & 0 & s & 0 \\ 0 & 1 & 0 & 0 \\ -s & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5)$$

Now we have $\mathbf{A}' = \mathbf{R}\mathbf{A}\mathbf{R}^{\mathrm{T}} = \mathbf{K}'\mathbf{T}\mathbf{B}\mathbf{T}^{\mathrm{T}}\mathbf{K}' = \mathbf{K}'\mathbf{B}'\mathbf{K}'$. From the equation, we can see that the transformation from symmetric matrix $\mathbf{A}$ to diagonal matrix $\mathbf{A}'$ is equivalent to the transformation from symmetric matrix $\mathbf{B}$ to diagonal matrix $\mathbf{B}'$. In other words, in order to figure out how many steps in the transformation of symmetric matrix $\mathbf{A}$, we only need to solve the step number in the transformation of symmetric matrix $\mathbf{B}$.

According to Jia's paper [10], setting the proper value of matrix $\mathbf{K}'$, the multiplication and the operation number of square root will cut in half through transforming matrix $\mathbf{B}$ We find from the above process that the computational complexity has been reduced by half when using this fast algorithm. Now we examine the case where a_{13} turns to zero, and turns into $\mathbf{A}'$. The received signal of (3) has the following expression

$$\tilde{\mathbf{r}} = \mathbf{K}'\mathbf{B}'\mathbf{K}'\mathbf{s} + \mathbf{K}'\mathbf{B}'\mathbf{K}'\mathbf{A}^{-1}\hat{\mathbf{n}}$$

$$= \underbrace{\begin{pmatrix} l_1'b_{11}' & 0 & 0 & 0 \\ 0 & l_2'b_{22}' & 0 & a_{24} \\ 0 & 0 & l_3'b_{33}' & 0 \\ 0 & a_{24} & 0 & l_4'b_{44}' \end{pmatrix}}_{\mathbf{A}'} \mathbf{s} + \tilde{\mathbf{n}} \quad (6)$$

Similarly, transform matrix $\mathbf{A}'$ with Jacobi methods can make $a_{24} = a_{42} = 0$. Then the matrix $\mathbf{A}'$ turns into diagonal matrix $\mathbf{A}''$. Thus we can use maximum likelihood decoding algorithm to decode the received signals. It can be found out that we use the Jacobi transform twice, but only need 16 multiplications and 16 additions and one square root operation when computing c and s. The comparison of computational complexity between Park's paper [13] and the fast algorithm is shown in Table 1. The expression of maximum likelihood estimation is given by (7), $\|\cdot\|$ denotes the norm of a matrix.

Table 1 Comparison of complexity for the two algorithms

Algorithm type	Addition	Multiplication	Open square
The algorithm proposed by Park [13]	16	32	2
Fast algorithm	16	16	1

$$\hat{\mathbf{S}} = \arg \min_{\mathbf{s}} \|\mathbf{r} - \mathbf{Sh}\|^2 = \arg \min_{\mathbf{s}} \|\mathbf{H}^H \hat{\mathbf{r}} - \mathbf{As}\|^2$$

$$= \arg \min_{\mathbf{s}} \|\mathbf{H}^H \hat{\mathbf{r}} - (\mathbf{K}')^2 \mathbf{B}' \mathbf{s}\|^2 \tag{7}$$

4 Simulation

In all simulations, we consider four transmit antennas and one receive antenna. The whole transmit slot $T = 4$. We provide simulation results to verify the efficiency of our proposed algorithm at the different signal power P and λ (which denote p and q values respectively in Fig. 1). Assuming that the communication system utilizes QPSK (Quadrature Phase Shift Key) to modulate and demodulate. The receiver uses fast Jacobi decoding algorithm to detect the received signals. The bit error rate (BER) of the proposed algorithm with the traditional decoding algorithm proposed by Park et al., Jia and Xiao, Xiao and Dai at different signal-to-noise ratio (SNR) is compared in Fig. 1. [10, 11, 13].
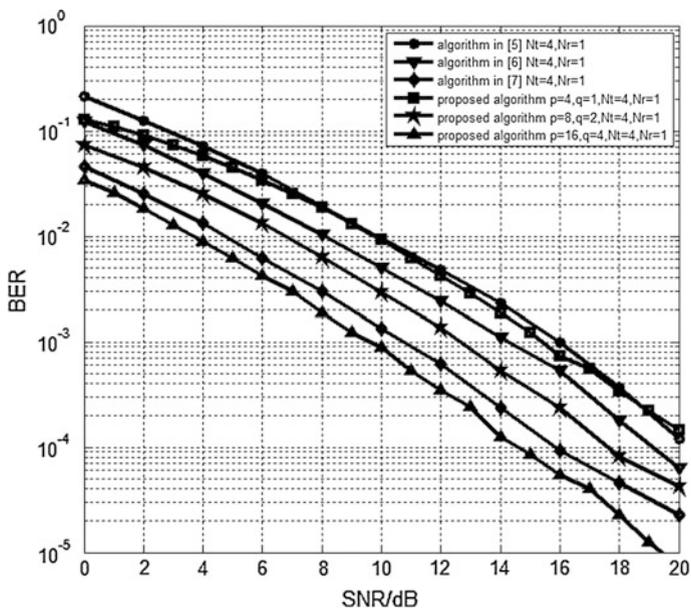
Fig. 1 BER comparison of algorithm proposed by Park et al., Jia and Xiao, Xiao and Dai [10, 11, 13] and proposed algorithm against $T = 4$ and QPSK

From these figures, we see that the BER performance of the system using the fast Jacobi decoding algorithm is improving with the increasing the signal power P than the algorithms in [10, 11, 13]. The proposed decoding algorithm can improve the performance but with lower computation complexity. Furthermore, the number of Multiplication and Open Square decreases by about 50 % compared to the algorithm proposed by Park et al. [13]. Furthermore, the algorithm proposed by Xiao and Dai [11] requires the computation of the inverse of the matrix, so it is more complex than the fast algorithm, which is approximately equal to the algorithm proposed by Jia and Xiao [10].

5   Conclusion

In this paper, we construct a new full-rate quasi-orthogonal space-time block code. Based on the traditional Bilateral Jacobi transformation decoding algorithm, we propose a new fast algorithm, which can reduce the multiplication and rooting computation complexity at the receiver. Simulation results show that the system bit error probability of our scheme is comparable to the previously proposed algorithm while the computation complexity is much lower. This has important significance for the signal processing of the MIMO communication system.

References

1. Jafarkhani, H.: A quasi-orthogonal space-time block code. IEEE Trans. Commun. **49**, 1–4 (2001)
2. Choi, M., Lee, H., Nam, H.: Error rate analysis of quasi-orthogonal space–time block codes with power scaling. Electron. Lett. **50**(25), 1940–1942 (2014)
3. Nimmagadda, S.M., Sajja, S.G, Bhima, P.R.: Quasi-Orthogonal space-time block codes based on circulant matrix for eight transmit antennas. In: International Conference on IEEE Communications and Signal Processing (ICCSP), , 1953–1957. IEEE, Piscataway (2014)
4. Shen, Q.J.: The faster algorithms of plane rotation. Numerical mathematics. J. Chin. Univ. **1**, 82–86 (1980)
5. Chen, M.Y., Cioffi, J.M.: Space-time block codes with symbol-by-symbol maximum-likelihood detections. IEEE J. Selected Topics Signal Process. **3**(6), 910–915 (2009)
6. Leuschner, J., Yousefi, S.: On the ML decoding of quasi-orthogonal space-time block codes via sphere decoding and exhaustive search. IEEE Trans. Wireless Commun. **7**(11), 4088–4093 (2008)

7. Ahmadi, A., Talebi, S., Shahabinejad, M.: A new approach to fast decode quasi-orthogonal space-time block codes. IEEE Trans. Wireless Commun. **14**, 165–176 (2015)
8. Astal, M.-T.E.L., Abu-Hudrouss, A.M., Olivier, J.C.: Improved signal detection of wireless relaying networks employing space-time block codes under imperfect synchronization. Wireless Pers. Commun. **82**, 1–18 (2014)
9. Lee, H.: Robust full-diversity full-rate quasi-orthogonal STBC for four transmit antennas. Electron. Lett. 24th, **45**(20), 1004–1005 (2009)
10. Jia, M.R., Xiao, L.P.: Improved QO-STBC for three and four transmit antennas. J. China Univ. Posts Telecommun. **17**(4), 42–46 (2010)
11. Xiao, L.P., Dai, X.X.: A novel QO-STBC scheme with linear decoding. J. China Univ. Posts Telecommun. **5**(5), 53–57 (2011)
12. Ranju, M.A.A.A., Rafi, R.S.: Analysis of rate one quasi-orthogonal space-time block coding over rayleigh fading channel. Int. J. Commun. Netw. Syst. Sci. **07**, 196–202 (2014)
13. Park, U., Kim, S.Y., Lim, K., Li, J.: A novel QO-STBC scheme with linear decoding for three and four transmit antennas. IEEE Commun. Lett. **12**(12), 868–870 (2008)

A MAC Protocol for Medical Emergency Monitoring of Wireless Body Area Networks

Chong-Qing Zhang, Ying-Long Wang, Yong-Quan Liang, Ming-Lei Shu, Chang-Fang Chen, Huan-Qing Cui and Sheng Fang

Abstract An adaptive medium access control (MAC) protocol is specially designed for medical emergency monitoring to improve energy efficiency and satisfy data transmission delay requirements. By adopting a long superframe structure to avoid unnecessary beacons and allocating most of the superframe to inactive periods, the duty cycle is reduced to an extremely low level. Short active time slots are inserted into the superframe and shared by all nodes to deliver emergency data in a low-delay and reliable framework. Experimental results show the proposed MAC protocol works well for WBANs with low emergency data flow.

C.-Q. Zhang (✉) · Y.-Q. Liang · H.-Q. Cui · S. Fang
College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China
e-mail: zhangchongqing@sdust.edu.cn

Y.-Q. Liang
e-mail: lyq@sdust.edu.cn

H.-Q. Cui
e-mail: cuihuanqing@gmail.com

S. Fang
e-mail: fangs99@126.com

C.-Q. Zhang · Y.-L. Wang · M.-L. Shu · C.-F. Chen
Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center in Jinan), Jinan 250101, China
e-mail: wangyl@sdas.org

M.-L. Shu
e-mail: shuml@sdas.org

C.-F. Chen
e-mail: chencf@sdas.org

1 Introduction

Medical emergency monitoring WBANs are helpful for the daily care of the elderly and chronically ill people [1]. Generally speaking, a medical emergency monitoring WBAN is composed of one coordinator and a small number of nodes. The nodes measure the events and deliver the data to the coordinator as needed [2]. The coordinator then informs the patient and related medical staff so that they can take timely measures to deal with the emergencies.

Medical emergency monitoring WBANs are expected to work for a considerable length of time, e.g., several years. This makes energy management becomes an apparent need for severely energy-constrained nodes. For a sensor node, radio is generally the part which consumes most energy. In order to extend the lifetime of a node from days to years, the only way is duty cycling the radio [3], and this is the task of the Medium Access Control (MAC) layer which drives the radio hardware.

Some existing MAC protocols [4–6] can be exploited to transfer the emergency traffic in the case of an emergency. Yet these protocols do not work well for the sleeping medical emergency monitoring WBANs because they are not specially designed for such applications. For one thing, the energy efficiencies of these protocols can be improved further. Secondly, these protocols lack an appropriate mechanism to immediately awaken the network. An energy-efficient MAC protocol was proposed by Shu et al. to transmit urgent data for WBANs with heavy traffic, yet this protocol did not consider WBANs with sparse traffic [7].

In this paper, we extend the MAC protocol proposed by Shu et al. [7] and specially design a beacon-enabled MAC protocol to meet the demands of sleeping medical emergency monitoring WBANs. The rarity of traffic in sleeping mode is taken into account to reduce the duty cycle to an extremely low level. Short active time slots are inserted into the superframe and shared by all the nodes to deliver the emergency data in a low-delay and reliable way. The effectiveness of this method is verified by simulation results.

2 MAC Protocol Design

Figure 1 displays the structure of the superframe. As a subsequent research of the work by Shu et al. [7], the MAC protocol proposed in this paper is also called I-MAC. The superframe begins with a beacon frame that may be followed by an optional broadcast period and an optional contention access period (CAP). The broadcast period is used by the coordinator to broadcast long frames to all the nodes. The CAP1 period can be used by the coordinator to transmit a long frame to one node, or can be used by one node to transmit a long frame to the coordinator.

Following the CAP1 period is a long inactive period into which many inserted short interrupt slots which are shared by all the nodes to transmit data or network commands. There is an optional CAP2 period at the end of a superframe. If there
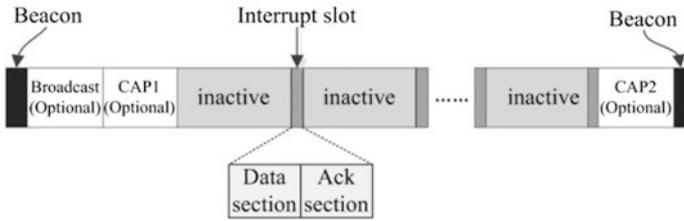
Fig. 1 Structure of I-MAC superframe

are frames collided, the coordinator activates such a CAP2 period to retransmit these frames. One interrupt slot contains two sections, one DATA section and one ACK section. A node can use the DATA section to transmit its data to the coordinator. The coordinator can also use to the DATA section to transmit its data to one node. The ACK section is used to acknowledge the data or commands.
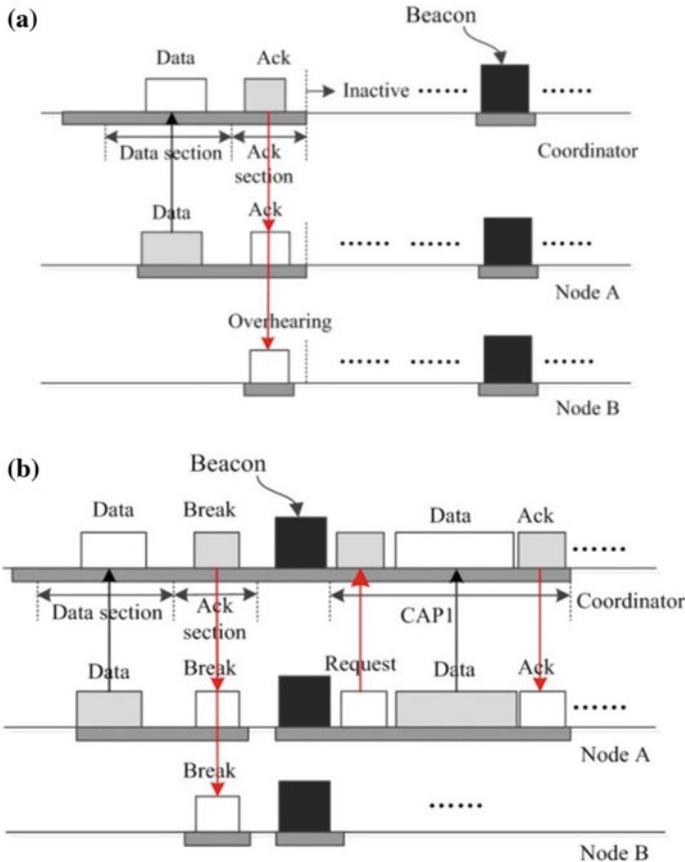
Fig. 2 A node sends a small data or a big data to the coordinator
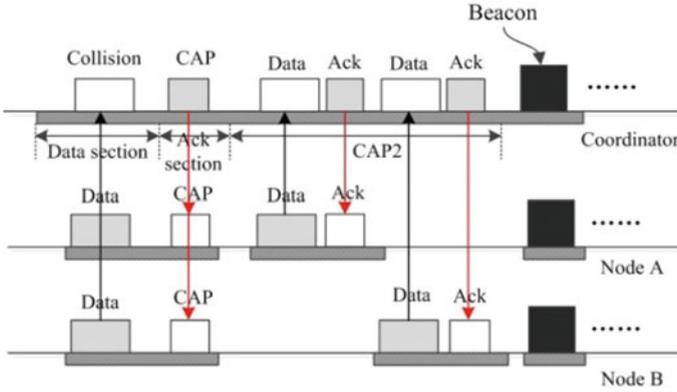
Fig. 3 The case that there is a collision happens

How a node *A* sends a small data or a big data to the coordinator is shown by Fig. 2a, b. As for a small data, it is encapsulated entirely into a data frame which can be transmitted using a DATA section. After this, the sender sends the data out. Then the coordinator acknowledges the frame using an ACK frame. If the sender gets a big emergency data to deliver, it issues a request frame containing a data size field to the coordinator. After receiving the request frame, the receiver breaks the current superframe with a "BREAK" frame. This "BREAK" frame opens a brand new superframe in which sender *A* will use CAP1 period to transmit its data.

Figure 3 illustrates how two sensors deliver their data to the coordinator in an interrupt slot. In such a case, the receiver cannot recognize these frames due to the frame collision. In the following ACK section the coordinator broadcasts a "CAP" frame. This "CAP" frame opens a CAP2 period for all nodes to send their data. Then the coordinator announces a new beacon to commence a new superframe.

3 Performance Evaluation

Four MAC schemes are compared when comparing the energy consumption of nodes. IEEE 802.15.4 is one scheme, and the others are three I-MAC variations. The three I-MAC variations are distinguished by the interrupt interval and the number of interrupt slots in a superframe. Let *BI* be the length of the superframe and I_{Int} be the length of an interrupt slot, then I_{Int} and $NI = BI/I_{Int}$ can be used to distinguish three I-MAC variations.

How the average powers of a node and the coordinator change with the average data occurrence interval is displayed by Fig. 4. As shown by Fig. 4a, as the data occurrence interval grows the averaged powers of all 4 MAC schemes decrease. Three I-MAC variations cost much higher energy than 802.15.4 when the data

Fig. 4 Effect of average event occurrence interval on node/coordinator average power
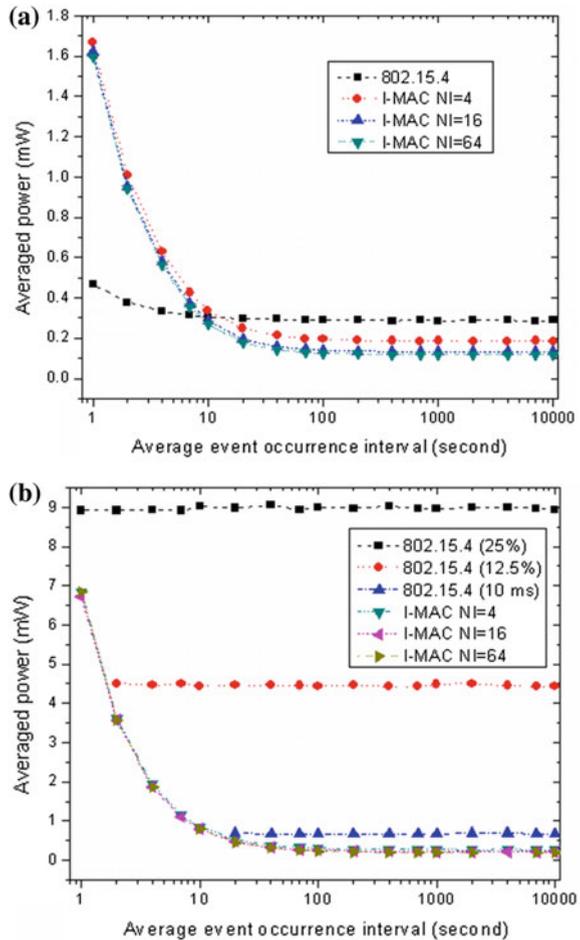
occurrence interval is very small. And these energy powers lower quickly to be less than 802.15.4 as the data interval increases. As Fig. 4b shows, for three 802.15.4 schemes, their coordinator energy consumptions keep constant because the duty cycle keeps constant. Three I-MAC schemes consume high powers as the data interval takes very small values, and these powers lower quickly to be less than 802.15.4 as the data interval increases. This can be explained by the frequent frame collisions when the data interval is small.

The average frame delays of four MAC schemes are shown and compared in Fig. 5. All four MAC schemes can meet the delivery delay demands of the emergency events. The average frame delay of IEEE 802.15.4 stays very stable as the data interval increases from 1 to 10,000 s. While for other three schemes, the average delays display a decrease as the data interval increases. In other words, these average delays show an increase as the data interval decreases. This reason

Fig. 5 The effect of the
average event interval on
average frame delay
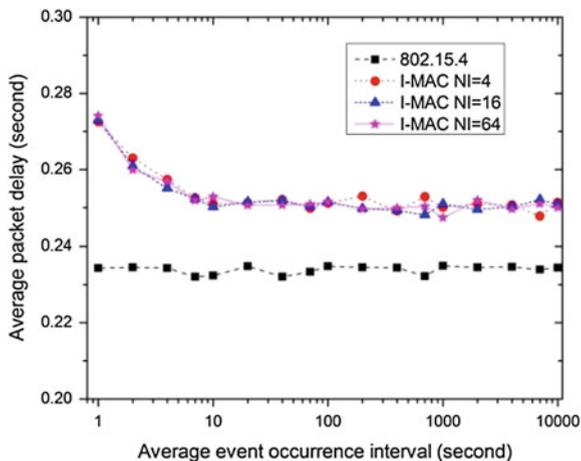
behind this phenomenon also is rooted in the additional CAP periods and beacons caused by collided frames.

4   Conclusion

By reducing the number of beacon frames and adopting short interrupt slots, I-MAC can obtain high energy efficiency. When the data occurrence probability is low, I-MAC is energy-efficient and can meet the real-time demands simultaneously. As for high emergency occurrence frequency, I-MAC cannot achieve high energy efficiency because the frequent collisions introduced by dense emergency data frames. As a consequence, I-MAC is not applicable to the applications that have high emergency event occurrence frequency.

References

1. Cavallari, R., Martelli, F., Rosini, R., Buratti, C., Verdone, R.: A survey on wireless body area networks: technologies and design challenges. IEEE Commun. Surv. Tutorials. **16**(3), 1635–1657 (2014)
2. Movassaghi, S., Abolhasan, M., Lipman, J., Smith, D., Jamalipour, A.: Wireless body area networks: a survey. IEEE Commun. Surv. Tutorials. **16**(3), 1658–1686 (2014)

3. Langendoen, K., Meier, A.: Analyzing MAC protocols for low data-rate applications. ACM Trans. Sens. Netw. **7**(2), 1–40 (2010)
4. Gama, O., Simoes, R.: A hybrid MAC scheme to improve the transmission performance in body sensor networks. Wireless Pers. Commun. **80**(3), 1263–1279 (2014)
5. IEEE Standard for Information Technology Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs), IEEE Standard 802.15.4 Working Group Standard (2006)
6. IEEE 802.15 (2013) WPAN task group 6 body area networks (BAN). http://www.ieee802.org/15/pub/TG6.html
7. Shu, M., Yuan, D., Zhang, C., Wang, Y., Chen, C.: A MAC protocol for medical monitoring applications of wireless body area networks. Sensors **15**(6), 12906–12931 (2015)

A New Method to Generate Semantic Templates Based on Multilayer Perceptron

Ya-Li Qi, Guo-Shan Zhang and Ye-Li Li

Abstract Content-based image retrieval pays more attention to reducing semantic gap. Semantic template is a promising method for reducing semantic gap, and consists of mapping between high-level and low-level visual features. The work presented here proposes a semantic template method via multilayer perceptron, which has three layers: an input layer, a hidden layer, and an output layer. In the proposed method, the pixel features of an interesting region are selected as input features, the features weights are originally designed randomly with random seeds, and softmax is selected as the activation function. Experiments show the proposed method has high accuracy for image retrieval, and the accuracy can be improved by adding samples to train the MLP (Multilayer perceptron) classifier until a relative stable state is achieved.

Keywords Content-based image retrieval (CBIR) · Semantic gap · Multilayer perceptron (MLP) · Semantic template

1 Introduction

In the early 1990s, the content-based image retrieval (CBIR) system was introduced [1]. The largest challenge to effective image retrieval using CBIR is a semantic gap between low-level feature and high-level semantics, which prevents its wide

Y.-L. Qi · G.-S. Zhang
School of Electrical Engineering and Automation, Tianjin University,
Tianjin, China
e-mail: zhanggs@tju.edu.cn

Y.-L. Qi (✉) · Y.-L. Li
The Computer Department, Beijing Institute of Graphic Communication,
Beijing, China
e-mail: qyl@bigc.edu.cn

Y.-L. Li
e-mail: liyl@bigc.edu.cn

application as a solution [2, 3]. In the past few decades, many algorithms have been proposed to reduce semantic gap. These methods can be classified into five categories [2]: object ontology, machine learning, relevance feedback (RF), semantic template (ST), textual information, and visual content, for Web image retrieval. In real image retrieval systems, a synthetic approach based on the five methods can be used in CBIR.

ST is not yet as widely used as the other techniques, but it is a promising approach in CBIR [1]. ST constructs a bridge between semantic features and visual features to reduce semantic gap. It is usually defined as a 'representative' concept, and calculated from a set of sample images [4, 5].

Smith and Li use composite region templates (CRTs) to decode image semantics [4], which defines the region as a semantic description, and describes the prototypal spatial arrangements of the semantic in the images. Chang et al. use semantic visual templates (SVT) to map low-level visual features to high-level semantics for video retrieval, which depends on the user's feedback [6]. Zhang et al. generate ST automatically in the process of RF [7]. In addition, wordnet is also used to construct a network of ST [8]. In the retrieval process based on the ST, once the user proposes a query concept, the system can find a corresponding semantic template. It uses the centroid feature and the corresponding weight to find similar images. Based on this method, Zhang et al. propose an algorithm to generate weak semantics [9]. Sarwar et al. propose a retrieval system using a corpus of natural scene images via imparting human cognition [10]. Hu et al. directly use Gradient Field HOG (GF-HOG) and Bag of Visual Words (BoVW) to define semantic template [11]. In addition in recent years, some scholars use annotation methods to extract the image semantic and generate semantic templates [12–14]. These annotation methods usually retrieve images according to the similarity distance of the semantic templates.

Here we propose an algorithm to generate semantic templates based on multi-layer perceptron (MLP). The algorithm uses the pixel features of interesting regions as input data and generates a random weight vector with random seed for the hidden layer of MLP. To achieve highly accurate classification, it selects softmax as an activation function for output layer of MLP.

2   MLP

MLP is an artificial neural network model based on feedforward, which maps input data to a series of appropriate outputs [15]. It consists of multiple layers with multiple nodes, and each layer is fully connects to the next one (see Fig. 1). MLP uses backpropagation to train the network, which is a supervised learning technique. It can solve the non-linear problems which cannot be solved by single-layer perceptron.

Fig. 1 Multilayer perceptron
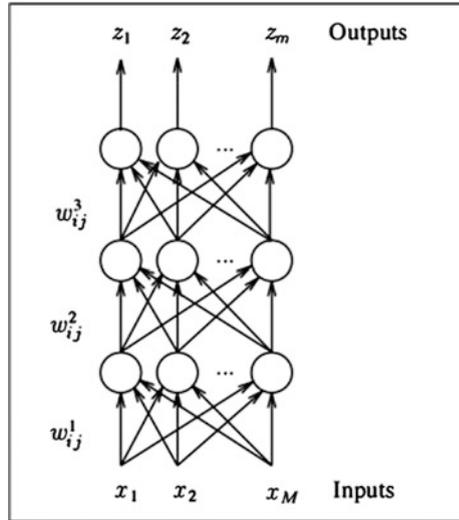
2.1 Layers Structure

The MLP usually consists of an input layer and an output layer. The output layer usually has one or more hidden layers, which include some nonlinearly-activating nodes. Each node in the front layer connects with a weight W_{ij} to every node in the following layer (Fig. 1).

2.2 Activation Function

Each neural node (except input nodes) has an activation function. There are three types of activation functions: linear, logistic and softmax. Linear activation functions are only copied the data by treating node, which is generally used for regression problems, and not suited for classification problems. Logistic activation functions are usually used for multiple classification problems, which generate multiple independent logical attributes as outputs. Softmax activation functions are usually used for common classification problems. The output of this type function is mutually exclusive classes [15].

2.3 Learning Through Back-Propagation

MLP is a supervised learning model based on back-propagation. After a piece of data is processed, the perceptrons change the connection weights by comparing the

difference between the error in the output and the expected result. The output layer weights are modified via the activation function derivation, and the hidden layer weights are changed according to the activation function back-propagation.

3   Experiment

3.1   Generating Semantic Template Based on MLP

To train the MLP classifier with three layers, we selected an interesting region to represent the semantic template features, and used the pixel feature of the interesting region to train MLP classifier. The corresponding output is the semantic perception.

- S1: For the input layer, the pixel feature is the color feature of the pixel, and the pixel feature is used to describe the interesting region as an object. As such, the feature vectors are color feature vectors of pixels in the region of interest. The feature vectors need to be normalized. To do so, the initial feature vectors need substract the mean of the training vectors, and divide the standard deviation of the training vectors components. Therefore, the processed feature vectors have a mean of 0 and a standard deviation of 1.
- S2: For the hidden layer, the weights are initialized randomly. To ensure the reproducible results, a random seed may be passed to the random number generator. When the training results have a relatively large error, the model may select a different value for random seed and retrain the MLP to achieve a smaller error.
- S3: For the output layer, the activation function selects softmax.
- S4: In the retrieval process, we segment the image by a watershed algorithm, which segments the objects form background in image following reference [16]. Then input sub region into MLP classifier. If there is relevant semantic perception in the image, the MLP classifier will classify this image to the relative semantic class as retrieval result.

3.2   Experimental Results

All experiments in presented here were performed on a Lenovo workstation with Intel Core2.5 GHz CPU and Halcon11. The experimental images were selected from http://image.baidu.com/.

In our experiments, we generated two semantic templates: one is 'orange', the other is 'lemon'. We selected a region of interest for each object as inputs for the semantic template classifier (see Fig. 2). Then train the MLP classifier for semantics: orange and lemon respectively.

We used the pixel feature of the region of interest to train MLP classifier. Then we normalized the feature vectors and designated 42 as the random seed, which is based on experience [15]. The region rounded by the orange line represents the orange semantic, and the region rounded by the yellow line represents the lemon semantic. There are two types error of the result: one is the orange image is classified to lemon semantic, or the lemon image is classified to orange semantic; the other is orange image or lemon image is classified into two semantic classes. The results of the classification by MPL are shown in Fig. 3.

For this experiment, we selected two samples to train MLP. One is orange, the other is lemon. The classifying data are 100 images, and the number of correct results is 66. To improve the accuracy of this algorithm, we added the samples to train semantic template classifier. One experiment selected four samples to train MLP, two orange, and two lemons. The other experiment selected six samples to train MLP-three orange samples and three lemon samples. We have performed the experiments with a larger number of samples. We selected eight, ten and twenty samples to train the algorithm. As Table 1 shows, the accuracy of the classifier can be improved via adding samples to train the MLP classifier until achieving a relative stable state, and the accuracy of classification is not significantly improved by increasing the size of training set.
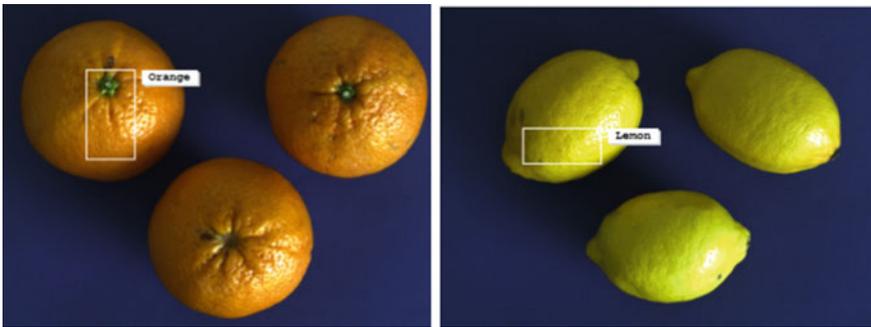
Fig. 2 Select interesting region of object as input for orange and lemon semantic
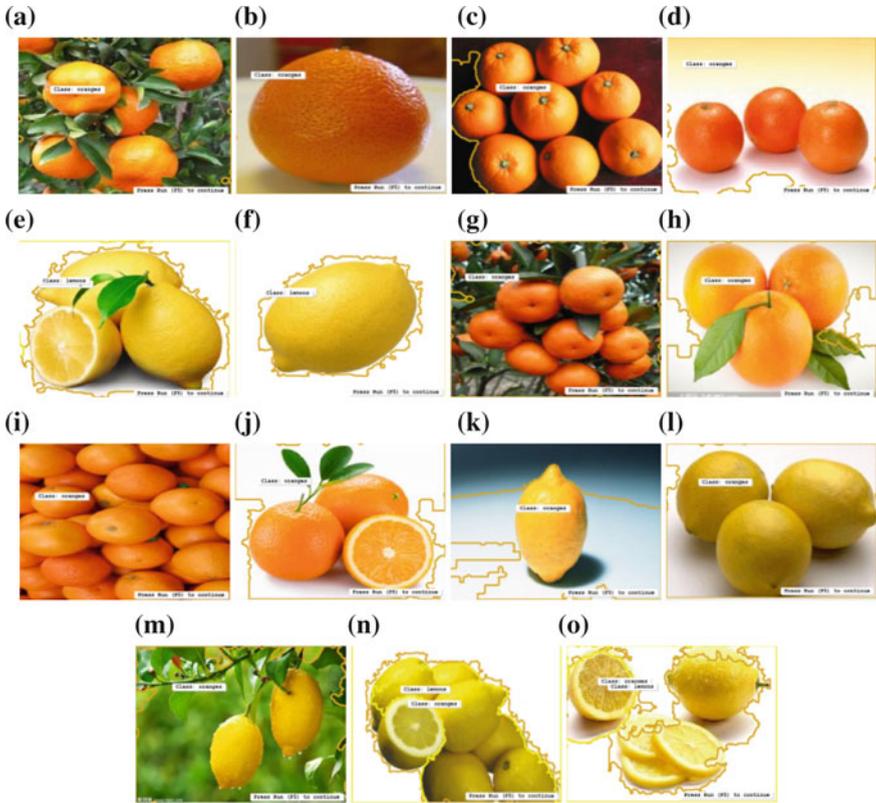
Fig. 3 The results of the classification based on the semantic template classifier. *1* The sub figures (**a–j**) show the correct classifying results; *2* The sub figures (**k–m**) show the error classifying results; *3* The sub figures (**n**, **o**) show the error classifying into two class result

Table 1 The results of the classify based on the semantic template classifier

The number of the samples to train MLP		2 (%)	4 (%)	6 (%)
The accuracy of the classifying		66	74	83
Error of classifying	Error classify to one class	26	20	15
	Classify to two classes	8	6	2

4 Conclusion

In this work we propose a semantic template method based on MLP to reduce the semantic gap for CBIR. The proposed method uses pixel features of a chosen region of interest, uses random weights and softmax to improve the classifier efficiency. Our experimental results show our method has a good efficiency to classify images

based on semantic. In addition, experiments shown that the accuracy of the classifier can be improved via adding samples to train the MLP classifier until attaining a relative stable state, and to achieve the stable state only need small amount of training samples. The proposed method can be applied into classification problem with two classes.

References

1. Eakins, J.P., Graham, M.E.: Content-based image retrieval: a report to the JISC technology applications programme. Technical report. Institute for Image Data Research, University of Northumbria, Newcastle (1999)
2. Liu, Y., Zhang, D., Lu, G., Ma, W.: A Survey of content-based image retrieval with high-level semantics. Pattern Recogn. **40**, 262–282 (2007)
3. Li, Z., Shi, Z., Zhao, W., Li, Z., Tang, Z.: Learning semantic concepts from image database with hybrid generative/discriminative approach. Eng. Appl. Artif. Intell. **26**, 2143–2152 (2013)
4. Smith, J.R., Li, C.S.: Decoding image semantics using composite region templates. In: IEEE workshop on content-based access of image and video libraries (CBAIVL-98), pp. 9–13, IEEE, Piscataway (1998)
5. Denman, S., Halstead, M., Fookes, C., Sridharan, S.: Searching for people using semantic soft biometric descriptions. Pattern Recogn. Lett. **68**, 306–315 (2015)
6. Chang, S.F., Chen, W., Sundaram, H.: Semantic visual templates: linking visual features to semantics. In: International Conference on Image Processing (ICIP), Workshop on Content Based Video Search and Retrieval. Chicago, Illinois. 10, 531–534 (1998)
7. Zhang, Y., Liu, X., Pan, Y.: Apply semantic template to support content-based image retrieval. In: Proceeding of the SPIE Storage and Retrieval for Media Database. San Jose, CA, USA. 3972(12), 442–449 (2000)
8. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: Introduction to wordnet: an on-line lexical database. Int. J. Lexicography. **3**, 235–244 (1990)
9. Zhang, C., Liu, J., Tian, Q., Liang, C., Huang, Q.M.: Beyond visual features: a weak semantic image representation using exemplar classifiers for classification. Neurocomputing. **120**, 318–324 (2013)
10. Sarwar, S., Qayyum, Z.U., Majeed, S.: Ontology based image retrieval framework using qualitative semantic image descriptions. Procedia Comput. Sci. **22**, 285–294 (2013)
11. Hu, R., Collomosse, J.: A performance evaluation of gradient field HOG descriptor for sketch based image retrieval. Comput. Vis. Image Underst. **117**, 790–806 (2013)
12. Zarchi, M.S., Monadjemi, A., Jamshidi, K.: A concept-based model for image retrieval systems. Comput. Electr. Eng. **46**, 303–313 (2015)
13. Kurtz, C., Beaulieu, C.F., Napel, S., Rubin, D.L.: A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations. J. Biomed. Inform. **49**, 227–244 (2014)
14. Ortiz, A., Gorriz, J.M., Ramirez, J., Salas-Gonzalez, D.: Improving MR brain image segmentation using self-organising mapsand entropy-gradient clustering. Inf. Sci. **262**, 117–136 (2014)
15. Wikipedia. http://en.wikipedia.org/wiki/Multilayer_perception
16. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Mach. Intell. **13**, 583–598 (1991)

A Novel Thermal Protection Circuit Based on Bandgap Voltage Reference

Wei Ding, Yong Xu, Rui Min, Zheng Sun and Yuan-Liang Wu

Abstract A novel thermal protection circuit based on a bandgap voltage reference is presented in this paper. Simulation was carried out using Cadence Spectre, based on a 0.25 μm CMOS (Complementary Metal-Oxide-Semiconductor Transistor) process, which indicated that the thermal protection temperature threshold is approximately 130 °C. It was also found that the hyteresis is nearly 20 °C in all types of process corners, and the designed bandgap reference voltage is 1.205 V with a temperature coefficient of 12.84 ppm/°C.

Keywords Thermal protection · Hysteresis temperature · Bandgap voltage reference · CMOS

1 Introduction

With the development of high performance and high speed integrated circuits, the thermal problem in chips has attracted increasing attention [1]. The chip, particularly the power-integrated circuit with large consumption, can be permanently damaged when the internal temperature exceeds the permitted temperature [2, 3].

W. Ding (✉) · Y. Xu · R. Min · Z. Sun · Y.-L. Wu
Institute of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China
e-mail: janedin@sina.com

Y. Xu
e-mail: xu_yong99@163.com

R. Min
e-mail: minrui_he@163.com

Z. Sun
e-mail: sunduoduolele@163.com

Y.-L. Wu
e-mail: Wuyuanliang_2006@163.com

Fig. 1 Block diagram of the
thermal protection circuit
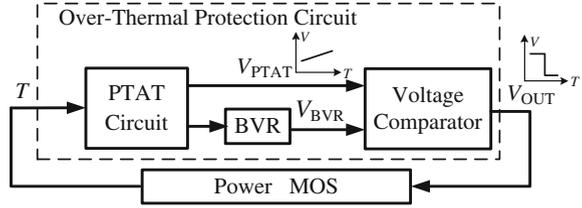
In order to ensure the reliability and lifetime of the circuit, a thermal protection circuit should be integrated into the chip [4, 5]. In this paper, a novel thermal circuit is proposed which consists primarily of PTAT (Proportional To Absolute Temperature) circuit, BVR (Bandgap Voltage Reference) and voltage comparator; the block diagram is shown within the dotted box in Fig. 1. The PTAT circuit beside the power MOSFET (Metal-Oxide-Semiconductor Field Effect Transistor) detects the internal temperature of the chip and transfers it to V_{PTAT}, which is proportional to the absolute temperature. When the temperature exceeds the acceptable temperature, the V_{OUT} would shut off the power MOS via the comparison of V_{PTAT} and V_{BVR}.

This work proposes a CMOS thermal protection circuit based on bandgap reference sources. The simulation results of the proposed circuit in a 0.25 μm CMOS process indicates that the circuit has the characteristics of simple structure, long-term stability, low power consumption and strong portability.

2   Circuit Design

The thermal protection circuit consists of four parts, including the PTAT & BVR circuit, temperature judgment circuit, bias circuit and startup circuit. The implemented structure is presented in Fig. 2.

2.1   PTAT & BVR Circuit

As the core of the circuit, the PTAT & BVR circuit consists of M_{11}–M_{17}, R_0–R_4, Q_1–Q_2, and op amp A_1. In order to increase the stability of BVR, the circuit adopts a cascade current mirror [6]. Furthermore, the output of A1 is set as the bias voltage of the casecode to introduce negative feedback while M_{11} and M_{12} have the same ratio of width and length, namely, $S_{11} = S_{12}$. Consequently, the current I_0 is given as [7]:
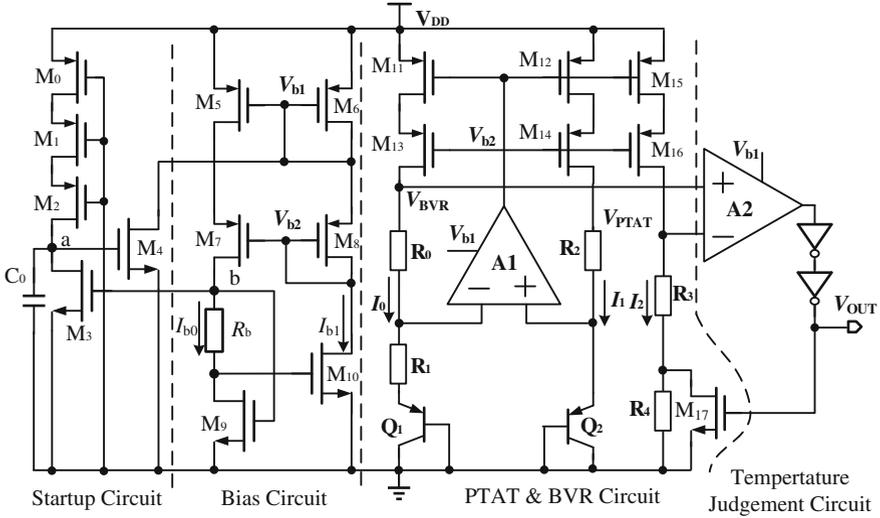
Fig. 2 Schematic of proposed thermal protection circuit

$$I_0 = I_1 = \frac{\ln n}{R_1} V_T \tag{1}$$

where V_T is the thermal voltage, n is the ratio between the emitter areas of the Q_2 and Q_1 bipolar transistors, $V_T = (k/q)T$, k is Boltzmann's constant, T is temperature and q is the conducted charge. V_T with the temperature coefficient +0.085 mV/°C at room temperature is proportional to the absolute temperature T.

It is clear that I_0 is proportional to the temperature, so that I_0, I_1, I_2 are all positive temperature coefficients. The PTAT voltage is expressed as:

$$V_{PTAT} = \ln n \frac{S_{15}}{S_{11}} \frac{R_3}{R_1} V_T \tag{2}$$

Equation (2) implies that V_{PTAT} is entirely independent of the power supply and process parameters. The ratio of resistances is not sensitive to temperature changes when using the same type of resistance. Therefore, V_{PTAT} has good linearity by ignoring the effects of resistance temperature characteristic.

V_{BVR} is the terminal voltage of R_0, and can be expressed as:

$$V_{BVR} = \ln n \frac{R_0}{R_1} V_T + V_{EB2} \tag{3}$$

At room temperature, V_{EB2} has a negative temperature coefficient of approximately –2.2 mV/°C. Therefore, V_{BVR} has little dependence on the power supply and

process parameters by adjusting the ratio of resistor R_0 and R_1. V_{BVR} realizes zero temperature coefficient, namely, the BVR [8].

2.2 Temperature Judgment Circuit

The temperature judgment circuit consists of a comparator A2 and two inverters. A2 adopts the classical two-stage CMOS op amp using M_{18}–M_{24} is shown in Fig. 3. The first stage consists of differential amplifier PMOS transistors; the second stage consists of a common-source MOSFET, and M18, M23 provide bias currents.

The integrated op amp A1 has the same structure. A miller compensation is adopted by connecting capacitor C1 between the outputs of two stages in order to achieve adequate phase margin (>60°) to ensure A1 working steadily [9].

At room temperature, V_{PTAT} is less than V_{BVR}. While the temperature rises, V_{PTAT} increases linearly until $V_{PTAT} = V_{BVR}$. And the comparator flips at the temperature, namely thermal temperature protection T_+ [10]. From this relationship, T_+ can be expressed as:

$$T_+ = \frac{q}{k} \frac{V_{BVR}}{\ln n} \frac{S_{11}}{S_{15}} \frac{R_1}{R_3} \tag{4}$$

It is shown that T_+ is determined by the ratio of R_1 and R_3. Simultaneously, the temperature characteristic of V_{BVR} also determines the stability of T_+.

V_{OUT} is the output of the comparator shaped by two-stage inverter. Once the temperature exceeds T_+, V_{OUT} jumps to low level which can shut off the power MOS and M_{17}. Meanwhile, V_{PTAT} is increased by the voltage drop of R_4. Once V_{PTAT} is equal to V_{BVR}, V_{OUT} turns to high level and opens power MOS and M_{17} at the threshold temperature named T_-. The hysteresis temperature ΔT can be solved as follows:

Fig. 3 Comparator circuit
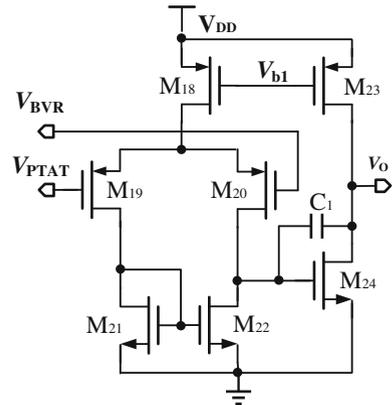
$$\Delta T = T_+ - T_- = \frac{1}{R_3/R_4 + 1} T_+ \tag{5}$$

It is clear that ΔT is determined by R_4 at certain T_+. The feedback circuit composed of M_{17} and R_4 introduces a hysteresis comparison that the risk of thermal oscillation phenomena can be avoided at the temperature T_+ [11].

2.3 Bias Circuit and Startup Circuit

As shown in Fig. 2, the bias circuit supplies bias current Ib0 and the bias voltages Vb1 and Vb2 for op amp A1, A2, PTAT and BVR circuit.

Once the bias circuit is unable to provide necessary voltages and currents, the entire circuit will fail to work normally [12]. So the startup circuit which consists of M_0–M_4 and C_0 is necessary. After power is on, C_0 will be charged until point a exceeds the threshold voltage of M_4. After M_4 is on, V_{b1} decreases and point b increases. Then, M_3 is on, point a is much less than the threshold voltage of M_4, and M_4 is reliably disconnected between the startup circuit and the bias circuit. Thus far, the entire startup process is complete and the bias circuit is in normal balance.

3   Simulation Results and Discussion

The proposed circuit is implemented in a CSMC 0.25 μm 2P5M process. As shown in Fig. 4, the layout covers an area of 0.27×0.29 mm^2. V_{BVR} and V_{out} can be scanned in five parameters: tt, ss, ff, sf and fs corns from −40 to 150 °C. The results of Cadence Spectre simulation is shown in Figs. 5 and 6.

Figure 5 shows that V_{BVR} is highly overlapping at all corners. This indicates that V_{BVR} is little influenced by the corners, in favor of the stability of the temperature of thermal protection. It is observed that the center value of V_{BVR} is 1.205 V with a temperature coefficient of 12.84 ppm/°C.

V_{OUT} is scanned in double direction in the range of −40 to 150 °C, the simulation results in Fig. 6 indicate that the protection temperature is 130 °C, hysteresis temperature is 20 °C, and the proposed circuit exhibits good temperature sensitivity. The maximum temperature error is 1.4 °C in different process corners, and the hysteresis temperature approximation is 20 °C in the same corner. This validates the stability and reliability of the proposed circuit.

Furthermore, the same performance was recorded from other simulation results. Table 1 lists the performance comparison with other works. It is shown that the proposed circuit in this paper has several advantages, such as the largest temperature range and power supply range, smallest circuit area, and good temperature coefficient.
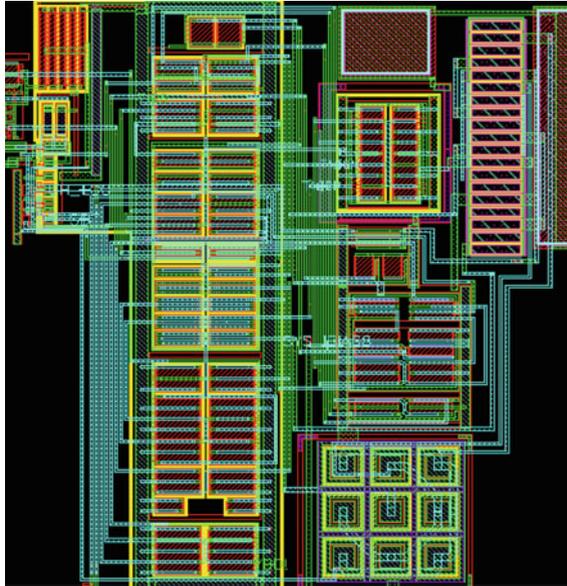
Fig. 4 Layout of the circuit
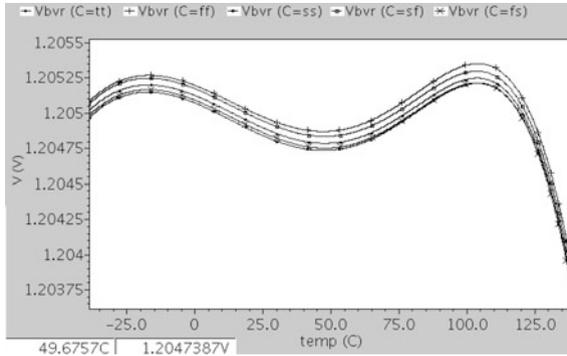
Fig. 5 Simulated temperature characteristic of V_{BVR}
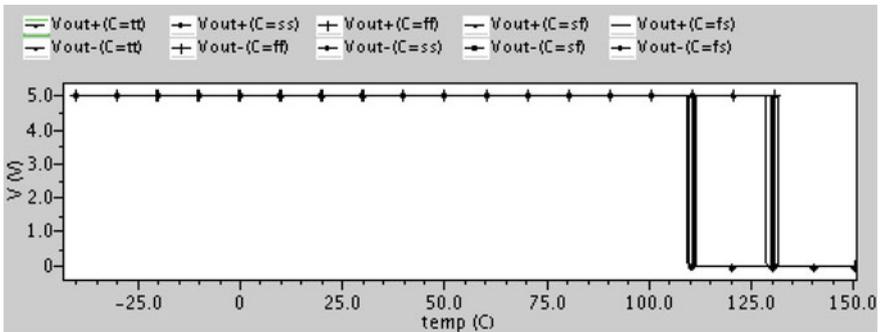
Fig. 6 Simulated temperature characteristic of the output

Table 1 Performance comparison with other works

	Current work	Dong et al. [13]	Wang et al. [14]
Process (µm)	0.25	0.6	0.18
Scale of temperature (°C)	−40 to 150	30–130	0–120
Power supply (V)	1.6–10	4.5–5.5	0.7–3.6
V_{REF} (V)	1.205	2.75	0.4306
Temperature coefficient (ppm/°C)	12.84	86	2.97
Area (mm^2)	0.27 × 0.29	1.9695 × 1.9683 include pad	0.02

4   Conclusions

A novel thermal protection circuit based on CMOS BVR is proposed. Cadence Spectre simulation results show that the thermal protection temperature is approximately 130 °C, the temperature error is less than 1.4 °C, the hysteresis temperature is nearly 20 °C, and the bandgap reference voltage is 1.205 V with a temperature coefficient 12.84 ppm/°C from −40 to 150 °C. Long-term stability, high sensitivity, small size, and high portability have been achieved by the proposed circuit which can be widely used in different application fields.

References

1. Hsia, S.C., Sheu, M.H., Jhou, J.J., Lai, S.Y.: Embedded temperature sensor for multilevel current LED driver. IEEE Sens. J. **14**(8), 2801–2806 (2014)
2. Wang, Z.J., Shi, L.Z., Cheng, K.: A LDO with thermal and over-current protection. Microelectronics **45**(3), 312–314 (2015)
3. Ahmad, K., Chandrajit, T., Gary, F., Jeyanandh, P., Yoed, R.: Characterization of a CMOS sensing core for ultra-miniature wireless implantable temperature sensors with application to cryomedicine. Med. Eng. Phys. **36**(9), 1191–1196 (2014)
4. Syed, A.J., Waqar, A.Q., Atia, S., Junaid, A.Q., Abdul, H., Moaaz, H.: Low-power area-efficient wide-range robust CMOS temperature sensors. Microelectron. J. **44**(2), 119–127 (2013)
5. Aiello, O., Fiori, F.: On the susceptibility of embedded thermal shutdown circuit to radio frequency interference. IEEE Trans. Electromagn. Compat. **54**(2), 405–412 (2012)
6. Xu, Y., Zhao, F., Xu, Y.B., Guan, Y.: Design of bandgap voltage reference with novel self-startup circuits. Res. Prog. SSE. **29**(4), 566–568 (2009)
7. Razavi, B.: Design of Analog CMOS Integrated Circuits, pp. 314–315. Xi'an Jiaotong University Press, Xi'an (2003)
8. Meijer, G.C.M., Wang, G.J., Fruett, F.: Temperature sensors and voltage reference implemented in CMOS technology. IEEE Sens. J. **1**(3), 225–234 (2001)
9. Philip, E.A., Douglas, R.H.: CMOS Analog Circuit Design, 3rd edn., pp. 255–256. Oxford University Press, New York (2011)

10. Chang, S.W., TSOU, C.F.: A novel Silicon-based LED packaging module with an integrated temperature sensor. IEEE Trans. Compon. Packag. Manuf. Technol. **4**(5), 769–776 (2014)
11. Wang, X.Q., Li, L.N., Ou, W., Zhai, P.F., Mao, S.B.: Design of a LDO with enable circuit and protection circuits. Microelectronics. **45**(4), 449–453 (2015)
12. Pang, J., Wu, J., Yan, L.H.: High reliability start-up circuits design. Chin. J. Electron Devices **28**(2), 390–393 (2005)
13. Dong, X.Y., Sun, Y., Zhu, D.Z.: Design of a thermal-shutdown circuit using 0.6 μm CMOS technology. Res. Prog. SSE. **28**(2), 281–283 (2008)
14. Wang, H.J., Wang, C.H., He, H.Z., Li, Z.: A low power voltage reference source based on subthreshold MOSFETs. Microelectronics **41**(5), 654–657 (2011)

A Real-Time 106 Mbit/S Visible Light Communication System Design Using General Lighting LED with Analog Pre-emphasis

Xiu-Qin Yang, Min Zhang, Da-Hai Han, Peng-Fei Luo, Zabih Ghassemlooy and Qing-Yuan Huo

Abstract This paper presents a real-time 106 Mbit/s visible light communication (VLC) system using only non-return-to-zero on-off keying (OOK-NRZ) modulation and general lighting LED. With a pre-emphasis module based LED driver, the 3-dB modulation bandwidth of VLC link is extended from 4 to 106 MHz. Given the BER upper Limit was 1.3×10^{-3}, the highest bit rates achieved through the designed system were 106 and 98 Mbit/s at the distances of 60 and 70 cm, respectively.

Keywords Pre-emphasis · White LED · OOK-NRZ · VLC

1 Introduction

The light emitting diodes (LED) have remarkably been developed in recent years, which are smaller in size, highly efficiency, with high output power and longer life-time. Therefore, LEDs are rapidly replacing incandescent and fluorescent lights

X.-Q. Yang (✉) · M. Zhang · D.-H. Han · Q.-Y. Huo
State Key Laboratory of Information Photonics and Optical Communications,
Beijing University of Posts and Telecommunication, Beijing, China
e-mail: yangxq19900102@163.com

M. Zhang
e-mail: mzhang@bupt.edu.cn

D.-H. Han
e-mail: dahaihan@gmail.com

Q.-Y. Huo
e-mail: 846137167@qq.com

P.-F. Luo · Z. Ghassemlooy
Optical Communications Research Group, Faculty of Engineering and Environment,
Northumbria University, Newcastle-upon-Tyne, UK
e-mail: olive.luo@northumbria.ac.uk

Z. Ghassemlooy
e-mail: z.ghassemlooy@northumbria.ac.uk

in many applications. In visible light communication (VLC) systems, LEDs have unique dual functionalities of illumination as well as data communications, thus making VLC system one of the promising technologies for wireless communications [1].

LEDs used in VLC systems have large parasitic capacitance as a result of a large area of the p-n junction. The parasitic capacitance and the output resistance of the driver circuit limit the 3-dB modulation bandwidth of LED from several MHz to several tens of MHz [2]. Phosphor-based blue LEDs and single color (red, green, blue) LEDs have 3-dB modulation bandwidth of ~ 2 and ~ 40 MHz, respectively [3]. For higher data rate applications, the bandwidth of LEDs needs to be extended. Pre-emphasis and post-emphasis techniques are possible options to extend 3-dB mod-ulation bandwidth [4]. In the research done by Fujimoto and Mochizuki, 477 Mbit/s VLC based on non-return-to-zero on-off keying (OOK-NRZ) and using a pre-emphasis circuit over a transmission span of 40 cm was reported. A 460 Mbit/s VLC OOK-NRZ link employing analog modulator with pre-emphasis was demonstrated over a distance of 100 cm [6].

In this paper, we demonstrate 106 Mbit/s VLC link based on the OOK-NRZ format with a simple pre-emphasis circuit. Using a commercially available white LED and a high sensitivity PIN-photodetector (PD), the measured BER is at 1.3×10^{-3} over a transmission span of 60 cm, which is well within the pre-forward error correction (FEC) limit of 3.8×10^{-3}. The designed pre-emphasis circuit is low-cost since it is simply composed resistors, capacitors and transistors.

Compared with the research done by Fujimoto and Mochizuki [5], a RGB-type white LED and post amplifier is not used in this paper, and the transmission distance is 60 cm, longer than 40 cm proposed by Fujimoto and Mochizuki [5]. Compared with what Li et al. [6] has done, the driver circuit in this paper is simple, and OOK-NRZ is directly generated by BERT 1400 GENERATOR.

2 Proposed Pre-emphasis Circuit in VLC System

2.1 The Pre-emphasis Circuit Diagram

In general, 3-dB modulation bandwidth of LED is limited to several MHz due to parasitic capacitance and the output resistance of the driver circuit. In order to compensate for the attenuation at high frequencies, we included a pre-emphasis circuit to the LED driver, as shown in Fig. 1. The pre-emphasis circuit is composed of two of stages, with each stage compensating certain high frequency bands. Firstly, we chose a proper value for resistors to set the static operating points of the transistors accord-ing to their characteristics. The compensation point and magnitude response of the 1st stage are determined by R_3, R_4, R_5, C_1. The alternating current voltage of the first stage could be expressed by
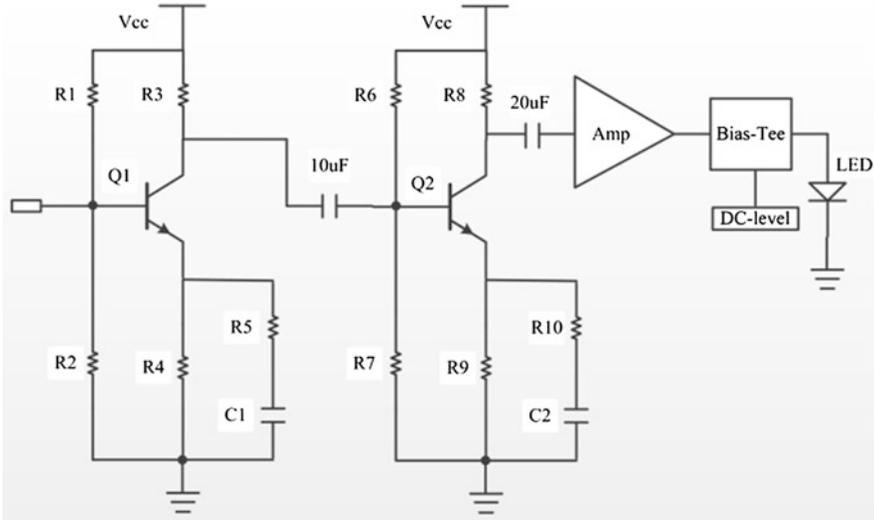
Fig. 1 The LED driver with a pre-emphasis circuit

$$A_1(j\omega) = \frac{R_3}{R_4//(R_5 + \frac{1}{j\omega C_1})}. \tag{1}$$

The magnitude response of the first stage amplifier is

$$|A_1(j\omega)| = \frac{R_3}{R_4}\left(1 + \frac{\omega R_4 C_1}{\sqrt{1 + \omega^2 R_5^2 C_1^2}}\right). \tag{2}$$

The drive capability of the 1st stage is relevant to its output impedance, especially the value of R_3, which is suitable for the 2nd stage circuit when R_3 is 10 Ω. The value of resistances of the 2nd stage is similar to the 1st stage. In addition, $R_8 = 49.9$ Ω, and $C_2 < C_1$.

2.2 Frequency Response of the VLC System with Pre-emphasis Circuit

The compensation frequency bands of two amplifying stages are about 40 and 80 MHz due to the value of resistances and capacitances of the pre-emphasis circuit. Finally, we extend 3-dB modulation bandwidth of the VLC system to about 106 MHz as shown in Fig. 2. Note that Fig. 2 presents the measured frequency

response of Electric/Optic/Electric (E/O/E) channel with/without the pre-emphasis circuit.

3 Experimental Setup of the VLC System

Figure 3 illustrates the block diagram and experimental setup of the proposed VLC system. In the transmitter, a pseudo random binary sequence (PRBS) generated from a signal generator (Tektronix, giga BERT 1400 GENERATOR) is emphasized by the proposed pre-emphasis circuit, and amplified (EA). The amplified signal is DC level shifted and used via a Bias-Tee for intensity modulation of a commercially available white-light LED (NICHIA, NSPW500BS).

At the receiver, a commercially available PIN-PD (Thorlabs PIN) combined with a blue-filter and a convex lens (50-mm-diameter) is used for collecting the light signal. Focusing the light onto the PD does lead to enhanced signal-to-noise ratio (SNR) performance of the link. The received optical signal is converted into an electrical signal and amplified using trans-impedance amplifier (TIA), the output of which is captured using the error detector (Tektronix, giga BERT 1400 ANALYZER) and a real-time digital oscilloscope (RIGOL, MSO4024) to observed the received signal (eye diagram).
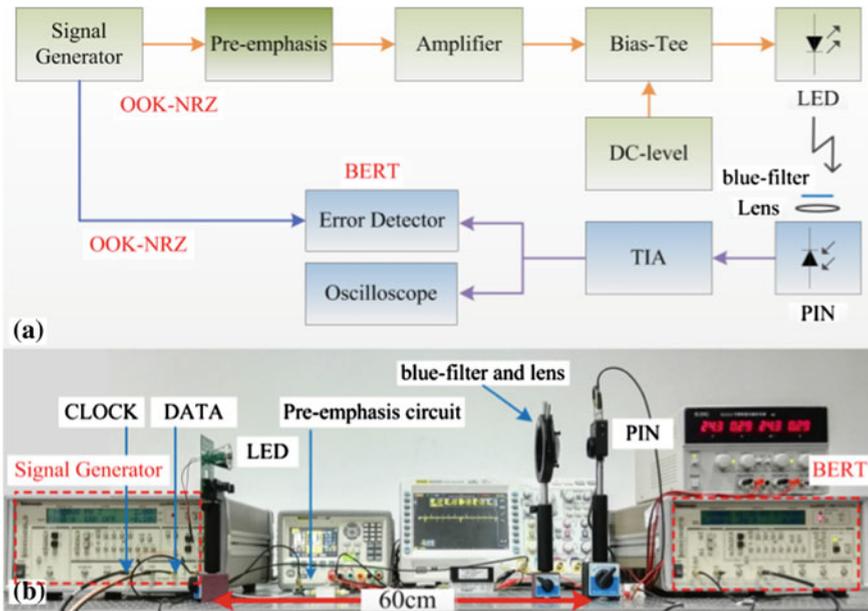
Fig. 3 **a** The block diagram, and **b** experimental setup of the VLC system

4 Results and Discussions

The eye diagram and BER are the performance criteria that are used to assess the quality of signal received at the receiver, from which the performance of the entire VLC system can be assessed. A real-time oscilloscope is used to capture the eye dia-gram for 106 Mbit/s data rate as shown in Fig. 4. The real-time BER of the received signal is measured using the error detector as we mentioned above. Figure 4 shows the BER as a function of the data rate for two transmission spans of 60 and 70 cm. The BER is less than 10^{-9} when the data rate is lower 90 Mbit/s, above which, the BER in-creases rapidly with the data rate. BER of 1.3×10^{-3} is observed at the data rate of 106 Mbit/s for the transmission distance of 60 cm, which is still below the FEC limit of 3.8×10^{-3}.

The BER is higher ($\sim 10^{-1}$) for the transmission distance of 70 cm for the same data rate of 106 Mbit/s. This is as expected since the SNR decreases with the transmission distance. Therefore, the maximal data rates achievable with the pro-posed system are 106 and 98 Mbit/s for the distances of 60 and 70 cm, respectively.

In order to improve the 3-dB bandwidth and bit rate of the VLC system, we could add a post-emphasis circuit to the receiver. The post-emphasis circuit should be designed according to the frequency response of transmitter. And we plan to design a MIMO VLC system based on the pre-emphasis circuit using a RGB-type white LED in the next step of the work.

Fig. 4 BER of the VLC system at the distance of 60 and 70 cm, and the eye diagram at the data rate is 106 Mbit/s
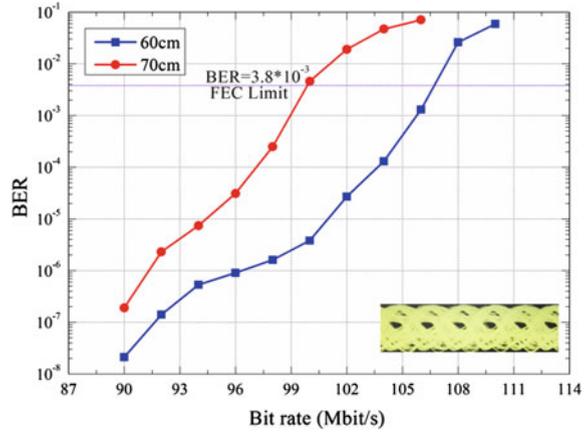
5　Conclusion

In this paper, we experimentally demonstrated a high speed VLC system employing OOK-NRZ and a simple pre-emphasis circuit. Using a commercially available white LED and a high sensitivity PIN-PD, the measured BER was 1.3×10^{-3} over a transmis-sion span of 60 cm, which is below the recommended FEC limit of 3.8×10^{-3}. We also showed that the same BER can be achieved the transmission span of 70 cm, but at a reduced data rate of 98 Mbit/s. The experimental results proved that a simple pre-emphasis circuit is effective for achieving a high-speed and low-cost VLC system.

References

1. Chien, L., Yuan, F., Chong, L., Meng, C., Chien, Y., Mau, P., Yuan, T., Cheng, F.: Light-emitting diodes for visible light communication. In: Wireless Communications and Mobile Computing Conference, pp. 665–667. IEEE Conference Publications (2015)
2. Chi, N., Haas, H., Kavehrad, M., Little, T.D.C., Huang, X.L.: Visible light communications: demand factors, benefits and opportunities. Wireless Commun. IEEE **22**, 5–7 (2015)
3. Zhang, M.L.: An experiment demonstration of a LED driver based on a 2nd order per-emphasis circuit for visible light communications. In: Wireless and Optical Communication Conference, pp. 1–3. IEEE, Piscataway (2014)
4. Karunatilak, D., Zafar, F., Kalavally, V., Parthiban, R.: LED based indoor visible light communications: state of the art. Commun. Surv. Tutorials IEEE **17**, 1649–1678 (2015)

5. Fujimoto, N., Mochizuki, H.: 477 Mbit/s visible light transmission based on OOK-NRZ modulation using a single commercially available visible LED and a practical LED driver with a pre-emphasis circuit. In: Optical Fiber Communication Conference, pp. 1–3. OSA Publishing, Washington D.C. (2013)
6. Li, H., Chen, X., Guo, J., Gao, Z., Chen, H.: An analog modulator for 460 Mb/s visible light data transmission based on OOK-NRS modulation. Wireless Commun. IEEE **22**, 68–73 (2015)

GNSS Receiver Anti-spoofing Techniques: A Review and Future Prospects

Ling Xiao, Peng-Cheng Ma, Xiao-Mei Tang and Guang-Fu Sun

Abstract Spoofing interference can mislead a target receiver to report a wrong position and time. This can pose a serious threat to the security of global navigation satellite system (GNSS) applications, and may cause undesirable consequences. As such, anti-spoofing techniques have become a hot research topic within the GNSS discipline. This paper provides a review of recent research in the field of GNSS anti-spoofing on the receiver side. The vulnerability of GNSS receivers to spoofing attacks is studied, and the anti-spoofing algorithms around the base band digital signal processing layer and the information processing layer of the receiver is discussed. The limitation, cost and applicability of these anti-spoofing methods are investigated and the trend of anti-spoofing research in the future is analyzed.

Keywords GNSS receiver · Spoofing interference · Anti-Spoofing technique

1 Introduction

With the development of GNSS, the position, navigation and time (PNT) services, provided by GNSS, have a large influence in our daily life. Nowadays, various applications such as aircraft navigation and landing systems, electrical power distribution grids, digital communication networks, stock exchange transactions,

L. Xiao (✉) · P.-C. Ma · X.-M. Tang · G.-F. Sun
Satellite Navigation R&D Center, School of Electronic Science and Engineering,
National University of Defense Technology, Changsha 410073, China
e-mail: xiaoling_nudt@163.com

P.-C. Ma
e-mail: mapengcheng1001@163.com

X.-M. Tang
e-mail: txm_nudt@hotmail.com

G.-F. Sun
e-mail: sunguangfu_nnc@163.com

police and rescue services and many more are relying on GNSS signals. With the increased use of GNSS, the security of these services is becoming more and more important. However, as the signals become extremely weak when they reach the earth, they are vulnerable to interference. In addition, because the working frequency band, the modulation type, the civilian pseudo-random noise (PRN) codes and data information are public, GNSS signals can be easily faked.

These counterfeit signals are termed spoofing interference. Among all the types of interference, spoofing is most harmful, because it can fool the target receiver into reporting wrong position or time results without perception, which may lead to serious consequences, for example, leading an unmanned aerial vehicle (UAV) off course [1], blocking digital communication networks [2], creating power grid equipment failure [3] and so on.

Therefore, there are many anti-spoofing techniques that have been proposed in recent years. This paper first investigates the vulnerability of GNSS receivers to spoofing attacks around the signal processing and information processing layers. Then, a brief summary of current anti-spoofing techniques in the above two layers will be provided. Finally, the trend of future research within this topic will be analyzed.

This paper is organized as follows: GNSS vulnerability against spoofing attacks is studied in Sect. 2. Anti-spoofing techniques will be discussed in Sect. 3. In Sect. 4, the study trends of anti-spoofing methods will be analyzed. Finally, the conclusion is drawn in Sect. 5.

2 GNSS Vulnerability Against Spoofing Attacks

As shown in Fig. 1, a GNSS receiver mainly has three functional modules: the radio frequency (RF) front end module, the base band signal processing module and the navigation generating module, which is also termed the information processing
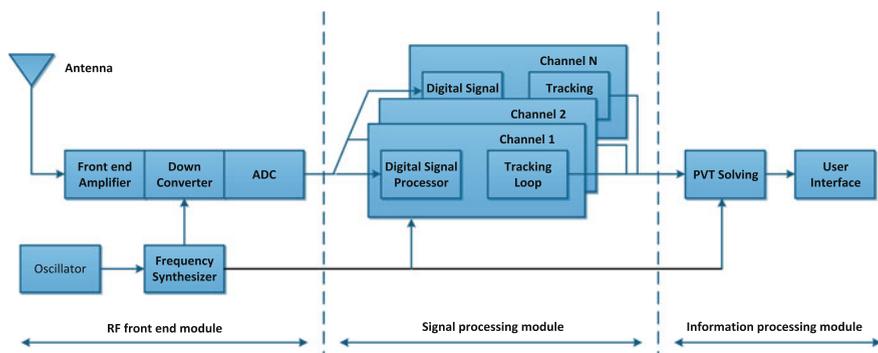
Fig. 1 The three function modules of a classical GNSS receiver

module. The tasks of the RF front end module are signal amplifying, frequency down-conversion and signal filtering; the incident signal almost hasn't been changed in this module. Thus, this module is vulnerable to all kinds of interferences that fall in its processing band. Spoofing signals are aimed to attack the last two modules and control the receiver to report false position or time. We will investigate the receiver vulnerability to spoofing at the two modules in the following.

2.1 GNSS Receiver Vulnerability in Signal Processing Module

The main tasks of the signal processing module are signal acquiring and tracking. In the signal acquiring phase, the spoofer can transmit counterfeit signals that are much more powerful than authentic ones (as shown in Fig. 2), which can cause the receiver to acquire the counterfeit signal. In the signal tracking phase, a more covert spoofing attack can take place, which transmits a counterfeit signal that slowly approaches the authentic one, and then drags the tracking loop away (as shown in Fig. 3). Once the receiver is working on the fake signals, the receiver is controlled by the spoofer.
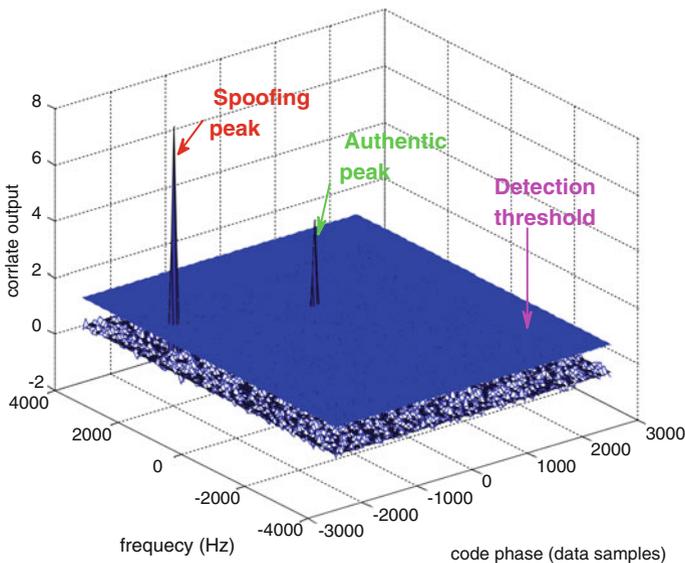
Fig. 2 The scenario of spoofing attack during signal acquiring phase
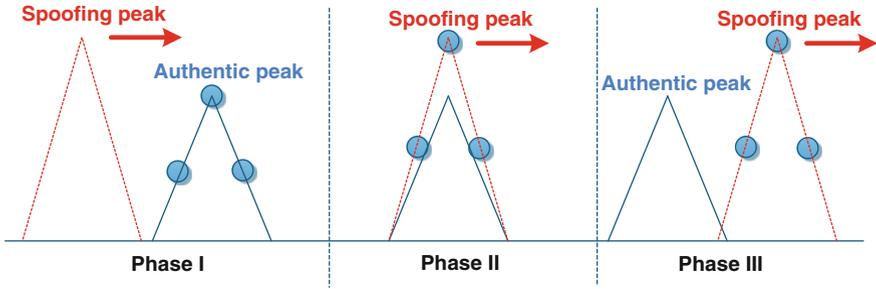
Fig. 3 The scenario of a spoofing attack during the signal tracking phase (the *three dots* denote signal tracking points)

2.2 GNSS Receiver Vulnerability in Information Processing Module

In the information processing module, the information is extracted from the data messages, and the PNT are solved using the measurement quantities provided by the signal processing module. As the framing structure of the data message is publicly known and the information does not change rapidly during some time intervals, the data message can be easily faked, which makes the receiver trust the faked message casually. During the PVT (Position, Velocity and Time) solving phase, the receiver autonomous integrity monitoring (RAIM) procedure can detect abnormal events based on range residuals. However, when the receiver is fully controlled, the range residuals are too small to trig alarms. Also, a well-designed spoofer can change the PVT results gradually and make the receiver not notice the danger.

3 GNSS Receiver Anti-spoofing Techniques

Anti-spoofing techniques can be classified into two major categories: the GNSS side and the receiver side. The GNSS side anti-spoofing techniques always need modifications of the GNSS structure, which can't be implemented promptly. This paper will discuss receiver side anti-spoofing techniques. In the following, anti-spoofing methods that take place in the signal processing module and information processing module are discussed respectively.

3.1 Anti-spoofing Methods in the Signal Processing Layer

In-band Power Monitoring The existence of spoofing signals will increase the in-band power, which will change the receiver's auto gain control (AGC) level. The

spoofing interference can be alarmed by monitoring the abnormal variance of the AGC gain level [4]. This method needs the information of AGC gain, so when the receiver only deals with digital intermediate frequency signals, the method can't be implemented. To make up for this limitation, Jafarnia-Jahromi et al. [5] have proposed a pre-despreading authenticity verification method. The delay and multiply (DAM) property of Gold codes is used in this method to generate a new Gold code that carries all the incident signals' power. Then, the in-band power component is filtered by a comb filter. The filter output is used to detect spoofing interference. This algorithm can sense the spoofing signal effectively, but it can't discriminate between spoofing interference and spectrum matched interference.

CNR (Carrier Noise Ratio) Monitoring Most GNSS receivers employ CNR measurements as a parameter that characterizes the received signal quality. Under normal conditions, the received signal power changes smoothly with the satellite movement and surroundings change. However, when a higher power spoofing signal controls the receiver tracking loop, the received CNR may experience a sudden change that can indicate the presence of spoofing interference [6, 7]. Wen et al. [8] shows that when the distance between the spoofer antenna and the receiver changes from 8 to 100 m, the received CNR reduces by 22 dB. Thus for a moving receiver, if its CNR measurements change considerably, there may be spoofing interference.

Multi-antenna Methods Montgomery et al. [9] have proposed a spoofing detection technique that compares the calculated phase difference of two fixed GNSS antennas to the theoretical one. This technique requires a calibrated antenna array, and it takes about one hour to do the detection. Borio [10] designed a double antenna receiver and developed a phase only analysis of variance (PANOVA) method in order to detect the phase difference coherency of spoofed PRN signals. This method can effectively recognize spoofing signals when the SNR (signal noise ratio) is larger than 10 dB, otherwise the detection performance is poor. Psiaki et al. [11] have proposed a method using a dual-antenna differential carrier phase. This method detects spoofing based on the fact that the quantities of authentic signals' carrier-phase single-differences are multiplicity, while the spoofing ones are identical.

Synthetic Array Methods Nielsen et al. [12] has proposed a spoofing detection algorithm that employs the synthetic antenna array technique. This algorithm detects spoofing signals by computing the correlation coefficient of the channel gain. The satellite signals arrive by passing different transmitting channels, so the channel gains are uncorrelated. However, as all the spoofing signals pass through the same channel, the channel gains for these signals are identical. This method works effectively even in multipath environments because all the spoofing signals experience the same fading path. The drawback is that it is only applicable to moving receivers.

Signal Quality Monitoring (SQM) Methods SQM techniques are widely used to monitor GNSS correlation peak quality in multipath fading environments. The

signal in the process of a spoofing attack on a receiver tracking loop is similar to the multipath component. Thus, the SQM techniques have been extended to detect spoofing attacks [13–15]. The ratio and delta SQM tests are employed to detect any abnormal asymmetry or flatness of GNSS correlation peaks. These techniques can only be used in line-of-sight propagation environments to detect spoof interference. In multipath environments, SQM methods might not be able to distinguish the spoofing signals or multipath reflections.

Code and Phase Rates Consistency Check For authentic signals, the Doppler frequency and the code rate are consistent, as they are both affected by the relative movement between GNSS satellite and the receiver. The relationship of these parameters is $f^a = -f_{RF}\dot{\tau}^a$, where f^a and $\dot{\tau}^a$ denote the Doppler frequency and code rate respectively, and f_{RF} is the radio frequency of the GNSS signal. Thus, this relationship can be used to detect spoofing [8]. This method is simple to implement. However, the spoofer can keep this relationship easily.

3.2 Anti-spoofing Methods in the Information Processing Layer

Received Navigation Data Check

Ephemeris Consistency Check. The ephemeris information, including eccentricity, orbital inclination, rate of right ascen and so on, will not change for about 2 h. Thus, we can compare the current received ephemeris with the save ones. If there are many differences, there may be a spoofing attack.

Satellites Clock Consistency Check. The data messages of every signal contain all the satellite's clock information. The information coming from different signals should be the same. Any abnormality may indicate a spoofing attack.

PNT Solution Check

Receiver Clock Variance Check. In normal cases, the receiver clock bias changes smoothly, which depends on the quality of the used crystal oscillator. However, in the spoofed case, when the receiver moves with respect to the spoofer antenna, the clock bias will change rapidly [16]. This is because all the spoofing signals experience a common delay from the spoofer to the receiver. In the PVT solving process, the common delay is reflected on the clock bias.

Multi-Receiver Position Consistency Check. Literatures [17–20] all proposed a multi-receiver system that detects spoofing by checking the position reported by the receivers. If the system is spoofed, all receivers will obtain the same position result. In order to detect spoofing successfully, it requires the distance between receivers to be at least as large as twice the position solution, and all the receivers to be spoofed.

Consistency Check with other Navigation System. Before the GNSS bearing, land radio navigation systems have been widely used, such as the Roland system

and tactical air navigation system. Therefore, whether or not the receiver is attacked by spoofing can be checked by comparing the GNSS solution with another navigation system's solution [21].

Consistency Check with Inertia Measurement Unit (IMU). Stand-alone inertia equipment can independently provide many high solution navigation parameters, such as position, velocity and attitude. These parameters can be used to detect spoofing by comparing with GNSS ones [22, 23].

3.3 Summary

The requirement, complexity, valid scope and performance of the above-discussed anti-spoofing methods are tabulated in Table 1.

The three performance levels are defined as: (1) alarming means that the method can't discriminate spoofing interference or other type interference; (2) detecting means that the method can recognize spoofing, but can't mitigate it; (3) suppressing means that the method can detect and mitigate spoofing.

Table 1 Summary of GNSS receiver anti-spoofing methods

Anti-spoofing methods	Required capability	Complexity	Valid scope	Performance
AGC gain monitoring	AGC output	Low	Confined	Alarming
Pre-despreading method	None	Low	Generally	Alarming
CNR monitoring	CNR measuring	Low	Generally	Alarming
Direction of arrival monitoring	Antenna array	High	Confined	Suppressing
PANOVA method	Dual antenna	High	Confined	Detecting
Synthetic array method	Receiver moving	Low	Confined	Suppressing
SQM method	None	Low	Generally	Detecting
Code and phase consistency check	None	Low	Generally	Detecting
Received ephemeris consistency check	None	Low	Generally	Alarming
Satellites clock consistency check	None	Low	Generally	Alarming
Receiver clock variance check	None	Low	Generally	Alarming
Multi-receiver position consistency check	Multi-receiver	Medium	Confined	Detecting
Consistency check with other navigation system	Multi-navigation system processing ability	High	Confined	Detecting
Consistency check with IMU	IMU equipment	High	Confined	Detecting

4 Prospect of Future Research

According to the above discussions of anti-spoofing techniques, the current research findings are mainly focusing on alarming or detecting the spoofing interference, and some findings have applicability limitations. For example, some require extra equipment, and some are only effective in special scenarios. Therefore, techniques that can be generally used, and can mitigate or eliminate the interference rather than only detecting it, are required. We think the future researches of this scope will be expanded in the following aspects:

1. That research will occur on different anti-spoofing techniques fusion strategies. A stand-alone method may have limitations, while methods combining together can extend the sphere of application. For example, the power monitoring method combines the SQM method and can detect not only high power spoofing but also covert spoofing attacks, and the applicability is not only confined to line-of-sight scenarios. Thus, how to fuse anti-spoofing methods will be a trend to be researched.
2. That research will occur on multi-GNSS anti-spoofing techniques. With the development of GPS, GLONASS, Galileo and Compass, many receivers have the ability to deal with multi-GNSS signals, which can help to detect spoofing signals. Spoofing interferences' detection and suppression can be realized by comparing and checking the characters of multi-signals (e.g., signal power) and processing results (e.g., the state of clock errors).
3. That anti-spoofing technique research by combining exterior assistants will occur. GNSS receivers are generally used on mobile phones, cars, airplanes, and steamships, on which there are other facilities to provide location, velocity and attitude information. How to use these messages to enhance the safety of the receivers' services should be researched.
4. That research on interference source localization techniques will occur. Techniques, localizing and further destroying the interference source are the most effective methods to protect GNSS receivers. The CNR, pseudo-range and Doppler measurements from different receivers are candidates for source localization.

5 Conclusion

With the wide use of GNSS services all over the world, their security and robustness become more and more important. This paper summarizes the current anti-spoofing techniques around the signal processing layer and information processing layer. As discussed in Sect. 3, the methods, such as in-band power monitoring, CNR monitoring, PNT check and so on, that have low complexity can be used generally. However, most of these methods can't tell whether there is a threat

or just a receiver failure. The multi-antenna technique can detect and mitigate spoofing threats effectively, but it needs extra equipment and space. In conclusion, low-cost and universal applicable GNSS receiver anti-spoofing techniques will be a research point.

References

1. UAVs Vulnerable to Civil GPS Spoofing. http://gpsworld.com
2. Recommended Minimum Performance Standards for CDMA 2000 Spread Spectrum Base Stations. C.S0010-C. Technical report, 3rd Generation Partnership Project 2 "3GPP2". http://www.docin.com/p-560878501.html
3. Zhang, Z.H., Gong, S.P., Dimitrovski, A.D., Li, H.S.: Time synchronization attack in smart grid: impact and analysis. IEEE Trans. Smart Grid. **4**(1), 87–98 (2013)
4. Akos, D.M.: Who's afraid of the spoofer? GPS/GNSS spoofing detection via automatic gain control (AGC). J. Navig. **59**(4), 281–290 (2012)
5. Jafarnia-Jahromi, A., Broumandan, A., Nielsen, J., Lachapelle, G.: Pre-despreading authenticity verification for GPS L1 C/A signals. J. Inst. Navig. **61**(1), 1–11 (2014)
6. Dehghanian, V., Nielsen, J., Lachapelle, G.: GNSS spoofing detection based on receiver C/N0 estimates. In: Proceedings of the 25th International Technical Meeting of The Satellite Division of the Institute of Navigation, pp. 2878–2884. The Institute of Navigation, Manassas, USA (2012)
7. Nielsen, J., Dehghanian, V., Lachapelle, G.: Effectiveness of GNSS spoofing countermeasure based on receiver CNR measurements. Int. J. Navig. Obs. **2012**, 1–9 (2012)
8. Wen, H., Huang, P.Y., Dyer, J., Archinal, A., Fagan, J.: Countermeasures for GPS signal spoofing. In: Proceedings of the 18th International Technical Meeting of The Satellite Division of the Institute of Navigation, pp. 1285–1290. The Institute of Navigation, Manassas, USA (2005)
9. Montgomery, P.Y., Humphreys, T.E., Ledvina, B.M.: Receiver-autonomous spoofing detection: experimental results of a multi-antenna receiver defense against a portable civil GPS spoofer. In: Proceedings of ION International Technical Meeting, pp. 124–130. The Institute of Navigation, Manassas, USA (2009)
10. Borio, D.: PANOVA tests and their application to GNSS spoofing detection. IEEE Trans. Aerosp. Electron. Syst. **49**(1), 381–394 (2013)
11. Psiaki, M.L., O'Hanlon, B.W., Powell, S.P.: GNSS spoofing detection using two-antenna differential carrier phase. In: Proceedings ION GNSS+2014, pp. 2776–2800. The Institute of Navigation, Manassas, USA (2014)
12. Nielsen, J., Broumandan, A., Lachapelle, G.: GNSS spoofing detection for single antenna handheld receivers. Navigation **58**(4), 335–344 (2011)
13. Cavaleri, A., Motella, B., Pini, M., Fantino, M.: Detection of spoofed GPS signals at code and carrier tracking level. In: Proceedings of the 5th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC '10), pp. 1–6. IEEE, Piscataway (2010)
14. Cavaleri, A., Pini, M., Presti, L.L., Fantino, M.: Signal quality monitoring applied to spoofing detection. In: Proceedings of the 24th International Technical Meeting of The Satellite Division of the Institute of Navigation, pp. 1888–1896. The Institute of Navigation, Manassas (2011)
15. Kuusniemi, H., Bhuiyan, M.Z.H., Kröger, T.: Signal quality indicators and reliability testing for spoof-resistant GNSS receivers. Eur. J. Navig. **11**(2), 12–19 (2013)

16. Jafarnia, A., Daneshmand, S., Broumandan, A., Nielsen, J., Lachapelle, G.: PVT solution authentication based on monitoring the clock state for a moving GNSS receiver. http://plan.geomatics.ucalgary.ca/papers/pvt_authentication_enc2013_alij_27apr2013.pdf
17. Swaszek, P.F., Hartnett, R.J., Kempe, M.V., Johnson, G.W.: Analysis of a simple, multi-receiver GPS spoof detector. In: Proceedings of the 2013 International Technical Meeting of The Institute of Navigation. pp. 884–892. The Institute of Navigation, Manassas (2013)
18. Swaszek, P.F., Hartnett, R.J.: Spoof detection using multiple COTS eceivers in safety critical applications. In: Proceedings of the ION GNSS + 2013, pp. 2921–2930. The Institute of Navigation, Manassas (2013)
19. Swaszek, P.F., Hartnett, R.J.: A multiple COTS receiver GNSS spoof detector—extensions. In: Proceedings of the 2014 International Technical Meeting of The Institute of Navigation, pp. 316–326. The Institute of Navigation, Manassas (2014)
20. Heng, L., Makela, J.J., Dominguez-Garcia, A.D., Bobba, R.B., et al.: Reliable GPS-based timing for power systems: a multi-layered multi-receiver architecture. Power Energy Conf Ill (PECI) **2014**, 1–7 (2014)
21. Zhu, X.W., Wu, Y.W., Gong, H., et al.: Timing receiver toughen technique in complicated jamming environments. J. Nat. Univ. Defense Technol. 37(3), 1–10 (2015) (in Chinese)
22. Khanafseh, S., Roshan, N., Langel, S., Chan, F.C., Joerger, M., Pervan, B.: GPS spoofing detection using RAIM with INS coupling. In: Proceedings of IEEE/ION PLANS 2014, pp. 1232–1239. The Institute of Navigation, Manassas (2014)
23. Tanil, C., Khanafseh, S., Pervan, B.: Impact of wind gusts on detectability of GPS spoofing attacks using RAIM with INS coupling. In: Proceedings of the ION 2015 Pacific PNT Meeting, pp. 674–686. The Institute of Navigation, Manassas (2015)

Adaptive Energy-Efficient Design of Cooperative Spectrum Sensing in Cognitive Radio Networks

Jing-Wei Liang, Jian-Xin Dai, Yi-Chen Liu, Xing-Zhou Zhou and Man Xu

Abstract Cognitive radio is a promising technology which can be used to solve the shortage of spectrum resource. As sensing nodes of the network, secondary users are usually battery powered. Making full use of the energy should be considered by balancing the tradeoff of energy and throughput. In this paper, we consider the overall throughput and energy consumption of the system. We propose a model for throughput and energy consumption with an adaptive factor to discuss the influences of sensing time and number of secondary users, and analyze the existence of an optimal point that get the maximum value of energy efficiency. Simulation results show that energy efficiency always have the optimal points in different fusion rules. Our findings indicate that cognitive radio network design, based on energy efficiency consideration, is feasible.

Keywords Cognitive radio · Energy efficiency · Adaptive factor

J.-W. Liang (✉) · Y.-C. Liu
College of Telecommunications and Information Engineering,
Nanjing University of Posts and Telecommunications, Nanjing 210003, China
e-mail: liang.jingwei@foxmail.com

Y.-C. Liu
e-mail: b13010314@njupt.edu.cn

J.-X. Dai · M. Xu
School of Science, Nanjing University of Posts and Telecommunications,
Nanjing 210023, China
e-mail: daijx@njupt.edu.cn

M. Xu
e-mail: xman_njupt@163.com

X.-Z. Zhou
Bell Honors School, Nanjing University of Posts and Telecommunications,
Nanjing 210023, China
e-mail: q14010122@njupt.edu.cn

1 Introduction

Cognitive radio (CR) is a potential solution to the scarcity of spectrum resource. Current inflexible spectrum allocation policy causes spectrum shortage problem and low utilization ratio [1].

In CR network, the spectrum bands are allocated to primary users (PU). Secondary users (SU) should sense the radio environment and adaptively choose the transmission parameters according to sensing result to avoid the interference to Pus [2]. It is a fundamental issue in CR networks that SUs should be able to efficiently and effectively detect the presence of Pus [3, 4]. The PU is the owner of the channel and SUs rent the spectrum when it is idle. To improve the opportunity of SUs to access the spectrum channel, we can use multiple SUs that work corporately to sense a single channel.

The basic tradeoff in the cooperative spectrum sensing method is as follows: if more SUs are assigned to sense one channel, higher sensing performance can be achieved, however, the total transmission power of the signal measurement and the overhead traffic in the secondary networks grows approximately linearly with the number of cooperating SUs [5]. It is an important issue to balance the two aspects, throughput and energy to achieve the maximum usage of spectrum and energy resource. Considering that SUs are always battery-powered, energy efficiency is important for CR networks.

Deng divided the sensors into a number of non-disjoint feasible subset such to extend the lifetime of sensors [6]. Hu et al. [7] focused on the optimization of the final decision threshold to maximize the energy efficiency for different signal channels. Nardelli et al. [8] provided a throughput analysis based on location information. Monemian and Mahdavi [9] supposed a new energy-based sensor selection algorithm was proposed to provide approximately the same lifetimes for sensors via the appropriate design of cooperative spectrum sensing.

In this paper, we provide an optimization model to discuss the system performance in with respect to energy efficiency and introduce an adaptive factor to make the model more flexible. We analyze the model and give the simulations of different fusion rules.

2 System Model

Since spectrum and energy are both precious resources that should be considered. We propose a model of throughput and energy consumption with adaptive factor to discuss the influences of sensing time and number of secondary users.
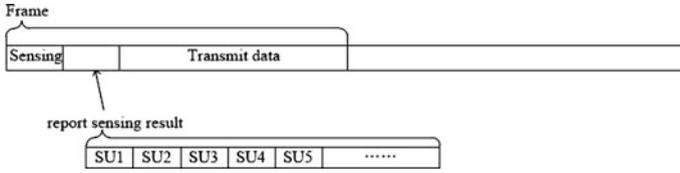
Fig. 1 Frame structure of cooperation spectrum sensing

2.1 Frame Structure of Cooperative Spectrum Sensing

Figure 1 shows the frame structure of cooperative spectrum sensing [1]. It can be divided into three different parts: sensing, reporting and transmitting slots. SUs periodically sense the spectrum for an ideal channel. In cognitive radio network (CRN), all SUs sense the spectrum. The primary channel sensing result of SUs will then be transmitted to the base station. The base station analyzes the data and determines if the spectrum is ideal.

2.2 Energy Detection

Each sensor performs spectrum sensing independently. The sensor will compare the collected energy E_i with a predefined threshold ε to determine if the channel is busy.

The false alarm probability and detection probability of the sensor are defined as:

$$P_f = \Pr\{D_i = 1|H_0\} = \Pr\{E_i > \varepsilon_i|H_0\} = Q\left(\frac{\varepsilon - f_s\tau}{\sqrt{2f_s\tau}}\right), \tag{1}$$

$$P_d = \Pr\{D_i = 1|H_1\} = \Pr\{E_i > \varepsilon_i|H_1\} = Q\left(\frac{\varepsilon - f_s\tau - \gamma}{\sqrt{2f_s\tau + 4\gamma}}\right), \tag{2}$$

where $\gamma = \sigma_x^2/\sigma_n^2$ is the received signal-to-noise ratio, H_1 indicates when the channel is busy and H_0 indicates when the channel is ideal.

2.3 Cooperative Spectrum Sensing

Multiple sensors can be coordinated to perform cooperative spectrum sensing to solve the hidden terminal problem by fusing the sensing result of all SUs to avoid errors. The sensors forward their decisions to the base station. The base station fuses these decisions to make the final decision. The decision fusion rules at base station are "OR", "AND" and "K/N". The "OR" rule can be stated as:

$$\begin{cases} Q_d = 1 - \prod_{i=1}^{N}(1 - P_{d,i}) \\ Q_f = 1 - \prod_{i=1}^{N}(1 - P_{f,i}) \end{cases}. \tag{3}$$

The "AND" rule can be stated as:

$$\begin{cases} Q_d = \prod_{i=1}^{N} P_{d,i} \\ Q_f = \prod_{i=1}^{N} P_{f,i} \end{cases}. \tag{4}$$

The "K/N" rule can be stated as:

$$\begin{cases} Q_d = \Pr\{D_i = 1|H_1\} = \Pr\left\{\sum_{i=1}^{N} D_i \geq k|H_1\right\} \\ Q_d = \Pr\{D_i = 1|H_1\} = \Pr\left\{\sum_{i=1}^{N} D_i \geq k|H_1\right\} \end{cases}. \tag{5}$$

2.4 Model Description

The cooperative radio network has four states. A SU can transmit data when it is p $(H_0|H_0)$ or $p(H_0|H_1)$. Since data transmission fails at $p(H_0|H_1)$, the normalized throughput of SU is:

$$R_S(L, \tau) = C_S\left(\frac{T - \tau - LT_R}{T}\right)\left(1 - Q_f(L, \tau)\right)p(H_0), \tag{6}$$

where C_S is the channel capacity, L represents the number of cooperative spectrum sensing users, τ is sensing time, T is the length of time frame, T_R is the result reporting time of one SU, $Q_f(L, \tau)$ is false positive probability. We suppose that data can only be transmitted successfully at $p(H_1|H_1)$, the normalized throughput of SU is:

$$R_P(L, \tau) = C_P Q_d(L, \tau)p(H_1). \tag{7}$$

Power consumption can be shown as four different parts:

1. Spectrum channel is ideal and SU transmits data. $p(H_0|H_0)$
   In this situation, the power cost is $C_{00}(L, \tau) = L\tau P_S + LT_R P_R + LTP_C + (T - \tau - LT_R)P_T + P_{CP}T$.

2. PU does not transmit data and fusion center gives an incorrect result. $p(H_1|H_0)$. The power cost of this situation is $C_{10}(L, \tau) = L\tau P_S + LT_R P_R + LTP_C + P_{CP}T$.
3. PU transmits data in the channel and fusion center does not detect it. $p(H_0|H_1)$ The power cost is $C_{01}(L, \tau) = L\tau P_S + LT_R P_R + LTP_C + (T - \tau - LT_R)P_T + P_{CP}T + P_{TP}T$.
4. PU transmits data and the fusion data gives the right result. $p(H_1|H_1)$ The power cost of this situation is $C_{11}(L, \tau) = L\tau P_S + LT_R P_R + LTP_C + P_{CP}T + P_{TP}T$.

Hence we get the average power cost:

$$P(L, \tau) = p(H_0|H_0)C_{00}(L, \tau) + p(H_1|H_0)C_{10}(L, \tau)$$
$$+ p(H_0|H_1)C_{01}(L, \tau) + p(H_1|H_1)C_{11}(L, \tau). \tag{8}$$

P_s is the power of sensing the spectrum, P_R is the power of reporting result to base station, P_T is the SU's power of transmitting packets, P_C is the power lost in circuit. P_{CP} is the circuit loss of PU, P_{TP} is the transmission power of PU.

The power efficiency can be represented as:

$$EE(L, \tau, \alpha) = \frac{R_P(L, \tau)(1 - \alpha) + R_S(L, \tau)\alpha}{P(L, \tau)} = \frac{R(L, \tau, \alpha)}{P(L, \tau)}. \tag{9}$$

Hence we get the optimization problem:

$$\begin{aligned} &\max_{L,\tau} EE(L, \tau, \alpha) \\ &s.t. Q_d(L, \tau) \geq \bar{Q}_d \\ &\alpha \in [0, 1] \\ &1 \leq L \leq N \\ &L \in N \\ &T - \tau - LT_R \geq 0 \\ &\tau \geq 0 \end{aligned} \tag{10}$$

$\alpha \in [0,1]$ is the adaptive factor.

Theorem If the number of SUs is fixed, analyze the relationship between energy efficiency and length of sensing time. We find the partial derivate of energy efficiency with respect of sensing time, if there exists a point that makes the partial derivate equal to 0, there exists an optimal point that get the maximum value of energy efficiency.

Proof Find the partial derivate of energy efficiency:

$$\frac{\partial EE(L, \tau, \alpha)}{\partial \tau} = \frac{P(L, \tau)\partial R(L, \tau, \alpha) - R(L, \tau, \alpha)\partial P(L, \tau)}{\partial \tau P(L, \tau)^2}. \tag{11}$$

Then we get the partial derivate of throughput and energy:

$$\frac{\partial R(L, \tau, \alpha)}{\partial \tau} = C_1\alpha(1 - Q_f(L, \tau)) + C_2\tau\alpha\frac{\partial Q_f(L, \tau)}{\partial \tau} + C_3(1 - \alpha)\frac{\partial Q_d(L, \tau)}{\partial \tau}, \quad (12)$$

$$\frac{\partial P(L, \tau)}{\partial \tau} = C_4 + C_5\frac{\partial Q_d(L, \tau)}{\partial \tau} + C_6\frac{\partial Q_f(L, \tau)}{\partial \tau}, \quad (13)$$

where C_1, C_2, C_3, C_4, C_5, and C_6 are constants. We use "AND" rule in (4) as an example to find the partial derivate of false alarm probability and detect probability: $\frac{\partial Q_f}{\partial \tau} = L(1 - P_f)^{L-1}\frac{\partial P_f}{\partial \tau}, \frac{\partial Q_d}{\partial \tau} = L(1 - P_d)^{L-1}\frac{\partial P_d}{\partial \tau}$. Since $\tau \in [0, T]$, we find the values of terminal points, $\tau = 0, \frac{\partial P_f}{\partial \tau} \to -\infty, \frac{\partial Q_f}{\partial \tau} \to -\infty$. In this model, we fix the value of detection probability, then we have $\frac{\partial Q_d}{\partial \tau} = 0$. Then we get $\frac{\partial EE(L, \tau)}{\partial \tau} > 0$ at the point $\tau = 0$ if $C_S\alpha P(H_0) > C_P(1 - \alpha)P(H_1)$. By the same method, we can get $\frac{\partial Q_f}{\partial \tau} \to 0$ at the point $\tau = T$. Then we have $\frac{\partial EE(L, \tau)}{\partial \tau} < 0$. Since it is a continuous function, there always exists a sensing time that satisfies the optimal energy efficiency. Similarly, we can also prove that when the length of sensing time is fixed, there are a specific number of SUs that satisfies the optimal energy efficiency.

3 Simulations

In the following simulations, we suppose the base station can confirm the location of every SU. Signals received by each SU have the same SNR. We set $\gamma_i = -20$ dB, $T = 20$ ms, $T_R = 0.1$ ms, $Q_d = 0.9$. We suppose $P_{TP} = 4P_T$.

3.1 Simulations Results of "AND" Rule

Horizontal axis indicates the value of sensing time in Fig. 2a and the number of SUs in Fig. 2b. We use different linetypes to distinguish the value of τ and L, use different symbols to distinguish the value of α. As we can see from the Fig. 2a, α effectively affects the value of energy efficiency. Maximum value is higher when α is 0.8. It is because of the high expense the PU paid while transmitting the same amount of data. As we can see from the Fig. 2b, we can see from the figure that if α is higher, which means consider less about the throughput of PU, the value of energy efficiency is lower when sensing time is 1 ms.

Fig. 2 Simulations results of "AND" rule, **a** energy efficiency versus the sensing time, **b** energy efficiency versus number of SUs
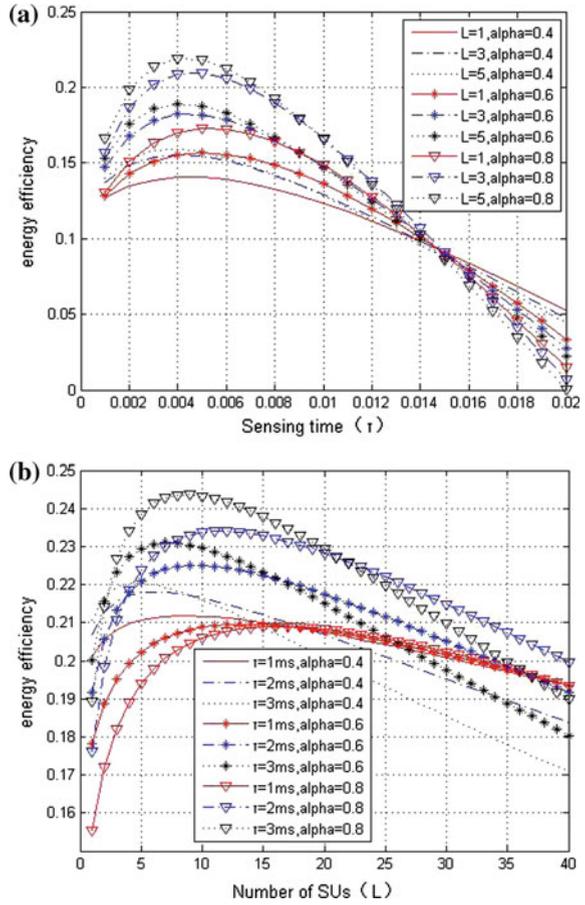
3.2 Simulations Results of "OR" Rule

The "OR" rule in (3) means if all SUs report the same result of spectrum sensing, then the fusion center determines the situation is true. The "OR" rule is always used to protect the access of PU. As such, the SUs have fewer chances to use the channel. As we can see from Fig. 3a, the energy efficiency of the model increase at first and after reaching the optimal point, decrease as the sensing time grows up. Figure 3b shows us as L increases the energy efficiency firstly increase and then decrease. The value of energy efficiency is affected by different parameters. Given three different specific sensing time to observe how the number of SUs influence the energy efficiency.
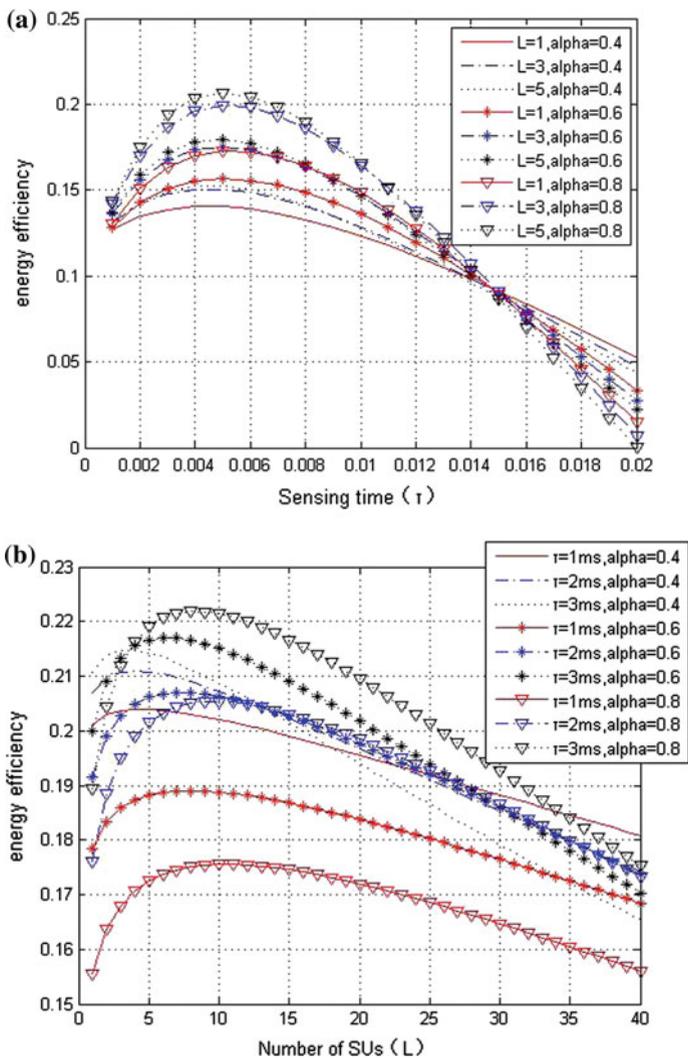
Fig. 3 Simulations results of "OR" rule, **a** energy efficiency versus the sensing time, **b** energy efficiency versus number of SUs

3.3   Simulations Results of "K/N" Rule

If more than half of the SUs report the same result, the fusion center decides the result to be true. We can see from the Fig. 4a that the curves of energy efficiency are also convex. There must have a point for every spectrum sensing to get the best

Fig. 4 Simulations results of "K/N" rule, **a** energy efficiency versus the sensing time, **b** energy efficiency versus number of SUs
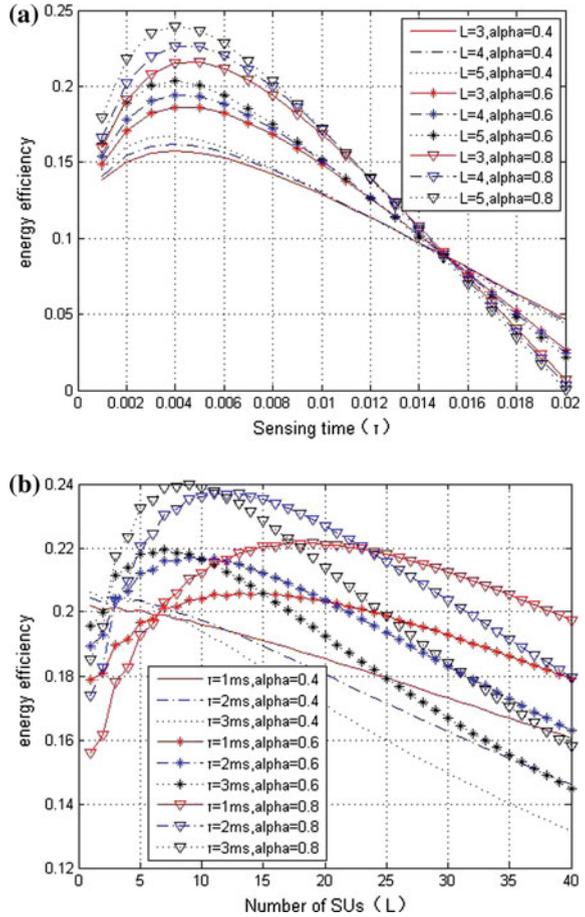
energy efficiency. As we can see from Fig. 4b, the curve is not smooth because of the half value rounded up if the number is not integer. Its influence is very obvious when the number of SUs is small.

4 Conclusion

Multiple SUs working cooperatively in cognitive radio improved the usage rate of spectrum, but spectrum is not the only resource needed to be considered. Energy consumption is also a very important when designing the CR network. We have proposed an energy efficiency model with an adaptive factor to discuss the influence of sensing time and number of SUs. The adaptive factor can be used to make the model fit different situations. The value of the adaptive factor indicates the degree of

PU's channel protection. From this model we can determine the proper values for the sensing time and the number of SUs to make the network worked effectively. By considering both energy and spectrum resources we can find the proper network to make the best profit. In this paper, we consider only one primary user and channel. The reality is complex, we will introduce more variables in our future works.

References

1. Ghasemi, A., Sousa, E.S.: Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs. Commun. Mag. IEEE **46**(4), 32–39 (2008)
2. Zhao, N., Sun, H.: Robust power control for cognitive radio in spectrum underlay networks. Ksii Trans. Int. Inf. Syst. **5**(7), 1214–1229 (2011)
3. Malhotra, M., Aulakh, I.K., Vig, R.: A review on energy based spectrum sensing in Cognitive Radio Networks. In: 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE 2015), pp. 561–565. IEEE, Piscataway (2015)
4. Jiang, C., Zhang, H., Ren, Y., Chen, H.H.: Energy-efficient non-cooperative cognitive radio networks: Micro, meso, and macro views. IEEE Commun. Mag. **52**(7), 14–20 (2014)
5. Hu., H., Xu, Y., Li, N.: Energy-efficient cooperative spectrum sensing with QoS Guarantee in cognitive radio networks. Ieice Trans. Commun. **96**(5), 1222–1225 (2013)
6. Deng, R., Chen, J., Yuen, C., Cheng, P., Sun, Y.: Energy-efficient cooperative spectrum sensing by optimal scheduling in sensor-aided cognitive radio networks. Veh. Technol. IEEE Trans. **61**(2), 716–725 (2012)
7. Hu, H., Zhang, H., Yu, H., Chen, Y., Jafarian, J.: Energy-efficient design of channel sensing in cognitive radio networks. Comput. Electr. Eng. **42**, 207–220 (2014)
8. Nardelli, P.H.J., Lima, C.H.M.D., Alves, H., Cardieri, P., Latva-Aho, M.: Throughput analysis of cognitive wireless networks with Poisson distributed nodes based on location information. Ad Hoc Netw. **33**, 1–15 (2015)
9. Monemian, M., Mahdavi, M.: Analysis of a new energy-based sensor selection method for cooperative spectrum sensing in cognitive radio networks. IEEE Sens. J. **14**(9), 3021–3032 (2014)

An Adaptive Spatial-Temporal Error Concealment Scheme Based on H.264/AVC

Dan Yang, Tao Liu, San-Min Liu and Fu-Chun Chen

Abstract When audio-video is in the process of wireless transmission, it is prone to losing packets. As a solution, this paper presents an adaptive spatial-temporal error concealment scheme based on H.264/AVC. It can adaptively select the spatial error or temporal error concealment on the basis of change in the surroundings, in which temporal error finds the best motion vectors with the spatial-temporal boundary matching, and spatial error concealment switches between the weighted average method and improved directional interpolation. Experimental results indicate good performance of the proposed scheme for error concealment.

Keywords Multi directional interpolation · Error concealment · Boundary matching algorithm

1 Introduction

Single bit error only will lead to packet loss. It can also cause error propagation phenomena in H.264/AVC (audio video coding). In order to solve the problem, H.264/AVC has used error concealment techniques. Error concealment techniques take advantage of the correlation of image sequences and human visual characteristics to restore the damaged macro-block [1]. Error Concealment techniques only take place on the decoder side, which does not affect real-time communication or increase

D. Yang (✉) · T. Liu · S.-M. Liu · F.-C. Chen
Key Laboratory of Computer Application Technology,
Anhui Polytechnic University, Wuhu, Anhui, China
e-mail: 503308723@qq.com

T. Liu
e-mail: 76627110@qq.com

S.-M. Liu
e-mail: 29278883@qq.com

F.-C. Chen
e-mail: cfcjsj@qq.com

the burden of the encoder. Error Concealment techniques comprise spatial error concealment (SEC) and temporal error concealment (TEC) [2]. There are two kinds of frames: I frame and P-frame. To increase efficiency and ease, I-frame exploits SEC, while P-Frame exploits TEC. SEC usually adopts binaries interpolation and direction interpolation, while TEC usually uses boundary matching algorithms [3].

When the scene changes, the correlation of adjacent frames is low. Therefore, methods based on frame types must have two different error concealment algorithms, which compromises the quality of restored data [4]. This paper presents an adaptive spatial-temporal error concealment scheme. It can adaptively select TEC or SEC according to the change of surroundings. In this solution, temporal concealment finds the best motion vectors with the spatial-temporal boundary matching, and spatial concealment makes use of multi directional interpolation.

2 Adaptive Spatial-Temporal Error Concealment Scheme

TEC cannot adapt changes of the scenes in P frame. The proposed method not only restores the effect to drop, but also takes the place of the phenomenon of error propagation. It estimates whether the scenes change, and then chooses different algorithms based on these estimations. If the scene changes, special error concealment is adopted. Otherwise, temporal error concealment is adopted. Figure 1 shows adaptive spatial-temporal error concealment algorithms.

Fig. 1 Proposed adaptive spatial-temporal error concealment scheme
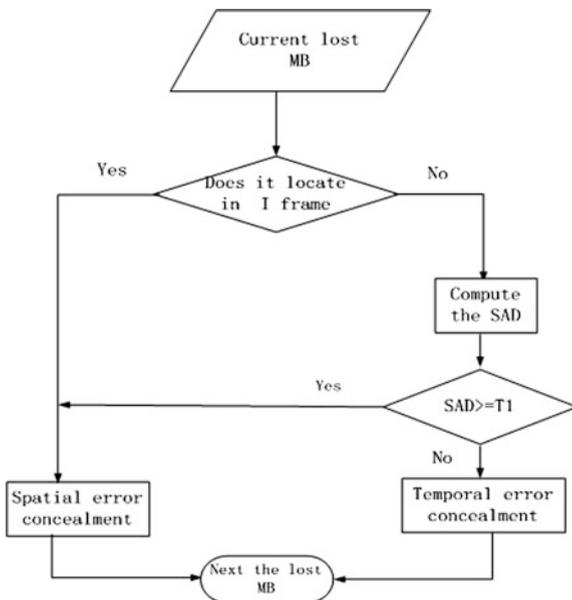
Firstly, the algorithm determines the type of the current frame. Then, it calculates the sum of absolute differences (SAD) for P frame [5], and SEC for I frame. The threshold T_1 is set by the following Eq. (1) to determine whether scenes change:

$$T_1 = N_{MB} \times G_{ave} \tag{1}$$

where N_{MB} is the number of intra-frame coding macro-block (MB) around the lost MB and G_{ave} is the mean gradient of edge of the lost MB. If SAD is larger than the threshold T1, spatial error concealment is chosen for the change of scenes. Otherwise, enhanced temporal error concealment is chosen. The proposed algorithm adaptively chooses TEC or SEC on the basis of the threshold T_1.

3 Enhanced TEC

Motion vector (MV) estimation plays an important role in TEC. TEC estimates missing motion vectors by correlation of continuous frames. With these estimations, the lost MB can be replaced with corresponding MB in a reference frame with motion compensation methods. The best MV is selected from the MV candidates by boundary matching arithmetic (BMA). BMA minimizes boundary matching distortion between the internal and external parts of recovered MB. The minimum distortion function D_1 is given by Eq. (2) as follows [6]:

$$
\begin{aligned}
D_1 = \frac{1}{(\omega_T + \omega_L + \omega_R + \omega_B)N} \Bigg\{ &\omega_T \sum_{x=x_0}^{x_0+N-1} (F(x, y_0 - 1, n) - F(x + \hat{m}_x, y_0 + \hat{m}_y, n - 1))^2 \\
&+ \omega_B \sum_{x=x_0}^{x_0+N-1} (F(x, y_0 + N, n) - F(x + \hat{m}_x, y_0 + N - 1 + \hat{m}_y, n - 1))^2 \\
&+ \omega_L \sum_{y=y_0}^{y_0+N-1} (F(x_0 - 1, y, n) - F(x_0 + \hat{m}_x, y + \hat{m}_y, n - 1))^2 \\
&+ \omega_R \sum_{y=y_0}^{y_0+N-1} (F(x_0 + N, y, n) - F(x_0 + N - 1 + \hat{m}_x, y + \hat{m}_y, n - 1))^2 \Bigg\}
\end{aligned}
\tag{2}
$$

where (x_0, y_0) is the coordinate of the top left corner pixel in the lost MB, $(\hat{m}_x, \hat{m}_y)$ is the candidate motion vector, $F(x, y, n)$ and $F(x, y, n - 1)$ stand for the current frame and reference frame, respectively, N is the MB size (N = 16), and ω_T, ω_R, ω_L, and ω_B respectively are the coefficients of the top, bottom, left, and right MBs adjacent to the lost MB. The coefficients are equal to 1 if the corresponding MB adjacent to the lost MB is available; otherwise, the result is 0. The motion vector that minimizes boundary matching distortion function D_1 is selected as the

replacement motion vector of the lost macro-block (MB). BMA mainly exploits the smoothness property between adjacent pixels. However, when the pixels suddenly change in the edge, D_1 is not suitable to this case and error increases. BMA is only applicable to horizontal and vertical edges of lost macro-blocks, but shows little effect on the other directions. BMA provides us with another matching criterion because there is strong structural similarity between MB adjacent to lost MB in the current frame and MB adjacent to replacement MB in the reference frame. This method minimizes boundary matching distortion between the inner boundaries of replacement MB and the external boundaries of the lost MB.

The minimum distortion function D_2 is given by Eq. (3) as follows:

$$
\begin{aligned}
D_2 = {} & \frac{1}{(\omega_T + \omega_L + \omega_R + \omega_B)N} \\
& \times \left\{ \omega_T \sum_{x=x_0}^{x_0+N-1} (F(x, y_0 - 1, n) - F(x + \hat{m}_x, y_0 - 1 + \hat{m}_y, n - 1))^2 \right. \\
& + \omega_B \sum_{x=x_0}^{x_0+N-1} (F(x, y_0 + N, n) - F(x + \hat{m}_x, y_0 + N + \hat{m}_y, n - 1))^2 \\
& + \omega_L \sum_{y=y_0}^{y_0+N-1} (F(x_0 - 1, y, n) - F(x_0 - 1 + \hat{m}_x, y + \hat{m}_y, n - 1))^2 \\
& \left. + \omega_R \sum_{y=y_0}^{y_0+N-1} (F(x_0 + N, y, n) - F(x_0 + N + \hat{m}_x, y + \hat{m}_y, n - 1))^2 \right\}
\end{aligned} \tag{3}
$$

D_1 can maintain the smoothness property of the adjacent pixels, while D_2 can maintain temporal continuity. This paper proposes the distortion function D, which includes D_1 and D_2. In other words, D not only maintains the smoothness property, but also maintains temporal continuity, as shown in Eq. (4):

$$
D = \alpha D_1 + (1 - \alpha)D_2 \tag{4}
$$

where α is the weight factor, which varies between 0 and 1. The quality of the restored image is good when α value is greater than 0.5.

4   Spatial Error Concealment

4.1   A Weighted Average Method (WAM)

The weighted average method replaces the lost pixels with the weight average of the boundary pixels of the adjacent MB. The weighted value and the distance

between interpolation pixels and reference pixels have a close correlation, which is inversely proportional. This relationship is shown in Eq. (5):

$$P_{ij} = \frac{d_1 \times P_1 + d_2 \times P_2 + d_3 \times P_3 + d_4 \times P_4}{d_1 + d_2 + d_3 + d_4} \tag{5}$$

where P_{ij} is the pixel values which will be restored, P_1, P_2, P_3, P_4 respectively are the pixel value of the top, bottom, left, and right closest to P_{ij}, and d_1, d_2, d_3, d_4 are the corresponding distance to P_{ij}. The better the results of the restoration, the smoother the lost MB will be. However, when more edges go through the lost MB, WAM can result in edge information loss.

4.2 Improved Directional Interpolation

The edges in the lost MB should be detected according to available boundary pixels. The Sobel operator is used to detect edges, which has the property of circularity. The Sobel operator is defined by Eq. (6):

$$S_x = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix}, \quad S_y = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix}. \tag{6}$$

The gradients are computed at the pixel F(i, j) by Eq. (7):

$$\begin{aligned} G_x(i,j) &= vec(S_x)^T vec(F(i,j)) \\ G_y(i,j) &= vec(S_y)^T vec(F(i,j)) \end{aligned} \tag{7}$$

where G_x(i, j) and G_y(i, j) represent the vertical gradient and the horizontal gradient, respectively, and (i, j) is the horizontal and vertical coordinate for pixel point F. Therefore:

$$\begin{aligned} G(i,j) &= \sqrt{G_x^2(i,j) + G_y^2(i,j)} \\ \theta(i,j) &= \arctan\left(\frac{G_y^2(i,j)}{G_x^2(i,j)}\right) \end{aligned} \tag{8}$$

where G(i, j) and θ(i, j) are the magnitude and angle of the gradient for each pixel F(i, j).

Some false edges will still remain due to noise. As the larger magnitude of the gradient for pixels reflects the edge information, it therefore sets threshold Tm to eliminate these false edges. Experimentally, Tm was found to be equal to 60. Considering the characteristics and complexity of the pixels, the valid edge is

Fig. 2 Eight directional edge
categories
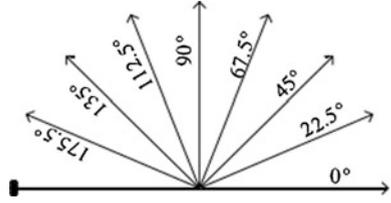
Fig. 3 The directional
interpolation
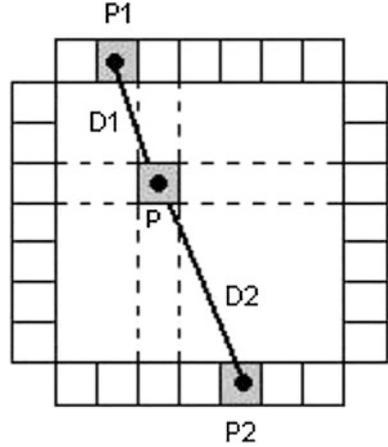
quantized into eight direction levels, as depicted in Fig. 2. Each direction level
covers 22.5° from 0 to 180°. The located direction of maximum magnitude for each
direction level means restructuring edge direction [7].

As depicted in Fig. 3, the pixel p_i^* is interpolated along the i edge direction. This
approximate value can be found by Eq. (9):

$$p_i^* = w_{di} \times \frac{D_1 \times p_2 + D_2 \times p_1}{D_1 + D_2} \qquad (9)$$

where p_1 and p_2 are two boundary pixels along the direction of the ith edge, and D_1
and D_2 are the corresponding distance to the interpolate pixel. The weighting factor
w_{di} is the degree of similarity between p_1 and p_2.

w_{di} is given by Eq. (10) as follows [8]:

$$w_{di} = \exp\left(-\frac{1}{\ell}|p_1 - p_2|\right) \qquad (10)$$

where ℓ is the defined scale factor. Experimental results indicate that ℓ is equal to
10. The value of w_{di} indicates the accuracy of the restructuring edge, and the
corresponding two boundary pixels are more similar.

4.3 Directional Entropy

The directional entropy [9] is defined by Eq. (11) as follows:

$$H_d = -\sum_{x \in EP} p(d_x) \log_2 p(d_x) \tag{11}$$

where EP is an ensemble of the boundary pixels x surrounding the lost MB, and d_x is the number of direction levels. When the probability of 8 directions is identical, the directional entropy is maximum (3 bits). If H_d has a larger value, the restructuring edge is a false edge. In that case, the weighted average method is adopted to obtain good results. Experimental results indicate that the threshold of directional entropy H_d should be set at 2.6. Thus, the proposed algorithm is as follows:

$$
\begin{aligned}
&If \quad |G(x,y)| < T_0 \quad or \quad H < H_{th}\\
&\qquad apply \quad WAM\\
&else\\
&\qquad apply \quad \text{Improved Directional Interpolation}
\end{aligned}
$$

5 Experiments and Comparisons

The proposed algorithm was implemented in H.264/AVC JM16.2. Under the conditions of the same packet loss rate, the proposed algorithm was compared with the JM standard [10]. Test sequences in QCIF format, such as Foreman, Mobile, Akiyo, News and carphone, were implemented in the experiment. The sequence type was IPPP, and each test sequence had 100 frames. Figure 4 shows a comparison of PSNR performance of the four algorithms of different sequences when the loss packet ratio was 15, 10, and 5 %, respectively. The results imply that the proposed algorithm is generally superior to other methods.

In Fig. 5, the restoration effects of the image in the second frame of foreman can be seen, where the original frame is (a) and the corrupted frame is (b). Figure 5c–e are the image restored with DI and BMA. These results clearly show that the proposed algorithm is superior to other algorithms. It can increase the image quality by means of adjustment of the quantization parameter value when video stream transmits in wireless channels.
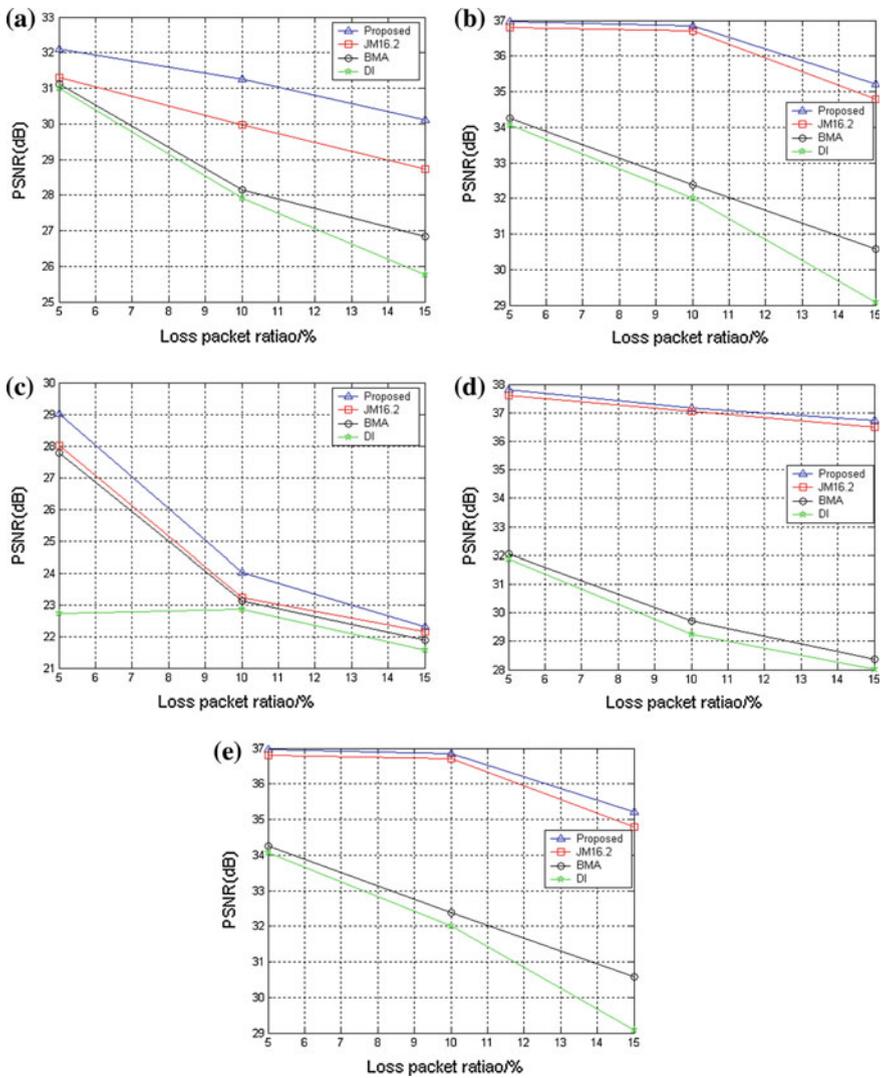
Fig. 4 Comparisons of PSNR curves with JM16.2. **a** Foreman sequences, **b** mobile sequences, **c** Akiyo sequences, **d** news sequences, **e** carphone sequences
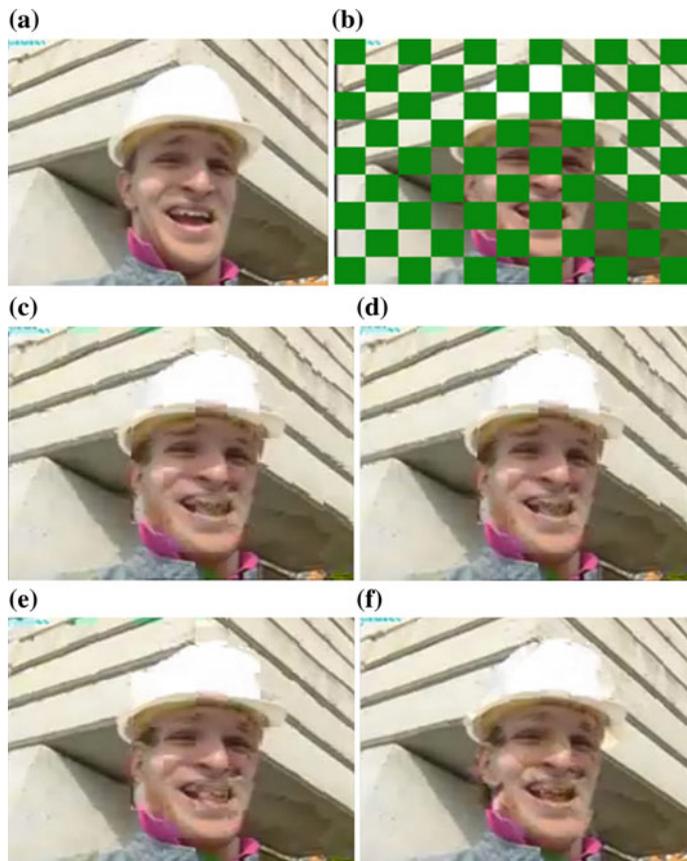
Fig. 5 Recovery performance for foreman (QCIF). **a** The error-free frame, **b** the corrupted frame, **c** DI, **d** BMA, **e** JM, **f** Proposed

6 Conclusion

An algorithm that adaptively selects TEC or SEC on the basis of the change of surroundings is proposed. SEC makes use of direction entropy to decide WAM or IDI, where TEC uses improved BMA to select the best MV. Experiment results strongly suggest that the PSNR performance of this scheme is better than that of JM16.2 when time is invariant.

References

1. Kung, W.Y., Kim, C.S., Kuo, C.C.J: Spatial and temporal error concealment techniques for video transmission over noisy channels. IEEE Trans. Circuits Syst. Video Technol. **16**(7), 89–803 (2006)
2. Sanchez, G., Saldanha, M., Zatt, B., Porto M., Agostini, L.: S-GMOF: a gradient-based complexity reduction algorithm for depth-maps intra prediction on 3D-HEVC. In: 2015 IEEE 6th Latin American Symposium on Circuits and Systems (LASCAS), pp. 1–4, 24–27. IEEE, Piscataway (2015)
3. Tekeli, N., Kamisli, F.: Intra-frame coding with recursive intra prediction and adapted block transforms. In: IEEE Visual Communications and Image Processing Conference, pp. 185–188, 7–10. IEEE, Piscataway (2014)
4. Rhee, C.E., Kim, T.S., Lee, H.J.: An H.264 high-profile intra-prediction with adaptive selection between the parallel and pipelined executions of prediction modes. J IEEE Trans. Multimedia **16**(4), 947–959 (2014)
5. Chen, J., Su, S.M., Zhou, J.K., Gao, L., Dong, X.S.: An adaptive spatio-temporal error concealment scheme based on H.264/AVC. J. Multimedia **6**(3), 293–299 (2011)
6. Kuan, Y.K., Li, G.L., Chen, M.J., Tai, K.H., Huang, P.C.: Error concealment algorithm using inter-view correlation for multi-view video. J. EURASIP J. Image Video Process. **38**(11), 187–198 (2014)
7. Xiang, Y.J., Feng, L.M., Xie, S.L., Zhou, Z.H.: An efficient spatio-temporal boundary Matching algorithm for video error concealment. Multimedia Tools Appl. J. **52**(1), 91–103 (2011)
8. Chen, Y., Hu, Y., Au, O.C., et al.: Video error concealment using spatio-temporal boundary matching and partial differential equation. J. IEEE Trans. Multimedia **10**, 2–15 (2008)
9. Correa, G., Assuncao, P., Agostini, L., Cruz, L.A.D.S.: Four-step algorithm for early termination in HEVC inter-frame prediction based on decision trees. In: IEEE Visual Communications and Image Processing Conference, pp. 65–68, 7–10. IEEE, Piscataway (2014)
10. Chen, X., Chung, Y., Bae, C.: Dynamic multi-mode switching error concealment algorithm for H.264/AVC video applications. J. IEEE Trans. Consum. Electron. **54**(1), 154–162 (2008)

An Autonomous Vehicle Navigation System Based on Hough Transform and Fuzzy Logic

Sheng-Zhi Du and Chun-Ling Tu

Abstract This paper proposes an autonomous vehicle navigation system based on Hough Transform and fuzzy logic techniques. Hough Transform is used to detect the lane lines and the position of the vehicle in the lane. The vehicle position and the deviation of the proceeding orientation are used to control the steering system. Fuzzy logic control scheme is used for steering the vehicle. The software simulation study validated the proposed system.

Keywords Hough transform · Lane line detection · Fuzzy logic · Auto navigation vehicle

1 Introduction

Building autonomous vehicles is becoming an active and challenging issue in the field of robotics. The advantage of autonomous is obvious that can reduce the traffic accident to save human lives, release human work loads, and so on. The developing of autonomous vehicles attracts a number of vehicle manufactures, research institutions even military department, where lots of financial and human resources were input to relevant projects. The basic and important goal to develop autonomous vehicles is the auto navigation under various uncertainties such as bad road, luminance and weather situations, etc.

S.-Z. Du (✉)
Department of Mechanical Engineering, Mechatronics and Industrial Design,
Tshwane University of Technology, Pretoria 0001, South Africa
e-mail: dushengzhi@gmail.com

C.-L. Tu
Department of Computers System Engineering, Tshwane University of Technology,
Pretoria 0001, South Africa
e-mail: tclchunling@gmail.com

Computer vision techniques are commonly used in the automatic navigation by visual data [1, 2]. In the vehicle navigation [3], lane lines are detected by analyzing the obtained visual data, then the auto navigation is achieved by tracking the lane lines. The detection speed and accuracy are the main concerns for the road lane detecting and tracking, but it is also a challenging and complicate work due to the problems from poor road marking, lightness, complicate multiple lanes and so on. As a robust line detector, the Hough Transform [4] and its variations are commonly used, such as lane detection technique [5]. Road lane recognizing systems were limited to semiautomatic.

In this paper, an automatic vehicle navigation system is proposed. We divided the proposed methods to three progresses: firstly, road lane detection based on an improved Hough transform [6] is employed; secondly, Fuzzy logic scheme is designed to steer the vehicle; at last, 3D vehicle and different road scenario are built using Virtual Reality (VR). The simulation results validated the proposed system.

2 Related Work

The computer vision based navigation of autonomous vehicles is challenging since the uncertainties in the vehicle locating/lane detection and navigation/lane keeping. For lane detection, some interferences should be considered, such as the imperfect lane lines, errors in lane recognition, and the unpredictable road orientation changes. In the aspect of navigation, the controller has to be tolerant to the uncertainties like the road profile variation, lane width, and acceptable spacing for driving, etc. All these factors negatively affect the reliability of the autonomous system. So it is important to carefully choose and tune the lane detection method and controller.

2.1 Hough Transform

The road detecting methods must be robust under imperfect lane lines and undesired luminance and weather condition. In straight line and segment detection methods, the Hough transform (HT) [4] is one of the most widely used techniques. The HT is defined by the mapping from the image space to the parameter space (i.e. the HT space) by Eq. (1).

$$\rho = x \cos\theta + y \sin\theta, \tag{1}$$

Sharp distinct peaks are desired when mapping feature points from the image space to the HT space. However problems exist such as peak splitting, flattening [7]. With the unreliable peaks, some methods were reported to extract information from the butterfly surrounding the peaks to determine straight line segment

parameters, such as the segment detection method making use of a butterfly's quadrangular neighborhood [6]. These methods do not use feature points in the image space therefore have high computationally efficiency. Another advantage of these methods is the robustness against the peak splitting and flatting problems.

2.2 Fuzzy Logic Control

Fuzzy logic [8] is a nature language like reasoning system. It expresses concepts by membership functions, that is, an operation is not an exact numerical value but can be cast in such terms like human language, so that the controller can be constructed based on human operators' experiences, such as controlling the position of a mobile robot [9], and robot navigation [10]. Since fuzzy logic control does not rely on exact measurement from the sensors, it makes the controllers feasible for noise and even fault tolerant applications, for instance, fault tolerant of satellite synchronization system using fuzzy control [11] and fuzzy control for nonlinear system with Multiplicative Noises [12].

3 The Proposed Method

The proposed navigation system and the vehicle are implemented in virtual reality (VR) for the purpose of validation. The scenario of the auto navigation system is shown in Fig. 1. To mimic the real world situation, real road images are used to build the testing road in VR as shown in Fig. 1a. A camera is mounted on the side of the vehicle to capture the road image for lane line detection. The captured image is shown in Fig. 1b. The total length of the testing road is 4 km with straight parts and turning corners as shown in Fig. 1c, and the designed vehicle driving speed is 80 km/h.
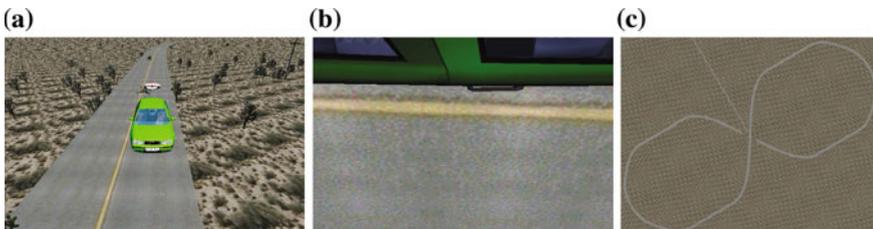
Fig. 1 The road and car scenario. **a** A car on a road, **b** lane line to be detected, **c** the map of the road

3.1 Lane Lines Detection

Considering the fact that the quality of the lane lines on real world roads is not always high, it is important to employ a robust lane line detection method. Because of the requirement of realtime navigation, the method with low computation load is preferable. In this research we apply the improved Hough transform [13] which has the capacity to detect segment directly and the computation load is slight.

3.2 Vehicle Auto Navigation

In the navigation, the vehicle must be kept in the lane without any part beyond the lane lines. To assure this requirement, the distances of the vehicle front and back corners to the lane line are chosen as the input variables of fuzzy controller, denoted as y_1 and y_2 respectively. The difference of y_1 and y_2 is also an input variable, which provides the relative orientation of the vehicle and is used for finer tuning the steering output. The membership functions of y_1 and y_2 are the same as shown in Fig. 2a, where 3 levels are defined as "small", "media" and "big". The membership function of $y_1 - y_2$ is depicted in Fig. 2b, where 5 levels are defined as "-big", "-media", "media", "+media", and "+big". According to the human driving experiences, the following 14 rules are defined.
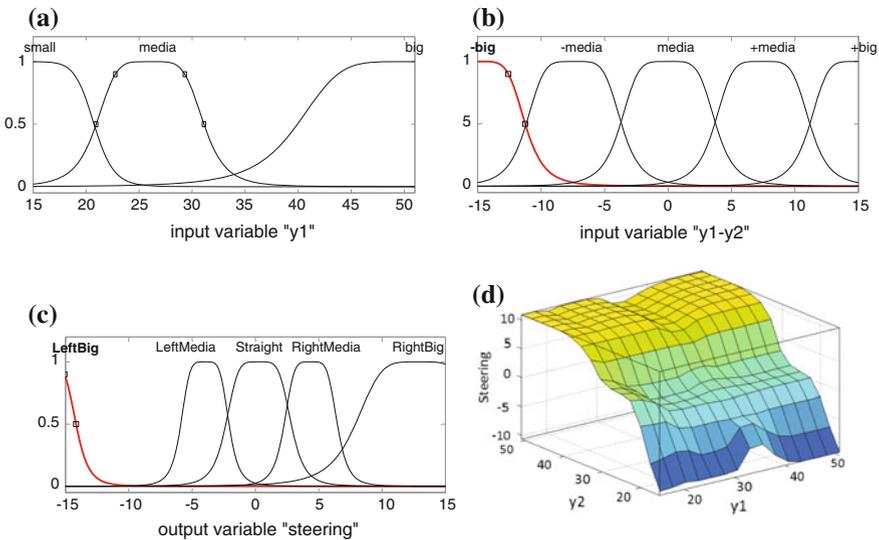
Fig. 2 The fuzzy logic control design. **a** Member function of input variable y_1, **b** member function of input variable $y_1 - y_2$, **c** member function of output variable (steering), **d** the surface of steering versus y_1 and y_2

1. *If (y_1 is small) and (y_2 is small) then (steering is LeftBig)*
2. *If (y_1 is small) and (y_2 is media) then (steering is RightMedia)*
3. *If (y_1 is small) and (y_2 is big) then (steering is RightBig)*
4. *If (y_1 is media) and (y_2 is small) then (steering is LeftBig)*
5. *If (y_1 is media) and (y_2 is media) then (steering is RightBig)*
6. *If (y_1 is media) and (y_2 is big) then (steering is RightBig)*
7. *If (y_1 is big) and (y_2 is small) then (steering is LeftBig)*
8. *If (y_1 is big) and (y_2 is media) then (steering is Straight)*
9. *If (y_1 is big) and (y_2 is big) then (steering is RightBig)*
10. *If (y_1 is media) and (y_2 is media) and ($y_1 - y_2$ is -media) then (steering is RightMedia)*
11. *If (y_1 is media) and (y_2 is media) and ($y_1 - y_2$ is media) then (steering is Straight)*
12. *If (y_1 is media) and (y_2 is media) and ($y_1 - y_2$ is +media) then (steering is LeftMedia)*
13. *If ($y_1 - y_2$ is +big) then (steering is LeftBig)*
14. *If ($y_1 - y_2$ is -big) then (steering is RightBig)*

Figure 2c is the steering variable membership function. According to the membership functions and rules, the steering surface is demonstrated in Fig. 2d.

4 Results

For the sake of comparison, the proportional-derivative (PD) controller is employed as well. Figure 3 shows the steering action and the distance from the vehicle to the lane line. In Fig. 3a, the fuzzy controller is more robust on the steering action, while the PD controller change the steering in high frequency and big magnitude. The difference comes from the robustness of the controllers against the noise, such as
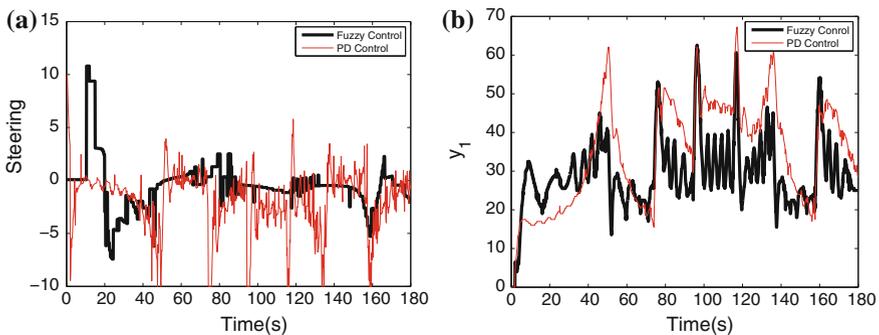
Fig. 3 Driving performance comparison. **a** The steering action, **b** the distance from the lane

the lane line detection errors. Because fuzzy controller is more tolerant to the noise, it is not very sensitive to the noise. This is not the case for PD controller.

In Fig. 3b, the distances between the vehicle and lane line indicate the steering profile of the two controllers. One can find that the fuzzy control steers the vehicle like human being, which keep steering the vehicle even on a straight road with slow and smooth steering actions. However, the PD controller shows big vibration during the steering, which is harmful to the mechanical system of the vehicle.

5   Conclusion

In this paper, an autonomous vehicle navigation system was proposed. An improved Hough transform was employed to detect the lane lines for locating the vehicle and provide lane information to steering system. Fuzzy logic control was applied to steering the vehicle. The result validated the robustness of the fuzzy controller, which steered the vehicle in such a way like human being. The simulation results indicated that the fuzzy controller got better performance than the PD controller.

References

1. Hoang, V.D., Hernandez, D.C., Hariyono, J., Jo, K.H.: Global path planning for unmanned ground vehicle based on road map images. In: IEEE International Conference, Human System Interactions (HSI) pp. 82–87. IEEE, Piscataway (2014)
2. Luo, C., Wu, Y., Krishnan, M., Paulik, M.: An effective search and navigation model to an auto-recharging station of driverless vehicles. In: Computational Intelligence in Vehicles and Transportation Systems (CIVTS), pp. 100–107. IEEE, Piscataway (2014)
3. Umamaheswari, V., Amarjyoti, S., Bakshi, T., Singh, A.: Steering angle estimation for autonomous vehicle navigation using Hough and Euclidean transform. In: IEEE International Conference, Signal Processing, Informatics, Communication and Energy Systems (SPICES 2015), pp. 1–5. IEEE, Piscataway (2015)
4. Hough, P.V.C.: A method and means for recognizing complex patterns. US Patent 3,069,654 (1962)
5. Vijay, G., Shashikant, L.: Lane departure identification for advanced driver assistance. IEEE Trans. Intell. Transp. Syst. **16**(2), 910–918 (2015)
6. Du, S., van Wyk, B.J., Tu, C., Zhang, X.: An improved Hough transform neighborhood map for straight line segments. IEEE Trans. Image Process. **19**(3), 573–585 (2010)
7. Shapiro, V.: Accuracy of the straight line Hough transform: the non-voting approach. Comput. Vis. Image Underst. **103**, 1–21 (2006)
8. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)

9. Cho, J.T., Nam, B.H.: A study on the fuzzy control navigation and the obstacle avoidance of mobile robot using camera. In: IEEE International Conference on Systems, Man, and Cybernetics, pp. 2993–2997. IEEE, Piscataway (2000)
10. Rusu, P., Petriu, M.: Behavior-based neuro-fuzzy controller for mobile robot navigation. IEEE Trans. Instrum. Meas. **52**(4), 1335–1340 (2003)
11. Li, J., Kumar, K.D.: Fault-tolerant control for satellite attitude synchronization. IEEE Trans. Fuzzy Syst. **20**(3), 572–586 (2011)
12. Chang, W., Tsai, Y., Ku, C.: Fuzzy fault tolerant control via Takagi-Sugeno fuzzy models for nonlinear systems with multiplicative noises. Int. J. Fuzzy Syst. **16**(3), 338–349 (2014)
13. Du, S., van Wyk, B.J., Tu, C., Chen, Z.: Collinear segment detection using HT neighborhoods. IEEE Trans. Image Process. **20**(12), 3520–3612 (2011)

An Efficient Study on Large
and Complex Network Modeling

Jun-Chun Ma, Min Li, Shan Chen and Wei Zhang

Abstract In order to support the automatic construction of an attack graph, a simple, flexible and complete network system model is necessary, which uses formal methods to describe network system resources and security elements. Since host reachability relations will be influenced by the filter device (router or firewall) in the network, making automatic construction difficult to obtain. Based on a detailed study of the relevant literature, a method of obtaining host reachability relations automatically is proposed. Based on the comprehensive analysis of network topology information and the configuration rules of the filter device, the proposed method first generates physical connection relations automatically; next, it combines the configuration rules of the filter device to obtain a set of host reachability relations. Experimental results indicate that this method exhibits high production efficiency and reduces consumption of system resources when constructing attack graphs, which can be used to evaluate the overall safety of large and complex network systems.

J.-C. Ma · M. Li · S. Chen (✉) · W. Zhang
Research Institute of High-Tech Hongqing Town, Xi'an, Shanxi, China
e-mail: chenshan1223@126.com

J.-C. Ma
e-mail: Mana119577@sohu.com

M. Li
e-mail: limin2314@126.com

W. Zhang
e-mail: weiz0124@sohu.com

1   Introduction

Network modeling is a prerequisite for network security assessments, which creates the initial attack scenario of the target network system [1, 2]. The parameters of the host, server, application and vulnerability, etc., can be automatically acquired by Nessus scanner and detection tools, but host reachability relations will be influenced by the filter device used in the network. Thus, it is difficult to be automatically obtain. In order to resolve the problems of obtaining host reachability relations, most existing methods assume that host reachability relations have exist [3]; the second method involves the use of Nessus and other scanning tools due to the existence of filter devices in the network, so this method cannot completely and accurately obtain host reachability relations [4–6]. The third commonly-used method is based on firewall rules, routing rules, NAT (Network Address Translation) and other network configuration information, which must generate the reachability matrix from the source host port to the destination host port. However, because there is no port information in the link layer, the network layer or the transport layer of a Transfer Control Protocol/International Protocol (TCP/IP) model, this method is also not comprehensive [7, 8]. Taking the above considerations into account, a method of automatically obtaining host reachability relations is proposed in order to support the automatic construction of an attack graph based on the in-depth study of the parameters of the automatic model. The attack graph depicts the process of network safety destruction by an attacker based on the network vulnerabilities, and supports quantitative vulnerability calculation. The proposed method also can help the network safety manager prioritize safety defense, which ensures network security with the least cost.

2   Method of Automatic Obtaining Host
Reachability Relations

One of the preconditions for an attacker to take advantage of network vulnerability to assault the network is that one or more vulnerabilities must definitely exist in the host; that is to say, there is a host reachability relation between the attacker host and the target host. However, in practice, host reachability relations will be influenced by filter devices [9, 10]. Therefore, it must consider two important factors for obtaining host reachability relations: (1) the network topology information [11, 12], which describes the physical connection between hosts and is the basis for host reachability relations; (2) the rule configuration information [13, 14], which reflects the basis for forwarding or filtering data packets and is the decisive factor in host reachability relations. Thus, this paper proposes a method to automatically obtain host reachability relations.

In the implementation process, the method of obtaining host reachability relations automatically proceeds in two steps. The first step is to automatically generate the physical connection, whose input consists of the target network topology information (the set of *Interface* and *Link*), and whose output is *Physical-Connection*; the second step is to automatically generate the logical connection, whose input is *Physical-Connection* and rule configuration information (the set of rule), and whose corresponding output is the set of host reachability relations *Connection*.

2.1 Algorithm of Generating the Physical Connection

Table 1 shows the algorithm used to automatically generate physical connection. Steps (1) through (5): it must be determined whether the type of device in the target network system is a host or a server. Then, the interface information of host or server is written into the set of Physical-Connection. Steps (6) through (28): the element of Physical-Connection named PCn is extracted in turn, and its physical connection is analyzed. Specifically, the functions of steps (9) through (13) determine the type of PCn of the last interface (host, server, or neither). If soothe type of PCn can be determined, it indicates that this connection has been analyzed; if not, the functions of steps (14) through (19) determine connection k from Link, of which the source interface is i, by first analyzing whether connection k has been analyzed. If so, PCn is deleted from Physical-Connection; if not, the target interface information of k is written into PCn. The functions of steps (20) through (28) determine whether the target interface information of k is host, server, or neither; if neither, connection l in Link is written into PCn, of which the source equipment is identical to the destination equipment of connection k. If l is greater than one, a new connection must be added into Physical-Connection. When the final elements of all connections in Physical-Connection have been identified as host or server, the algorithm ends.

2.2 Algorithm Used to Generate the Logical Connection

In the implementation process, two rule filter functions are employed: the access control list (ACL) rule filter function and the routing rule filter function. Table 2 depicts the ACL rule filter function. The functionality of step (1) is to extract one rule a from A, with A denoting a subset of the ACL rule set which is generated in the main function. Step (2): the set of host reachability relations C_0 before filtering; the functionality of this step is to extract the host connection relation c from C_0, of which the source IP address *HostS* and destination IP address are identical to those of a. Steps (3) through (10): if the protocol is identical, it must determine whether the Action of a is Deny or not; if so, c will be deleted. The function of step (11) is to

Table 1 Algorithm used to generate the physical connection

```
     name:algorithm of generating the physical connection
     input: network topology information:Interface,Link
     output: Physical-Connection
```

(1)	$ip=0$;
(2)	for ($i×Interface$, $i.Type=="Host"$ || $i.Type=="Server"$)
(3)	$\{PC_{ip} =\{<i.Name_{device}$, $i.Name_{interface}$, $i.Type>\}$;
(4)	$Physical\text{-}Connection=Physical\text{-}Connection\cup PC_{ip}$;
(5)	$ip++$;}
(6)	$flag$=true;
(7)	while ($flag$)
(8)	{$flag$=false;
(9)	for ($PC_{n}\in Physical\text{-}Connection$)
(10)	{$Y=PC_{n}$;
(11)	$i=Y_{last}$;
(12)	if ($i!=Y_{first}$ && ($i.Type =="Host"$ || $i.Type$ =="Server"))
(13)	break;
(14)	for ($k\in Link$, $k.NameS_{device}==i.Name_{device}$, $k.NameS_{interface}==i.Name_{interface}$)
(15)	{if ($Y\cap<k.NameD_{device}$, ANY, ANY>$!=\Phi$)
(16)	{DEL PC_{n} FROM $Physical\text{-}Connection$;
(17)	break;}
(18)	$PC_{n}= PC_{n}\cup<k.NameD_{device}$, $k.NameD_{interface}$, $k.TypeD>$;
(19)	$Y= PC_{n}$;
(20)	if ($k.TypeD!="Host"$ && $k.TypeD !="Server"$)
(21)	{$flag1$=true;
(22)	for ($l×Link$, $l.NameS_{device}==k.NameD_{device}$)
(23)	{if ($flag1$)
(24)	{$PC_{n}= PC_{n}\cup<l.NameS_{device}$, $l.NameS_{interface}$, $l.TypeS>$;
(25)	$flag1$=false; }
(26)	else
(27)	$Physical\text{-}Connection_{last+1}=Y\cup<l.NameS_{device}$, $l.NameS_{interface}$, $l.TypeS>$;}}
(28)	$flag$=true; }}

extract the intersection of port between c and a. Finally, C_0 is returned, generated to Connection. Table 3 shows the routing rule filter function. The purpose of step (1) is to define an empty set C; the purpose of step (2) is to extract a rule r from the set of rule R, which is a subset of the router rule generated in the main function; the purpose of step (3) is to extract the host connection relation c from C_0, with an IP address identical to that of r; the purpose of step (4) is to write c into C and finally, to return C to *Connection*.

Table 2 ACL rule filter function

Name: ACL rules filter function $Connection=f_A(C_0,A)$
input: the set of host reachability relations before filtering C_0, A
output: the set of host reachability relations after filtering $Connection$

(1)	{for $(a \in A)$
(2)	{for $(c \in C_0,$ $c.HostS==a.HostS$ && $c.HostD==a.HostD)$
(3)	{if $(a.\ Pro==$ANY && $a.Action==$Deny)
(4)	{DEL c FROM C_0;
(5)	break; }
(6)	if $(c.\ Pro!=a.\ Pro$ && $a.\ Pro!=$ANY)
(7)	break;
(8)	if $(a.Action==$Deny)
(9)	{DEL c FROM C_0;
(10)	break; }
(11)	$c.Port =c.Port \cap a.Port;$}}
(12)	return C_0;}

Table 3 Routing rule filter function

Name: routing rules filter function $Connection=f_R(C_0, R)$
Input: the set of host reachability relations before filtering C_0, R
Output: the set of host reachability relations after filtering $Connection$

(1)	{$C=\Phi$;
(2)	for $(r \in R)$
(3)	{ for $(c \in C_0,$ $c.\ HostD/c.MaskD \in r.Host/r.Mask)$
(4)	$C=C \cup c;$}
(5)	return $C;$}

Table 4 shows the algorithm used to automatically generate logical connections. Steps (1) through (5): based on the physical connection *Physical-Connection*, the initial set of host reachability relations *Connection* is established. Steps (6) through (11): first, the connection p in *Physical-Connection* and interface i of p are extracted repetitionally; then the rule set A in the ACL rule set is extracted, which is associated with i. Steps (12) through (22): based on the position of i, if it represents the

Table 4 Algorithm used to generate the logical connections

```
        name: algorithm of generating the logical connection
        input: Physical-Connection, Rule
        output: the set of host reachability relations Con-
        nection
(1)     Connection=Φ;
(2)     for (p∈Physical-Connection)
(3)     {i=p. Interface_{first};
(4)       j=p. Interface_{last};
(5)       Connection=Connection∪{<i.Host, j.Host,ANY,ANY>};}
(6)     for (p∈Physical-Connection)
(7)     {for (i∈p)
(8)       {A=Φ;
(9)         R=Φ;
(10)      for(a∈Rule.ACL, a.Name_{device}==i.Name_{device} &&
        a.Name_{interface}==i.Name_{interface})
(11)         A=A∪a;
(12)       if (i==p.Interface_{first})
(13)         Connection =f_{A}(Connection,A);
(14)       else if (i==p.Interface_{last})
(15)       {Connection=f_{A}(Connection,A);
(16)          break;}
(17)       else
(18)       {if ((i-1).Name_{device}==i.Name_{device})
(19)       { for (r×Rule.Router, r.Name_{device}==i.Name_{device} &&
        r.Name_{interface}==i.Name_{interface})
(20)            R=R∪r;
(21)          Connection =f_{R}(Connection,R);}
(22)          Connection=f_{A}(Connection,A);}}}
```

first or last interface, this indicates that the type of device in which *i* is located is the host or server. Then, the ACL rules are used to filter *Connection*; if it is the middle interface, when the device in which *i* located is identical to the device in which the front interface is located, the routing rules are used to filter, then the ACL rules are used to filter; otherwise, on the ACL rules are necessary to filter.

3 Experimental Results and Analysis

The method of automatically obtaining host reachability relations proceeds in two steps: generating the physical connection and generating the logical connection. This paper conducts numerous experiments using these two algorithms separately. The experimental environment is described as follows: the server is PowerEdge R710, the operating system is RetHat v5.4, memory is 32G, and CPU is 2.26 GHz.

3.1 Experimental Results and Analysis of Physical Connection Generation

The performance of automatically generating physical connection will be influenced by the following parameters: the number of hosts H, the maximum number of devices connecting any two hosts D, and the number of interfaces of each device F, among other parameters. To verify the influence of these parameters under different network environments, this paper designs three groups of experiments. Figures 1, 2 and 3 depict the experimental results.

The three groups of experimental results indicate that in the algorithm used to generate physical connection, the maximum number of devices which connect any two hosts D has the greatest impact, the number of interfaces of each device F exercises the second greatest impact, and the number of hosts H has the least impact.

Fig. 1 Influence of H
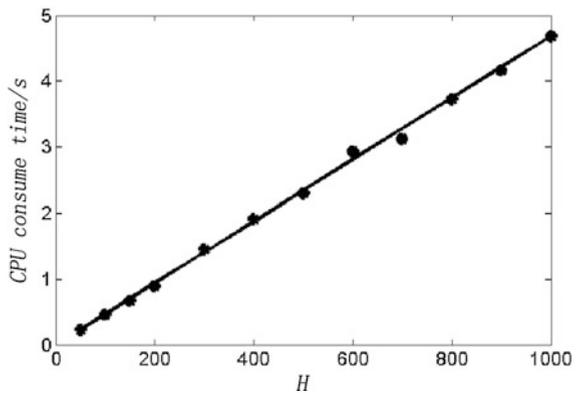
Fig. 2 Influence of D
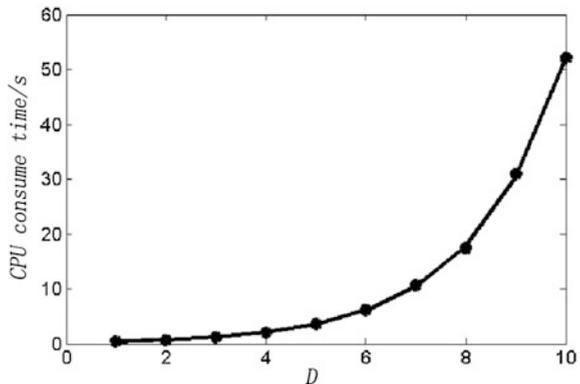
Fig. 3 Influence of F
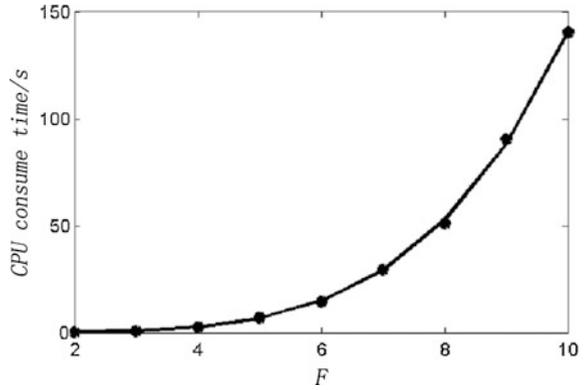
3.2 Experimental Results and Analysis of Logical Connection Generation

The experiments conducted to generate logical connection are similar to the experiments investigating the generation of *Physical connection*. This paper designs five groups of experiments; Figs. 4, 5, 6, 7 and 8 depict the experimental results.

The five groups of experimental results indicate that in the algorithm used to generate logical connection, the maximum number of devices which connect any two hosts D has the greatest impact, the number of interfaces of each device F has the second greatest impact, followed by the number of hosts H, the amount of rule information in the *ACL* rule set $|ACL|$ and the amount of routing information in the router rule set $|Router|$.

Fig. 4 Influence of H
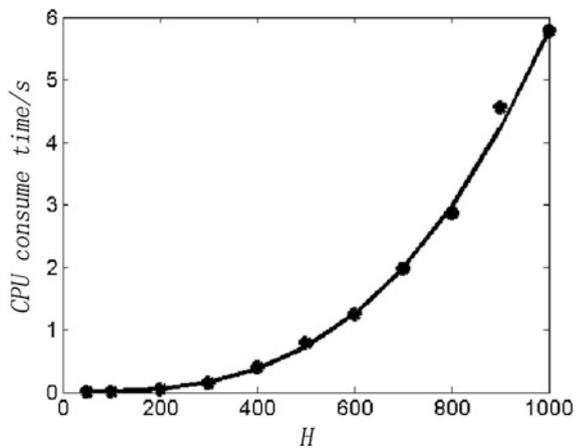
Fig. 5 Influence of D
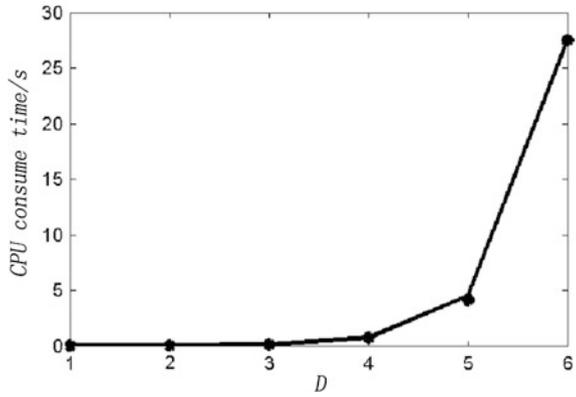
Fig. 6 Influence of F
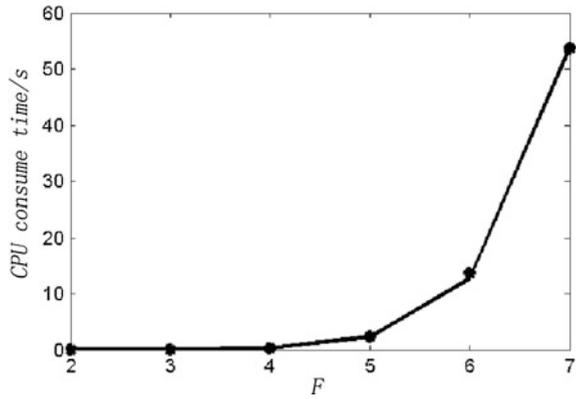
Fig. 7 Influence of |ACL|
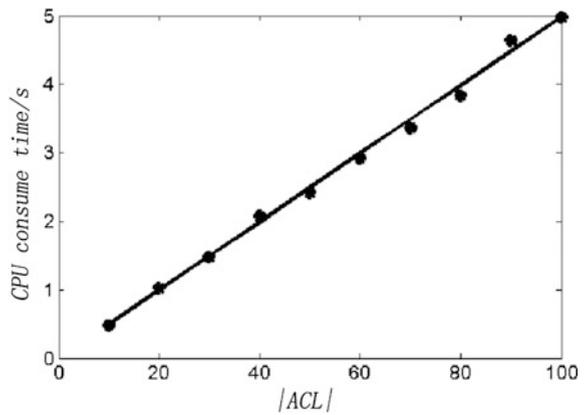
Fig. 8 Influence of |Router|
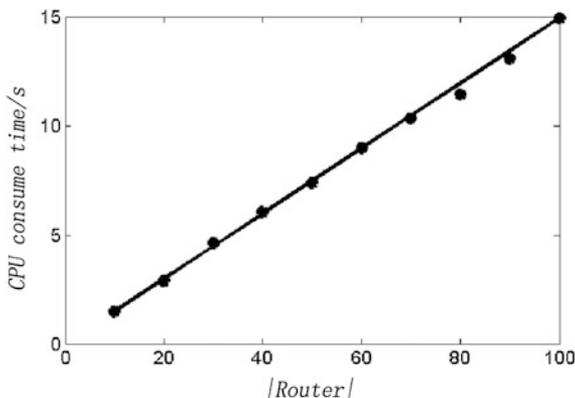
4   Conclusion

In order to establish a simple, flexible and completely automatic target network system model which can support the attack graph construction of large and complex networks, this paper presents a method to automatically obtaining host reachability relations. This proposed method first generates the physical connection, whose input is the target network topology information; based on this, it then combines with the configuration rules of the filter device to obtain the set of host reachability relations. Experimental results indicate that this method can effectively support the scalability of attack graph construction, and it is of potential application value for assessing the security of large-scale complex networks.

References

1. Ma, J.C.: Research on key technology of network vulnerability assessment based on attack graph. Ph.D. thesis, Second Artillery Engineering University, Xi'an (2012)
2. Ma, J.C., Wang, Y.J., Sun, J.Y.: A novel method of constituting network security policy based on attack graphs. Chin. High Technol. Lett. **22**(4), 17–24 (2012)
3. Wu, D., Lian, Y.F., Chen, K., et al.: A security threats identification and analysis method based on attack graph. Chin. J. Comput. **39**(9), 1938–1950 (2012)
4. Hui, D.P.: Research on network systems security assessment technology. Ph.D. thesis, Harbin Engineering University, Harbin (2009)
5. Chen, F.: A hierarchical network security risk evaluation approach based on multi-goal attack graph. Ph.D. thesis, Graduate School of National University of Defense Technology, Changsha (2009)
6. Jajodia, S., Noel, S.: Advanced cyber attack modeling, analysis, and visualization. Technical Report AFRL-RI-RS-TR-2010-078, Air Force Research Laboratory (2010)
7. Lippmann, R.P., Ingols, K.W., Scott, C., et al.: Evaluating and strengthening enterprise network security using attack graphs. Technical Report ESC-TR-2005-064, Lincoln Laboratory (2005)

8. Ingols, K., Lippmann, R., Piwowarski, K.: Practical attack graph generation for network defense. In: Proceedings of the 22nd Annual Computer Security Applications Conference (ACSAC'06), pp. 121–130. IEEE Computer Society, Washington DC (2006)
9. Yang, M., Yang, P.L.: Access control list technology studying and application. Comput. Technol. Dev. **21**(6), 145–149 (2011)
10. Evans, K.N., Kilcoyne, K.G., et al.: Predisposing risk factors for non-contact ACL injuries in military subjects. Knee Surg. Sports Traumatol. Arthrosc. **20**(8), 1554–1559 (2012)
11. Kgopa, A.T., Kekwaletswe, R.H.: Implementations of network security in the internet cafes around Pretoria, South Africa. Int. J. Comput. Sci. Electron. Eng. **2**(1), 44–49 (2014)
12. Nahir, A., Orda, A., Freund, A.: Topology design of communication networks: a game-theoretic perspective. IEEE/ACM Trans. Networking **22**(2), 405–414 (2014)
13. Deng, Y., Ling F.: A credit-based incentive routing scheme in DTMSN. J. Sichuan Univ. (Natural Science Edition). **48**(6), 1312–1318 (2011)
14. Fan, Y.J.: Improved routing strategies in complex networks based on local information. Nanjing University Posts Telecommunication, Nanjing (2013)

An Improved SVM-Based Motion Detection Algorithm Using an Accelerometer

Xu-Kang Wu, Xu-Guang Yang, Zheng-Guo Cai and Shan-Shan Luo

Abstract This paper presents an effective detection algorithm for wrist motions based on an improved SVM (support vector machine) classifier. Firstly, a novel windowing method is proposed to enhance the consistency of sampled motion. After extracting characteristic features in both time and frequency domains, a feature scaling process is applied and a C-SVC (C-Support Vector Classifier) is trained by a threshold-based grid-search with cross-validation, to achieve higher accuracy and faster convergence. The experiments demonstrate that the proposed algorithm outperforms the traditional SVM in accuracy and convergence.

Keywords Motion detection · SVM · Grid-search · Cross-validation

1 Introduction

Motion detection has been a popular topic in recent years. With the smart bracelets prevailing more and more quickly, consumer electronics encounter a smarter demand of motion detection. In order to realize such a goal and make it as convenient

X.-K. Wu
Shanghaitech University, 100 Haike Road, Pudong, Shanghai, China

X.-K. Wu (✉) · Z.-G. Cai · S.-S. Luo
Shanghai Institute of Microsystem and Information Technology,
865 Changning Road, Shanghai, China
e-mail: wuxk@shanghaitech.edu.cn

Z.-G. Cai
e-mail: caiziee@gmail.com

S.-S. Luo
e-mail: shary3@163.com

X.-G. Yang
Shanghai Internet of Things Co. Ltd., No.1455 Pingcheng Road,
Jiading District, Shanghai, China
e-mail: yxg@shiot.org

as possible for persons to use, the accelerometers are considered to be a preferred device. Many researchers apply fall detection, gait detection, etc. [1]. Studies have discussed about method using acceleration data in which the main idea is to build a classifier. A wide variety of classification methods are deployed: KNN (k-Nearest Neighbor) algorithms, neural networks, decision trees and support vector machines [2]. In this paper, an improved SVM classifier is introduced to detect wrist motions via acceleration data.

The rest of the paper is organized as follows. Section 2 introduces the fundamental theorem of SVM classifier. The main contributions, including data pre-processing, feature pre-processing, feature extraction and classifier training is presented in Sect. 3. Experimentation is carried out in Sect. 4. Conclusions are drawn in the last.

2 SVM Classifier

In this paper, an improved support vector machine (SVM) is introduced to recognize motions from acceleration data. SVM is widely used in data classifications with its superior performance and usability. Given a set of specific sample data with labels in pairs (x_i, y_i), $i = 1, 2, \ldots, k$ where $x_i \in R^n$ indicates sample variable and $y \in \{1, -1\}^k$ represents label, the SVM constructs an optimization problem [3] as follows:

$$
\begin{aligned}
\min_{w,b,\varepsilon} \quad & 1/2 w^T w + C \sum_{i=1}^{k} \varepsilon_i \\
subject\ to \quad & y_i(w^T \phi(x_i) + b) \geq 1 - \varepsilon_i, \\
& \varepsilon_i \geq 0
\end{aligned}
\tag{1}
$$

where x_i is mapped into a high-dimensional space after function $\phi(x_i)$ and $C \geq 0$ is the regularization parameter. Thus, the SVM mentioned above is also named C-SVC (C-support vector classifier). Structure of the C-SVC is shown in Fig. 1.

In order to obtain a good performance in dealing with high dimensionality, a radial basis function (RBF) is deployed as the kernel function.

$$
K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2), \gamma > 0
\tag{2}
$$

and the decision function is

$$
\mathrm{sgn}(w^T \phi(x) + b) = \mathrm{sgn}\left(\sum_{i=1}^{k} y_i \alpha_i K(x_i, x) + b \right)
\tag{3}
$$
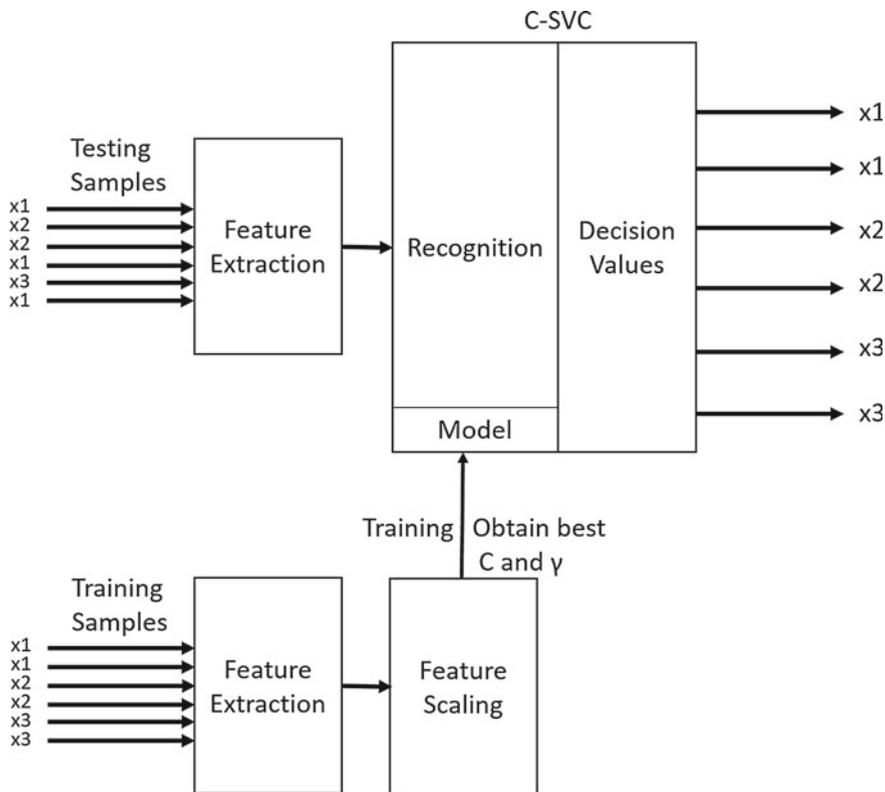
Fig. 1 Whole block diagram of constructing C-SVC

In fact, the output of the decision function has to compete against each other and the winner will be the final factor to determine the class of corresponding input element (testing samples). Specific features of testing samples with labels will be extracted as the input of the classifier. And the C-SVC must be trained by training samples first to construct a model. Traditionally, a C-SVC directly accepts the feature matrix and process with kernel functions to output the final result. No scaling is applied and the grid-search method is greedy search. An improved method consists of scaling and threshold-based grid-search is applied to increase the accuracy and to speed up convergency in obtaining parameters C and γ. Training C-SVC will be depicted in the following sections. For all inputs, the model will compute a decision matrix which has the same rows as the input. The final result will be the column number which the winner value belongs to.

3   Detection Strategy

Traditional method in motion detection consists of two phases [4]: pre-classifier for static and dynamic motions separation, and the main classifier for the dynamic motions. In the pre-classifier phase, a feature named signal magnitude area (SMA) [5] is commonly used to distinguish static motion from dynamic motion. However, whether it can succeed totally depends on the SMA threshold. A static motion usually does have a small value of SMA, while sometimes some disturbance will contribute to a large value of SMA, which is shown in Fig. 2. Consequently, a fixed threshold is not flexible and robust when facing dynamic environment. An important sense should be noticed that a dynamic state can be quite different from a static state in many aspects, such as energy, variance, mean, etc. Thus it is not necessary to add such a phase as the second phase can handle it quite good due to the experimental result. In this paper, three motions have to be classified: static motion, flip over motion and swing motion.

In order to obtain an accurate, reliable and robust classifier, one motion sample will be cut into three windows that overlap in time domain during the data pre-processing step. It is considered that motion is continuous. If one motion is covered in one window, it will lose some information even though the window overlaps either. The main idea of the windowing is to cover as many details as possible to help build a full understanding of the motions which can enrich the features [6]. After that, features of the samples will be computed. And a necessary scaling step will be applied on feature matrix. Afterwards a threshold-based grid-search under 3-fold cross-validation strategy is applied to acquire the best parameters in Eqs. (1) and (2). Finally the C-SVC predicts the testing samples.
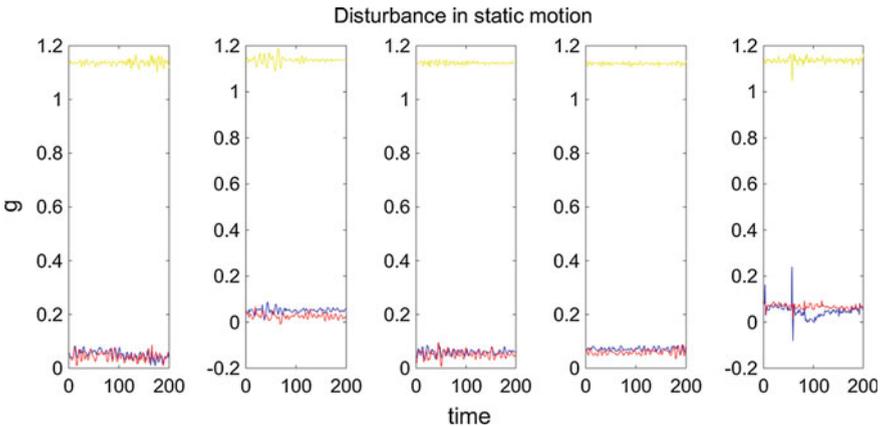
Fig. 2 Disturbance in static motion
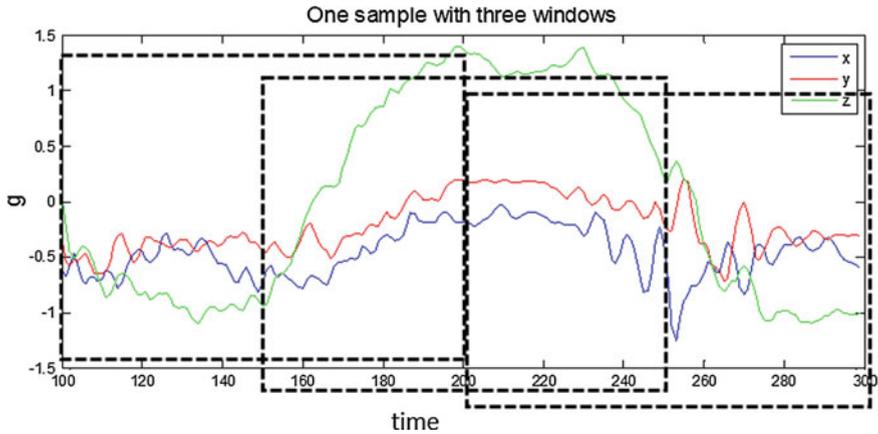
Fig. 3 One sample with three windows

3.1 Data Pre-processing

The acceleration data collected from the triaxial accelerometer is sampled at the rate of 100 Hz. And in this paper, the duration of one motion is defined as two seconds. Since the real motions are continuous and long series, to cut it into segments is necessary. As a result, three windows are set to cover one whole motion, shown in Fig. 3. The overlap of the windows is half of the width [7]. Using overlapping windows provides some information redundancy as the same samples will be processed multiple times into the result.

3.2 Feature Extraction

Features of samples are essential to classification. Representative features are selected to describe the character of the samples that can differ from others. In this paper, seven effective features are extracted, consisting of mean, correlation, energy, variance, minimum value, interquartile and peaks. Detailed information follows below: (1) Mean: the mean value of the acceleration data over one window. (2) Correlation: the relationship between three axes [8]. (3) Energy: energy from FFT (Fast Fourier Transform). Information in frequency domain always has sharp details and can be quite unique according to specific motion. (4) Minimum value: the minimum value of energy of three axes. When the motion is about flipping over or any rotating movement, the energy is different among three axes. (5) Variance: the variance over three axes in time domain. (6) Interquartile range: how data spreads out from median can be a good character. (7) Peaks: different motions are always have different amount of peaks in frequency domain.

Since the triaxial accelerometer provides data of three axes, and for each axis, three windows are computed. Consequently, a total features of $57(19 \times 3)$ are obtained.

3.3   Feature Pre-processing

Before training the C-SVC classifier, it is quite necessary to apply scaling processing [9]. The advantage of scaling is to avoid elements with greater numeric ranges dominating those in smaller ranges. Because the value of the kernel function depends on the inner products of feature vectors. Greater values can cause numeric problems.

In this paper, the scaling strategy is to narrow down the range of the feature values into $[-1,1]$. On account of rapid processing, a maximum number from the whole feature matrix (including training and testing features) is chosen as the scaling factor, in Eq. (4), where **A** and **B** are training and testing feature matrix. The experimental result shows how it works efficiently in obtaining the classification result.

$$scalingNum = \max\left\{abs\left(\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}\right)\right\} \qquad (4)$$

3.4   Classifier Training

After feature pre-processing, the scaled feature matrix is passed to train the C-SVC and construct a model with a suitable C and γ. C and γ are the critical parameters of the model that an optimized pair of them is required. As a result, a 3-fold cross-validation method is applied to optimize the parameters.

A threshold-based grid-search iteration will be the strategy under 3-fold cross-validation. The n-fold cross-validation [10] means the training samples will be divided into n subsets of equal size. And one subset is tested with a model that is trained by the rest $n - 1$ subsets. Such a cross validation method is able to avoid the overfitting problem in case that the number of the samples is not large enough. In the grid-search iteration, the parameters C and γ range from negative to positive and for each pair of C and γ, a cross-validation is run to output a total accuracy. When the threshold or the running times is reached, the iteration terminates and the optimized parameter is obtained. The experimental result shows the running time of proposed grid-search is less than traditional one with greedy search that have to finish all the possible values in the grid.

4 Experimental Results

In our experiment, all the acceleration data was collected from the ADXL362Z-DB triaxial accelerometer which is manufactured by Analog Device corporation, as is shown is Fig. 4. It can sense the acceleration from −8 g to 8 g. All the data is sampled at a rate of 100 Hz and processed on Windows 10 operating system.

In this paper, three wrist motions are defined: static motion, flip over motion and swing motion. Samples of these motions are shown in Fig. 5. And 621 samples are obtained.

The result shows that the C-SVC can perform efficiently with scaling, as is shown in Table 1. Without scaling, the accuracy of the C-SVC is 67.1498 % (621 samples). The accuracy with proposed scaling method comes to 94.2029 %.

Fig. 4 ADXL362Z-DB triaxial accelerometer
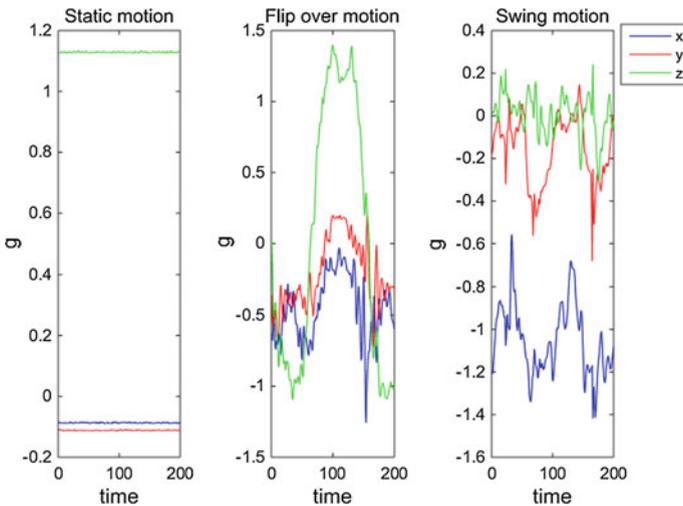
Fig. 5 Three defined motions

Table 1 Confusion matrix without/with scaling (621 samples)

Prediction	Static	Flipover	Swing	Prediction	Static	Flipover	Swing
Static	207	0	0	Static	207	0	0
Flipover	0	207	0	Flipover	20	179	8
Swing	0	204	3	Swing	0	8	199
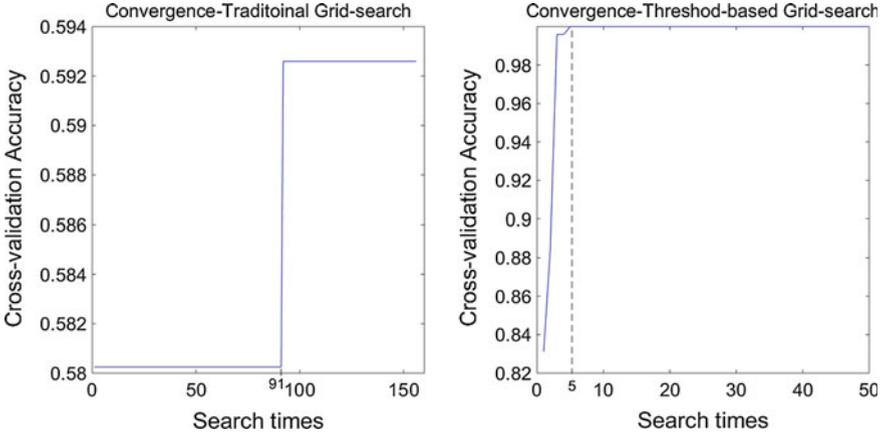
Fig. 6 Convergence in different grid-searchs

We can find that almost all the swing motions (204 in 207 total) are classified as flip over motions when no scaling is applied. That is because both motions have great energy which is considered a dominant feature among all the features.

The reason that scaling contributes a lot is all about the great numerical differences among features. For instance, the scaling factor is $1.7307e + 04$ for energy feature while less than 0.001 for some other features. Particularly, the test samples are contaminated by some disturbance to validate the robustness. The result shows that the C-SVC classifier proposed in this paper can achieves higher accuracy in motion detection. Convergence in our grid-search is faster than the traditional one, shown in Fig. 6. The threshold-based method can achieve a very high cross-validation accuracy (approximately 100 %) while the traditional one can only achieve a lower accuracy with much more running time. And note that no matter the features are scaled or not, the static motions are always recognized accurately.

5   Conclusions

In this paper, an improved SVM classifier is presented, including overlapping multi-windowing in one sample, representative features extraction, feature scaling and threshold-based grid-search optimization under 3-cross validation for training.

Experimental result shows a high accuracy (94.2029 %) in recognizing motions and better convergence ability (converge in 5 times). With this method, any devices with accelerometer can function as a motion monitor. In the future, we plan to implement this method in mobile device and develop a real time motion detection method.

References

1. Hsieh, S.L., Chen, C.C., Wu, S.H., Yue, T.W.: A wrist-worn fall detection system using accelerometers and gyroscopes. IEEE Tran. Networking Sens. Control **11**, 518–523 (2014)
2. Fuentes, D., Gonzalez-Abril, L., Angulo, C., Ortega.: Online motion recognition using an accelerometer in a mobile device. Expert. Syst. Appl. **39**(3), 2461–2465 (2012)
3. Cortes, V.: Support-vector network. Mach. Learn. **20**, 273–297 (1995)
4. Yang, J.Y., Wang, J.S., Chen, Y.P.: Using acceleration measurements for activity recognition: an effective learning algorithm for constructing neural classifiers. Pattern Recogn. Lett. **29**(16), 2213–2220 (2008)
5. Lugade, V., Fortune, E., Morrow, M., Kaufman, K.: Validity of using tri-axial accelerometers to measure human movement Part I: posture and movement detection. Med. Eng. Phys. **36**(2), 169–176 (2014)
6. Bakshi, A., Kopparapu, S.K., Pawar, S., Nema, S.: Novel windowing technique of MFCC for speaker identification with modified polynomial classifiers. IEEE Tran. Confluence **5**, 292–297 (2014)
7. Parhi, K., Ayinala, M.: Low-complexity welch power spectral density computation. IEEE Tran. Circ. Syst. I Regul. Pap. **61**(1), 172–182 (2014)
8. Xiao, P., Qu, W., Qi, H., Li, Z.: Detecting DDoS attacks against data center with correlation analysis. Comput. Commun. **67**, 66–74 (2015)
9. Grandvalet, Y., Canu, S.: Adaptive scaling for feature selection in SVMs. Adva. Neur. Info. Proc. Syst. **11**, 553–560 (2002)
10. Li, Q., Salman, R., Test, E., Strack, R., Kecman, V.: Parallel multitask cross validation for support vector machine using GPU. J. Parallel. Distr. Com. **73**(3), 293–302 (2013)

Analysis and Estimation of Vertical Beam Characteristic Parameters for Phased Array Radar

Long Xiang, Jian-Jiang Ding, Xiao-Wei Shi and Zhi-Fang Zuo

Abstract Vertical Coverage Pattern (VCP) is an important index in radar engineering design, which determines the performance and parameters of the radar. Based on the VCP of phased array radar, the variation and characteristic parameters of a vertical beam are analyzed deeply. In this paper work, we discuss a prediction method and propose estimation formulas for the characteristic parameters. We estimate the unknown parameters of a specific radar and its VCP by comparative analysis of two typical radars as a case study. The results of this analysis provide a novel method for new radar jamming technology, offering better theoretical significance and reference value in the practice of engineering.

Keywords Phased array radar · VCP · Parameter estimation · Vertical beam

1 Introduction

Phased array radar (PAR) has the merit of beam agility, waveform agility, and flexible operation patterns. As such, it is difficult to locate and jam PAR effectively. PAR has always been a research focus and difficult problem in the field of radar ECM (Electronic Counter Measure) [1–5]. Here, we quantitatively analyzed the vertical beam parameters based on the VCP of PAR. The estimation and prediction to the detection capability of radar has also been made by VCP simulation.

L. Xiang (✉) · J.-J. Ding · Z.-F. Zuo
Air Force Early Warning Academy, Wuhan, Hubei, China
e-mail: dick_500@163.com

J.-J. Ding
e-mail: djj1965@163.com

Z.-F. Zuo
e-mail: zzf1982@163.com

X.-W. Shi
The 29th Research Institute of CETC, Chengdu, Sichuan, China
e-mail: sxw1979@163.com

2   Beam Characteristic of BES-PAR

BES (Beam Elevation Scanning)-PAR denotes specifically the 3D air surveillance phased array radars with electronic scan in elevation (vertical) and mechanical rotation in azimuth (horizontal). The width and shape of the azimuth beam is usually fixed. The width and shape are determined by the technical index of BES-PAR. However, the vertical beam is raster scanning and can form multiple beams within VCP [6].

The multiple beams in vertical plane have some remarkable characteristics, and their values mainly depend on the selection of the operation mode and working parameters of PAR, which also embody the flexibility of PAR operation. The characteristics of BES-PAR can be described with the following formula:

$$C_{eb,i} = \left\{ \beta_{e,i}, \theta_{e,i}, n_{t,i}\tau_{t,i}, f_{0,i}, t_{r,i}, p_{t,i} \right\}(i = 1, 2, \ldots, N_e) \tag{1}$$

Here, $N_e, \beta_{e,i}, \theta_{e,i}, n_{t,i}\tau_{t,i}, f_{0,i}, t_{r,i}, p_{t,i}$ denotes the number of beams in vertical plane, beam direction angle, beam width, number of transmitted pulses in each beam, carried frequency, PRF (Pulse Repetition Frequency) and pulse power, respectively.

3   Analyses and Estimation to the Characteristic Parameter of BES-PAR

Since the vertical beams of BES-PAR are varied, radar ECM cannot obtain accurate and overall information about radar. However, there is another way to analyze and estimate these parameters of the BES-PAR. The general law can be explored and predicted in this paper.

3.1   Full-Range Elevation Sector Width

For BES-PAR, under the Full-range Elevation Sector (FES), the transmitting beam can reach or exceed the radar maximum detection range. The approximate calculation formula for FES width is [7]:

$$\theta_1 = \arcsin\left(\frac{H_{\max}}{R_{\max}} - \frac{R_{\max}}{2k_a}\right) \tag{2}$$

Here, K_a denotes the effective radius of the Earth (approximately, 8500 km), $H_{\max}$ denotes the maximum height, and $R_{\max}$ denotes the maximum range.

3.2 Beam Width and Beam Direction Angle of FESB

The beams in FES are called Full-range Elevation Sector Beam (FESB); we need to calculate the beam width and direction angle of FESB. Firstly, the definition of the first full-range elevation sector beam (FFESB) is as follows: the beam which has the lowest beam pointing angle in the FES. The beam width of FESB is calculated by the following formula:

$$\theta_{e1} = \frac{\theta_1}{n_1 \cdot \Delta_e} \tag{3}$$

In formula (3), n_1 denotes the number of beams which are used to cover the FES of $[0, \theta_1]$, the value of n_1 is also required to provide the desired resolution, or aperture height. Δ_e denotes the beam spacing between those elevation beams. Thus, the beam pointing angle of the other FESBs can be calculated by this formula:

$$\beta_{ei} = \beta_{e1} + (i - 1) \cdot \Delta_e \cdot \theta_{e1} \tag{4}$$

In formula (4), β_{ei} denotes the pointing angle of the ith FESB; β_{e1} denotes the pointing angle of FFESB; other parameters are same as formula (3).

3.3 Beam Width and Beam Direction Angle of UES

Above the FES, the region is called Upper Elevation Sector (UES). The scope of UES is from θ_1 to the max elevation angle (θ_{max}). Therefore, the whole width of UES is:

$$\theta_{e2} = \theta_{max} - \theta_1 \tag{5}$$

If we assume the number of the beam covering the whole UES to be n_2, which is also called Upper Elevation Sector Beam (UESB), each UESB's width can be calculate by an iteration method [8]. The iterative formula is given as:

$$\frac{\theta_{ehj}}{\theta_{e1}} = \frac{\sin(\beta_{ehj})}{\sin\theta_1}, \quad (j = 1, \ldots, n_2) \tag{6}$$

$$\beta_{ehj} = \beta_{en_1} + \Delta_e \cdot \sum_{k=1}^{j} \theta_{ehk} \tag{7}$$

In formula (6), θ_{ehj} is the beam width of jth UESB; in formula (7), β_{en_1} denotes the pointing angle of n_1th FESB, which can be deduced from formula (4). Thus, each UESB's pointing angle can be calculated in the following formula:

$$\beta_{ej} = \beta_{en_1} + j \cdot \Delta_e \quad (j = 1, \ldots, n_2) \tag{8}$$

3.4 Constraint Relationship Between ESST and ADT

Elevation Sector Scanning Time (ESST). For radar's FESB, the time needed to scan the whole elevation sector is called Elevation Sector Scanning Time (ESST), noted by t_{es}. Based on the above analysis, the calculation formula of t_{es} is:

$$t_{es} = \sum_{i=1}^{n_1+n_2} (\tau_{t,i} + t_{r,i}) n_{t,i} \tag{9}$$

The definitions of parameters in formula (9) are given in formula (1).
Azimuth Dwell Time (ADT). The definition of Azimuth Dwell Time (ADT) is given as:

$$t_a = \frac{t_s \cdot \Delta_a \cdot \theta_a}{A_m} \tag{10}$$

In formula (10), A_m denotes the azimuth sector width; t_s denotes the radar antenna scanning period; θ_a denotes the azimuth beam width; Δ_a denotes the spacing between adjacent azimuth beams. The relationship formula of them is given by:

$$t_{es} < t_a \tag{11}$$

4 A Case Study to Typical BES-PAR

In this case study we will examine TPS-59 radar (TPS-59), designed by Lockheed-Martin Company, as a case of BES-PAR. TPS-59's specification parameters and VCP can be found in Ref. [9]. Its derivative product is TPS-117 radar (TPS-117). The unknown specification parameters and VCP of TPS-117 can be estimated, by means of comparative analysis of the two radars' specifications and using the discussed in this paper.

4.1 Arrangement of Beam Positions and Pulse Number in FES & UES

TPS-59 arranges three FESBs in its FES ($\theta_1 = 1.7°$), and arranges another five UESBs in the UES. According to the formulas (3–8), it can be concluded that TPS-117 will arrange four FESBs in FES ($\theta_1 = 2.2°$). The relative parameters can be obtained by MATLAB's numerical calculation [10], as shown in Table 1.

TPS-59 transmits 11 short-range beams in the short-range elevation sector and places a three-pulse sequence with PRF of 625 Hz in the three low elevation sectors for MTI processing. The estimated number of beams is about 9–10, enough to cover the short-range elevation sector for TPS-117. It's also assumed that TPS-117 is similar to TPS-59 for MTI with PRF of 1100 Hz. So the number of short-range

Table 1 Beam position of TPS-117 in VCP

Secor name	Beam point angle (°)	Half-power down point (°)	Half-power up point (°)
Full-range elevation sector I	0.344	0	0.688
Full-range elevation sector II	0.983	0.55	1.237
Full-range elevation sector III	1.444	1.1	1.788
Full-range elevation sector IV	1.993	1.65	2.337
Upper elevation sector I	2.793	2.293	3.293
Upper elevation sector II	3.914	3.214	4.614
Upper elevation sector III	5.482	4.502	6.462
Upper elevation sector IV	7.072	6.078	8.066
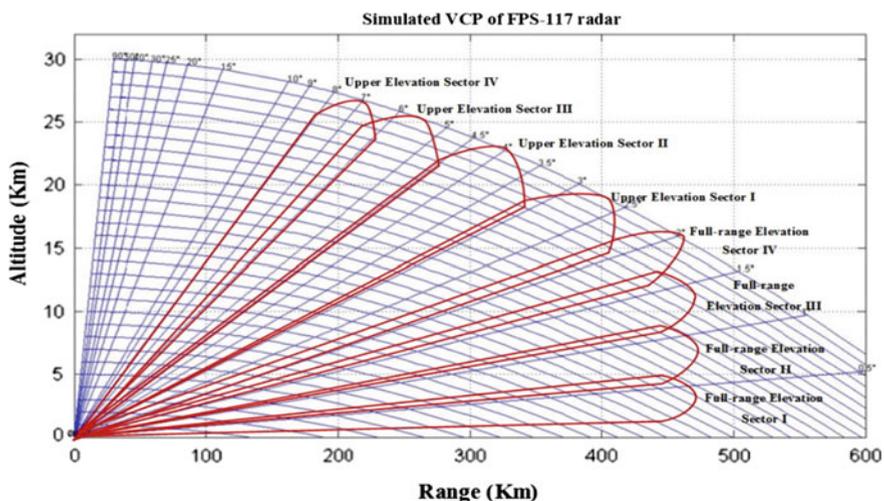
Fig. 1 VCP of TPS-117 (in FES)
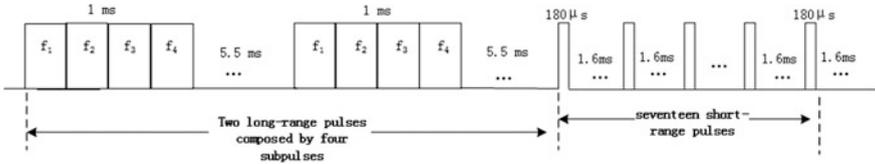
Fig. 2 Transmitting pulses train of TPS-59
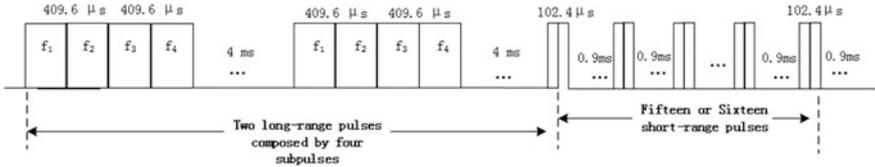
Fig. 3 Transmitting pulses train of TPS-117

pulses for TPS-117 is about 15–16. Based on the results of the calculations above, we get the VCP of TPS-117, as shown in Fig. 1.

4.2 Calculation of ESST and ADT

The transmitting pulses train of TPS-59 is shown in Fig. 2, the transmitting pulses train of TPS-117 is shown in Fig. 3, the ESST of TPS-117 can be calculated as $t_{es-117} = 24.67 - 25.67$ ms, and the ADT is $t_{a-117} = 61.1$ ms, it also has $t_{es-117} < t_{a-117}$.

5 Conclusion

In this paper, the vertical beam characteristics and parameters of BES-PAR have been analyzed and researched. The exploratory idea of digging the performance parameters and operation mode of an unknown radar, provides a new method for the research and development of 'adaptive, precise, intelligent' jamming technology.

References

1. Chen, C.X., He, M.H., Li, H.F.: An improved radar emitter recognition method based on D-S theory. Chin. J. Electron. **24**(3), 611–615 (2015)
2. Petrov, N., Jordanov, I., Roe, J.: Radar emitter signals recognition and classification with feed forward networks. Comput. Sci. **22**(5), 1192–1200 (2014)
3. Liang, H.D., Han, J.H.: Clustering and sorting radar signals based on multi-wavelet packets characteristics of bispectrum. Acta Photonica Sinica **43**(3), 1–7 (2014)
4. Focke, R.W., de Villiers, J.P., Inggs, M.R.: Interval algebra—an effective means of scheduling surveillance radar networks. Inf. Fusion **23**(2), 81–98 (2015)
5. Yang, D.: Technology researches on airborne fire control radar emitter identification. Ph.D. thesis, Nanjing University of Aeronautics and Astronautics, Nanjing, China (2014)
6. David, K.B.: Radar System Analysis and Modeling. Artech House, London (2005)
7. Wang, X.M., Kuang, Y.S., Chen, Z.X.: Surveillance Radar Technology. Publishing House of Electronics Industry, Beijing (2008)
8. Stewart, G.T., Zheng, Y.L., Wong, W.K.: Machine learning for activity recognition: hip versus wrist data. Physiol. Meas. **35**(11), 2183–2189 (2014)
9. Ground-Based Air Surveillance Radars. http://www.lockheedmartin.com/us/products/ground-based-air-surveillance.html
10. Bassem, R.M.: MATLAB Simulations for Radar Systems Design. Chapman & Hall/CRC Press, NewYork (2006)

Blind Image Quality Assessment Based on Mutual Information

Bao Liu, Li-Xiong Liu, Hong-Ping Dong and Yong-Gang Lin

Abstract Mutual information can indicate the degradation severity of natural images by capturing images' local structural properties. In this paper, we analyze the dependence between neighboring pixels and propose a no reference image quality assessment method based on mutual information in wavelet domain. The proposed image quality assessment (IQA) method, named MIQA-II, computes mutual information between neighboring pixels in the same subband, across different orientations and scales as features. A quadratic function is used to fit the mutual information values along a distance and the features are then transformed to a final predicted quality score through a two-step framework. MIQA-II is tested on the LIVE IQA database. The experimental results show that MIQA-II has a competitive subjective relevance performance and acceptable time consumption. In addition, MIQA-II still achieves good performance with a small training set.

Keywords Image quality assessment · Mutual information · Wavelet

1 Introduction

The amount of Internet data grows rapidly and big data is becoming a new research hotspot. In the big data, images and videos occupy the largest proportion. However, there exist masses of distortions in the image and video data. Visual information

B. Liu (✉) · L.-X. Liu · H.-P. Dong · Y.-G. Lin
School of Computer Science and Technology, Beijing Institute of Technology,
Beijing, China
e-mail: liubao@bit.edu.cn

L.-X. Liu
e-mail: lxliu@bit.edu.cn

H.-P. Dong
e-mail: hiyuki@bit.edu.cn

Y.-G. Lin
e-mail: bitailin@bit.edu.cn

degradation commonly occurs during the process of acquisition, compression, transmission, decompression, storing and displaying, leading to difficulties for human to obtain and understand the visual contents. Therefore, it is important to research in objective image quality assessment (IQA) [1, 2].

According the dependence on original image, IQA methods can be divided into three categories: full reference (FR), reduced reference (RR) and no reference (NR). In recent years, many researchers have devoted to exploring efficient visual quality assessment methods, yielding a number of outstanding IQA methods such as the FR methods SSIM [3] and SR-SIM [4] and NR methods DIIVINE [5], CBIQ [6], NSS-GS/NSS-TS [7], Q metric [8], PIPs-based BIQA [9] and BRISQUE [10]. Further, Saha et al. [11] proposed a new NR method, requiring no training. Saliency based image quality assessment is being researched in [12] and then extended to video saliency detection [13, 14]. Information theory approach also has been applied into visual quality assessment and achieved good results. Sheikh et al. [15] explored the relationship between image information and visual quality. Saad et al. [16] computed Renyi entropy to represent the anisotropy in BLIINDS. Hu et al. [17] proposed a function based on the product of entropy and contrast to predict image quality. Very recently, we proposed an NR IQA method based on NSS in curvelet domain [18], and used spatial and spectral entropies to conduct an NR IQA method named SSEQ [19].

Further, we have done some previous work on exploring the relationship between visual quality and pixels' mutual information, i.e., MIQA-I, which is a primary version of the method proposed in this paper [20]. MIQA-I used the original image and its corresponding normalized luminance map and local standard deviation map as inputs. Self-correlated mutual information between neighboring pixels was computed to quantify the dependence between neighboring pixels. Compared with MIQA-II, MIQA-I only considers the mutual information at distance 1 and does not fit the mutual information along distance. In addition, MIQA-II further analyzes the mutual information between different scales and orientations and only uses original image as the input data. In this paper, we study this relationship and propose a general-purpose NR IQA method called MIQA-II which utilizes mutual information within a wavelet subband and across different subbands as features. A quadratic function is used to fit the mutual information values along distance and then the features are transformed to final predicted quality score through a two-step framework [21]. Experimental results show that our method correlates well with subjective opinions. MIQA-II shows a advantage especially when the size of the training set is relatively small. The advantage is precious, since it is difficult to obtain sufficient training data.

The rest of the paper is organized as follows. In Sect. 2 we give the theory foundation of our method, describe the flow of our algorithm and detail the features extracted from the Wavelet domain. In Sect. 3, we evaluate the performance of the proposed algorithm and give the corresponding results and discussions. Finally, we conclude the paper in Sect. 4.

2 The Proposed Method

2.1 Mutual Information

Natural images are highly structured [3]. That is to say, there are redundancies between neighboring pixels in an image, yielding strong dependence between neighboring pixels. The image degradation however will destroy the statistical regularity of this dependence. For example, blurring effect will equal the values of local pixels, increasing the dependence. On the contrary, noisy effect will add high frequency signals to the image, destroy the structural properties and reduce the dependence.

Our basic hypothesis is that mutual information can capture images' local structural properties, i.e., dependence between neighboring pixels reflects images' distortion level. Ruderman [22] computed Shannon's mutual information between the two pixel values as below,

$$MI = \int p(\varphi_1, \varphi_2) \log \frac{p(\varphi_1, \varphi_2)}{p(\varphi_1)p(\varphi_2)} d\varphi_1 d\varphi_2, \tag{1}$$

where φ_1 and φ_2 are two pixels at a given separation in an image, $p(\varphi_1, \varphi_2)$ is the joint distribution of two pixels, and $p(\varphi_1)$ is the marginal distribution of a pixel [22]. According to Eq. 1, we find that MI is based on the joint distribution of two corresponding pixels and is a measure of the dependence between two pixels. Ruderman [22] also found that neighboring pixels in natural images have a certain statistical regularity and that the mutual information and the separation distance between two pixels obey a power-law below,

$$I(d, \theta) \approx d^{-\alpha(\theta)}, \tag{2}$$

where d is the separation distance between two pixels, θ is the angle of the separation axis, $\alpha(\theta)$ is a function of θ and $I(d, \theta)$ is the mutual information of two pixels with separation distance d and angle θ.

Ruderman revealed that the statistical property of the dependence between neighboring pixels is based on an immense ensemble of images. We believe that for one single natural image, this statistical property also exists if we regard this single natural image as the ensemble of many "tiny" natural image patches. In order to describe mutual information's ability to identify distortion, we plotted the self-correlated mutual information distribution curves for five distortion versions of a natural image (see Fig. 1), and observe that different distortions change the self-correlated mutual information in their own ways and hence their distribution curves are varied. This suggests that structural statistics property exists in the correlations between neighboring pixels and a quadratic distribution may be chosen to fit the mutual information values along distance. The quadratic distribution form is shown below,
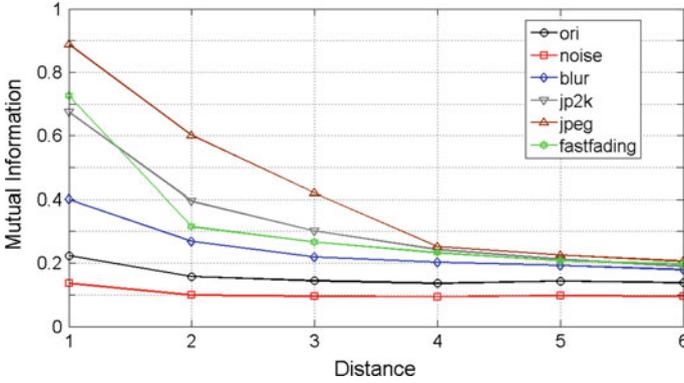
Fig. 1 Mutual information values of different distances for five distortion types

$$MI = ad^2 + bd + c, \tag{3}$$

where MI is the mutual information value at distance d, and a, b and c are fitted parameters which will be treated as features.

In our method, we employ a quadratic distribution to fit the mutual information value between pixels at the same orientation within a wavelet subband and further analyze the mutual information of pixels at the same location of different subbands. In addition, the mutual information across different orientations and scales are also considered. The process of our method can roughly be divided into three steps: (1) transforming the tested image using discrete wavelet transform. The wavelet used here contains 2 scales and 3 orientations. (2) Extracting mutual information features from the wavelet subbands. This step is the key step of our method and will be detailed below. (3) Predicting image's quality score using a two-step framework [21].

2.2 Feature Extraction

The discrete wavelet transform has been the preferred choice for image processing and analysis. It divides the information of an image into approximate and detailed subbands. Wavelets provide a mathematical way of encoding image information in an efficient and effective way that it is layered according to level of detail [23]. This layered structure not only represents images at multi-resolutions, but also helps to analyze images' properties in both spatial and frequency domains. So we preprocessed input images with discrete wavelet decomposition before feature extraction. The feature vector F considered here can be divided into three groups: fitted parameters of mutual information between two pixels within each subband, mutual information between two pixels across different scales and orientations.

Mutual Information between Two Pixels within a Subband (f1–f24). The two-scale wavelet decomposition first produced 6 subbands in 3 orientations (i.e., the horizontal, the vertical and the main diagonal). We then computed two separated pixels' mutual information in each subband. The locations of these two pixels are (i, j) and $(i, j + d)$ in the horizontal subband, (i, j) and $(i + d, j)$ in the vertical subband and (i, j) and $(i + d, j + d)$ in the main subband respectively. According to the finding as is seen in section introduction, we choose quadratic distribution to fit the mutual information along distance (see Eq. 3). The distance d between two pixels ranges from 1 to 3. In addition, we add a full orientation value by application of a filtering operator H with distance parameter d to image. Then we compute the fitted parameters of mutual information between the central pixel and its full orientation value similarly.

Thus, we get 3 mutual information values for each orientation (including the full orientation). These 3 values are fitted by a quadratic distribution and the 3 fitted parameters (a, b and c) are used (see Eq. 3) as features, yielding a 24-dimensional feature vector (2 scales × 4 orientations × 3 fitted parameters).

Mutual Information between Two Pixels across Different Scales (f25–f28). Two pixels at the same location of two subbands correspond to the same original pixel in the uncompressed image. Therefore, there exists strong redundancy and relevance between them and we can utilize mutual information to model this structure information. The size of subbands at 2 different scales are different, so we resize the low-scale subband from size M × N to 2 M × 2 N to match the sizes of different scales. In detail, the pixel value of low-scale subband at location (i, j) is used to fill the pixel value of the resized subband at location $(2i, 2j)$, $(2i + 1, 2j)$, $(2i, 2j + 1)$ and $(2i + 1, 2j + 1)$. Then in order to capture the dependence of two corresponding pixels across different scales, we compute the mutual information of two pixels at the same location in two subbands of the same orientation but different scales. Thus we get 4 features.

Mutual Information between Two Pixels across Different Orientations (f29–f34). We compute the mutual information of two corresponding pixels across different orientations, i.e., we compute the mutual information of pixels that locate in two subbands of the same scale but different orientations. In detail, we compute the mutual information between the horizontal and the vertical, the horizontal and the diagonal and the vertical and the diagonal. The locations of the two corresponding pixels are both (i, j). So we get another 6 features (2 scales × C_2^3).

As a conclusion, we get a 34-dimensional feature vector F. Table 1 lists the brief description of the feature vector F.

Table 1 Summary of extracting features

Features	Feature description
f1–f24	Dependence between pixels within each subband
f25–f28	Dependence between pixels across different scales
f29–f34	Dependence between pixels across different orientations

3   Evaluation and Results

We conducted a series of evaluation experiments on the well-known LIVE IQA database [24]. The performance indexes of correlation with human opinions used here consist of the Spearmen's Rank Order Correlation Coefficient (SROCC), the Kendall's Rank Order Correlation Coefficient (KROCC), the Pearson's (Linear) Correlation Coefficient (LCC) and the Root Mean Squared Error (RMSE) between the objective quality scores (predicted DMOS) and the subjective quality scores. A value close to 1 for SROCC, KROCC and LCC and a value close to 0 for RMSE indicate good performance in terms of correlation with human opinion. In our implementation, we repeat 1000 times training and test experiment for each NR IQA method to eliminate performance bias. The partition between training and test sets is random and do not overlap on content for each time. The training set occupies 80 % of the reference images and corresponding distorted counterparts while the rest data set composes the test set. The median values of each performance index are used as the final results.

3.1   Correlation with Human Opinions

We compared our method with other popular FR and NR IQA methods such as PSNR, DIIVINE, BRISQUE and so on.

Table 2 summarizes the median evaluation results of each FR and NR IQA methods. The italic ones correspond to FR methods, and others are NR methods. LCC and RMSE are computed after passing the algorithm scores through a logistic nonlinearity as described in [25].

The results show that our method is superior to other IQA methods except FR IQA approach VIF and indicate that there exist a strong relationship between pixels' mutual information and visual quality. What is more, we can see that MIQA-II is superior to MIQA-I, which illustrates that the wavelet transform has helped on improving the performance.

Table 2 Median evaluation results of FR/NR IQA methods on the LIVE IQA database

	SROCC	KROCC	LCC	RMSE
PSNR	*0.8756*	*0.6865*	*0.8723*	*13.3597*
SSIM	*0.9104*	*0.7311*	*0.9042*	*11.6694*
VIF	*0.9636*	*0.8282*	*0.9604*	*7.6137*
SR-SIM	0.9618	0.8299	0.9553	8.0811
BIQI	0.7343	0.5533	0.7589	20.3678
CBIQ	0.8936	0.7144	0.8953	12.7790
DIIVINE	0.9162	0.7617	0.9206	12.2138
SSEQ	0.9344	0.7953	0.9417	10.5052
MIQA-I	0.9304	0.7853	0.9389	10.7744
BRISQUE	0.9385	0.8018	0.9473	10.0070
MIQA-II	0.9483	0.8195	0.9539	9.4137

Italic denotes FR methods

3.2 Variation of Performance with the Number of Training Images

The NR IQA methods mentioned here are all training-based. In practice, however, it is not easy to get enough training images and the labeling work on the training images is a costly and time-consuming work. So it is necessary to have a stable and high performance with a relative small training set. We present our experiment results for 4 NR methods with different training images in Fig. 2. The x-axis and y-axis represent the number of training reference images and the median SROCC values of 1000 experiments, respectively. There are 29 reference images in the LIVE IQA database in total, if n reference images and their distorted counterparts are selected for training, the rest images are collected for testing. The results show that our method has a great advantage against other 3 NR IQA methods with different training set scales. In addition, our method still keeps a relatively higher subjective correlation than other methods with only 5 training reference images.

3.3 Statistical Significance Testing

Furthermore, to compare the statistical performance of the above NR IQA methods, we conducted an experiment based on t-test [26] on the SROCC values across 1000 training and test trails. The confidence level is set to 95 %. The detailed results are shown in Table 3.

Table 3 shows that our method is statistical superior to all other methods on the SROCC values across 1000 trails. So we can replace the other methods with our method in IQA applications without any performance loss.
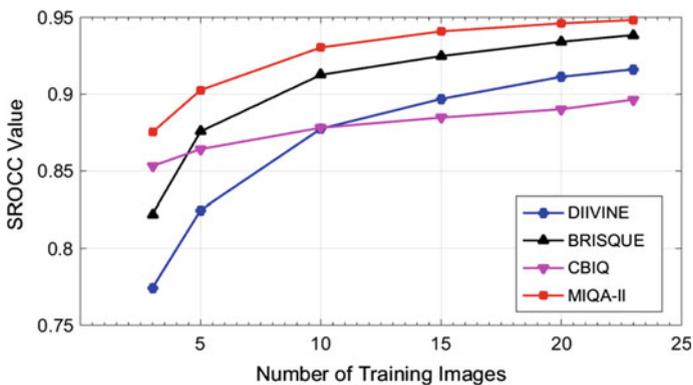
Fig. 2 Median SROCC values of several NR methods with different number of training images

Table 3 The result of T-test on the SROCC values across 1000 trails

	BIQI	CBIQ	DIIVINE	SSEQ	MIQA-I	BRISQUE	MIQA-II
BIQI	0	−1	−1	−1	−1	−1	−1
CBIQ	1	0	−1	−1	−1	−1	−1
DIIVINE	1	1	0	−1	−1	−1	−1
SSEQ	1	1	1	0	1	−1	−1
MIQA-I	1	1	1	−1	0	−1	−1
BRISQUE	1	1	1	1	1	0	−1
MIQA-II	1	1	1	1	1	1	0

A value of "1" indicates that the row method is statistical superior to the column method, a value of "−1" indicates that the row method is statistical worse than the column method and a value of "0" indicates that there is no statistical difference between row and column methods at the 95 % confidence level.

3.4 Performance on the TID2013 Database

We conducted a similar 1000 times 80 % training and 20 % test trails on the TID2013 Database. We selected 4 similar types of distortions with LIVE from the TID2013 Database to compare the performances of BIQI, MIQA-I, BRISQUE and our method (Table 4). It is also shown that our method performs better than BRISQUE even on the different database and with different natural scenes.

3.5 Computational Complexity

As an auxiliary tool to be integrated into the practical application, computational complexity is an important factor. We conducted a comparison on the time complexity performance (only consider the time of feature extracting). The test image is the "bikes.bmp" with a size of 768 × 512 in LIVE IQA database. The results are summarized in Table 5 and show that our algorithm consumes less time than DIIVINE and CBIQ. Note that our Matlab code has not been optimized.

Table 4 Median evaluation results of NR IQA methods on the TID2013 database

	SROCC	KROCC	LCC	RMSE
BIQI	0.7773	0.5855	0.7962	0.8440
MIQA-I	0.9040	0.7391	0.9256	0.5291
BRISQUE	0.8976	0.7301	0.9216	0.5448
MIQA-II	0.9140	0.7550	0.9339	0.5007

Table 5 Computational
complexity of several
NR IQA methods

	Time (s)
CBIQ	73.35
DIIVINE	25.40
BRISQUE	0.142
MIQA-II	1.420

4 Conclusion

In this paper, we proposed MIQA-II, a blind IQA method based on mutual information in Wavelet domain. The major contributions of the paper are included below:

(1) MIQA-II assumes that statistical properties of natural images and the mutual information is a powerful tool to identify and quantify image distortion, and utilizes mutual information between neighboring pixels in the same subband, across different orientations and scales as features. (2) MIQA-II has high perceived consistency statistically and high robustness to the number of training images. (3) MIQA-II also has a relatively low computational complexity.

In the future, we will apply the idea of entropy and mutual information to video and 3D quality assessment.

References

1. Lin, W., Kuo, C.C.J.: Perceptual visual quality metrics: a survey. J. Vis. Commun. Image Represent. **22**(4), 297–312 (2011)
2. Gao, X., Lu, W., Tao, D., Li, X.: Image quality assessment and human visual system. In: Visual Communications and Image Processing 2010, vol. 7744, pp. 77440Z-1–77440Z-10. SPIE, Bellingham (2010)
3. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
4. Zhang, L., Li, H.: SR-SIM: a fast and high performance IQA index based on spectral residual. In: IEEE International Conference on Image Processing (ICIP), pp. 1473–1476. IEEE, Piscataway (2012)
5. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans. Image Process. **20**(12), 3350–3364 (2011)
6. Ye, P., Doermann, D.: No-reference image quality assessment based on visual codebook. In: IEEE International Conference on Image Processing (ICIP), pp. 3150–3153. IEEE, Piscataway (2011)
7. Gao, X., Gao, F., Tao, D., Li, X.: Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning. IEEE Trans. Neural Netw. Learn. Syst. **24**(12), 2013–2026 (2013)
8. Narwaria, M., Lin, W.: Objective image quality assessment based on support vector regression. IEEE Trans. Neural Netw. **21**(3), 515–519 (2010)

9. Gao, F., Tao, D., Gao, X., Li, X.: Learning to rank for blind image quality assessment. IEEE Trans. Neural Netw. Learn. Syst. **99**, 1–16 (2015)

10. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **1**(1), 4695–4708 (2012)

11. Saha, A., Wu, Q.M.J.: Utilizing image scales towards totally training free blind image quality assessment. IEEE Trans. Image Process. **24**(6), 1879–1892 (2015)

12. Tong, Y., Cheikh, F.A., Konik, H., Tremeau, A.: Full reference image quality assessment based on saliency map analysis. Int. J. Imaging Sci. Technol. **54**(3), 030503–030514 (2010)

13. Tong, Y., Cheikh, F.A., Guraya, F.F.E., Konik, H., Tremeau, A.: A spatiotemporal saliency model for video surveillance. J Cogn Comput. **3**(1), 241–263 (2011)

14. Tong, Y., Konik, H., Cheikh, F.A., Guraya, F.F.E., Tremeau. A.: Multi-Feature based visual saliency detection in surveillance video. In: Visual Communication and Image Processing, vol. 7744, pp. 7744041–7744049. SPIE, Bellingham (2010)

15. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. **15**(2), 430–444 (2006)

16. Saad, M.A., Bovik, A.C., Charrier, C.: A DCT statistics-based blind image quality index. IEEE Signal Process. Lett. **17**(6), 583–586 (2010)

17. Hu, Q., Xie, Z.X., Wang, Z.F., Liu, Y.H.: Constructing NR-IQA function based on product of information entropy and contrast. In: International Symposium on Information Science and Engineering. vol. 2, pp. 548–550. IEEE Computer Society, Washington DC (2008)

18. Liu, L., Dong, H., Huang, H., Bovik, A.C.: No-Reference image quality assessment in curvelet domain. Sig. Process. Image Commun. **29**(4), 494–505 (2014)

19. Liu, L., Liu, B., Huang, H., Bovik, A.C.: No-Reference image quality assessment based on spatial and spectral entropies. Sig. Process. Image Commun. **29**(8), 856–863 (2014)

20. Dong, H., Liu, L.: No reference image quality assessment in mutual information domain. J. Image Graphics **19**(3), 484–492 (2014)

21. Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. IEEE Signal Process. Lett. **17**(5), 513–516 (2010)

22. Ruderman, D.L.: The statistics of natural images. Netw. Comput. Neur. Syst. **5**(4), 517–548 (1994)

23. Talukder, K.H., Harada, K.: Haar wavelet based approach for image compression and quality assessment of compressed image. IAENG Int. J. Appl. Math. **36**(1), 49–56 (2007)

24. Sheikh, H.R., Wang, Z, Cormack L., Bovik, A.C.: LIVE Image Quality Assessment Database Release 2. http://live.ece.utexas.edu/research/quality

25. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. **15**(11), 3440–3451 (2006)

26. Sheskin, D.: Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall, London, U.K. (2004)

Compressive Tracking via Weighted Classification Boosted by Feature Selection

Li Zhang, Hanzhun Chen and Yuanyuan Hu

Abstract The drifting problem of object tracking is efficiently alleviated in our research. In this paper, an advanced compressive tracking algorithm based on a weighted classifier boosted by feature selection is proposed. The compressed features with high discrimination are selected from the target information of previous and current frames by a discrimination evaluating strategy. These discriminating features are used to train a weighted classifier, which is composed of two sub-classifiers based on previous and current samples bags. Finally, the weighted classifier is used to tell the target object from the background. Experimental results show that the performance in terms of accuracy and robustness hugely improves in tracking via the proposed classification method.

Keywords Object tracking · Compressed features · Weighted classifier

1 Introduction

Object tracking has been applied to surveillance, human computer interaction, medical imaging, etc. During past decades, numerous algorithms have been proposed in literature [1–7]. Object tracking, however, faces great challenges in the practical applications, such as illumination change, posture change, occlusion. Tracking algorithms can be generally categorized as either generative [4, 6] or discriminative [1, 2, 7] based on their appearance models. Generative tracking algorithms typically learn a model to represent the target object, and then use it to predict the object location with minimal reconstruction error in the subsequent

L. Zhang · H. Chen (✉) · Y. Hu
Institute of Information Engineering, Shenzhen University, Shenzhen, China
e-mail: 2130130411@szu.edu.cn

L. Zhang
e-mail: wzhangli@szu.edu.cn

Y. Hu
e-mail: huyyszu@szu.edu.cn

frames. Recently, Mei and Ling [6] proposed l_1-tracker, which is motivated by the idea that appearance of the target can be sparse represented in a template subspace. Discriminative tracking algorithms regard tracking as a binary classification problem. The target object is found by a combination of foreground and background information and separated from the background. Babenko et al. [7] introduced an online multiple instance learning (MIL) algorithm, which used multiple positive samples and negative ones to construct a classifier.

Zhang et al. [1, 2] proposed a real-time compressive tracking (CT) method. Although CT has performed effectively, it only uses positive samples of the current frame to update the appearance model. Therefore, it may lead to tracking failure or drifting problems due to the features extracted from the target being occluded by other objects. In order to improve the performance of CT tracker, an improved compressive tracking based on a weighted classifier boosted by feature selection is proposed. The compressed features that have a high discrimination act as an appearance model to update the classifier.

The rest of the paper is organized as follows. In Sect. 2, the original algorithm of Compressive Tracking is reviewed. In Sect. 3, our proposed algorithm is presented in detail. Experiments are shown in Sect. 4. At last, the conclusion is drawn in Sect. 5.

2   Related Work

In this section, a review over previous work in Compressive Tracking [1] is given.

2.1   Compressive Features Extraction

According to the compressed sensing theory [8], if the random measurement matrix $R \epsilon \boldsymbol{R}^{n \times m}$ satisfies restricted isometry property (RIP), the origin signal can be reconstructed with the minimal error, when the signal is projected from a high-dimensional space $x \epsilon \boldsymbol{R}^m$ to a lower-dimensional space $v \epsilon \boldsymbol{R}^n$, that is:

$$v = Rx \tag{1}$$

where $n \ll m$. In CT, a quite sparse random measurement matrix $R \epsilon \boldsymbol{R}^{n \times m}$ is adopted to efficiently extract the features for the appearance model. The sparse random measurement matrix is defined as in Eq. (2):

$$r_{i,j} = \sqrt{s} \times \begin{cases} 1 & \text{with probability } \frac{1}{2s} \\ 0 & \text{with probability } 1 - \frac{1}{2s} \\ -1 & \text{with probability } \frac{1}{2s} \end{cases} \tag{2}$$

where $r_{i,j}$ is the element in the matrix R, and s is set to $s = \frac{m}{4}$.

2.2 Classifier Construction and Online Update

In CT algorithm, a binary classifier is used to distinguish the target from the background. After dimensionality reduction, for each random sample $z \in \mathbf{R}^m$, the compressed features are represented as $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T \in \mathbf{R}^n$ with $n \ll m$. The way to update a naive Bayes classifier with these features is defined as following form:

$$
\begin{aligned}
H(\mathbf{v}) &= \log\left(\frac{\prod_{i=1}^{n} p(v_i|y=1)p(y=1)}{\prod_{i=1}^{n} p(v_i|y=0)p(y=0)}\right) \\
&= \sum_{i=1}^{n} \log\left(\frac{p(v_i|y=1)}{p(v_i|y=0)}\right) = \sum_{i=1}^{n} h(v_i)
\end{aligned}
\tag{3}
$$

where $y \in \{0, 1\}$ is a binary variable which represents the sample label. A uniform prior is assumed that $p(y=1) = p(y=0)$. $h(v_i)$ is a weak classifier, which is trained by the compressed features v_i.

According to the CT [1], the conditional distributions $p(v_i|y)$ in the classifier $H(\mathbf{v})$ are assumed to be Gaussian distributed with four parameters $(\mu_i^1, \sigma_i^1, \mu_i^0, \sigma_i^0)$, that is:

$$
p(v_i|y=1) \sim N\left(\mu_i^1, \sigma_i^1\right), p(v_i|y=0) \sim N\left(\mu_i^0, \sigma_i^0\right)
\tag{4}
$$

where μ_i^1 and σ_i^1 are the mean and variance of positive samples, μ_i^0 and σ_i^0 are the mean and variance of negative samples. The scalar parameters in Eq. (4) are incrementally updated

$$
\begin{aligned}
\mu_i^1 &= \lambda\mu_i^1 + (1-\lambda)\mu^1, \sigma_i^1 = \sqrt{\lambda(\sigma_i^1)^2 + (1-\lambda)(\sigma^1)^2 + \lambda(1-\lambda)(\mu_i^1 - \mu^1)^2} \\
\mu^1 &= \frac{1}{n}\sum_{k=0|y=1}^{n-1} v_i(k), \sigma^1 = \sqrt{\frac{1}{n}\sum_{k=0|y=1}^{n-1} (v_i(k) - \mu^1)^2}
\end{aligned}
\tag{5}
$$

where $\lambda > 0$ is a learning parameter. In CT, the maximum classifier score of the detected samples bag is viewed as the tracking location.

3 Proposed Algorithm

In this section, our proposed algorithm is presented in detail.

3.1   Online Feature Selection

The discrimination of the compressed features between positive and negative samples is computed, and the compressed features that have a high discrimination as an appearance model are selected to update the classifier. According to Eq. (3), it can be seen that each dimension of the compressed features v_i corresponds to a weak classifier $h(v_i)$, so the compressed feature selection process is equivalent to the process of the selected weak classifier. N features $\boldsymbol{f} = (f_1, f_2, \ldots, f_N)^T \in \boldsymbol{R}^N$ that have a high discrimination in the compressed features of $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T \in \boldsymbol{R}^n$ (with $N < n$) are selected to construct a classifier. The strong classifier can be represented as following:

$$H(\boldsymbol{f}) = \sum_{i=1}^{N} h(f_i) \tag{6}$$

where $h(f_i)$ is a discriminative weak classifier in $H(\mathbf{v})$.

3.2   Evaluating Feature Discrimination

A simple method is used to select the feature. The variance of between-sample is the variance between positive and negative samples. The variance of within-sample is the variance of either positive or negative samples. A method is applied to evaluate the feature discrimination, which is defined as follows:

$$VR(h(v_i)) = g(\sigma) \times g(\mu) \tag{7}$$

where $VR(h(v_i))$ is the discrimination of the compressed features between positive and negative samples. The lager value of $VR(h(v_i))$ represents the high discrimination of the compressed features. $g(\sigma)$ and $g(\mu)$ are defined as follows:

$$g(\sigma) = \frac{\sigma\big(h_{pos}(v_i), h_{neg}(v_i)\big)}{\sigma\big(h_{pos}(v_i)\big) + \sigma\big(h_{neg}(v_i)\big)} \times \varDelta\sigma$$
$$\varDelta\sigma = \big|\sigma\big(h_{neg}(v_i)\big) - \sigma\big(h_{pos}(v_i)\big)\big|, \quad g(\mu) = \big|\mu\big(h_{pos}(v_i)\big) - \mu\big(h_{neg}(v_i)\big)\big| \tag{8}$$

where $\sigma\big(h_{pos}(v_i), h_{neg}(v_i)\big)$ is the variance of between-sample. $\sigma\big(h_{pos}(v_i)\big)$ and $\sigma\big(h_{neg}(v_i)\big)$ are the variance of positive and negative samples feature distribution. $\mu\big(h_{pos}(v_i)\big)$ and $\mu\big(h_{neg}(v_i)\big)$ are the mean of positive and negative samples feature distribution.

It can be concluded that the high discrimination features will have the smaller value of $\sigma\big(h_{neg}(v_i)\big)$, but the lager $\sigma\big(h_{pos}(v_i)\big)$. The bigger value of $\varDelta\sigma$ means the high discriminatory power of the corresponding feature. According to the sorted

value of $VR(h(v_i))$, N dimension of feature vector that has a high discrimination is selected to update the classifier.

3.3 Weighted Classifier Construction

The positive samples (positive samples bag) of the previous frames and the negative samples (negative samples bag) of the current frame are used to train the classifier, which is defined as a sub-classifier. The current samples are used to train the other sub-classifier. Two categories of samples with the current and previous samples are cropped. According to two categories of samples, a sub-classifier $H_{cur}(f)$ of the current samples and a sub-classifier $H_{pre}(f)$ of the previous samples can be built. The amount of the previous and current positive samples is represented as K and D respectively. Finally, the final classifier can be defined as a weight of $H_{cur}(f)$ and $H_{pre}(f)$ in follows from Eq. (6):

$$H_f(f) = \frac{K}{K+D} H_{pre}(f) + \frac{D}{K+D} H_{cur}(f) \qquad (9)$$

In Eq. (9), the final classifier of $H_f(f)$ can demonstrate the contributions of current and previous samples on training classifier. Obviously, it can be seen that the features with high discrimination will contribute more to the final classifier.

4 Experiments

In this section, our method is compared with several tracking algorithms [1, 6, 7] on challenging sequences which are publicly available [9]. Animatingly, experimental results have demonstrated the superior performance of our tracker in some degree.

4.1 Experimental Results

Two metrics are applied to evaluate our tracking algorithm with some other tracking algorithms [1, 6, 7]. The first one is the Success Rate (SR), $score = \frac{area(R_T \cap R_G)}{area(R_T \cup R_G)}$, where R_T and R_G are the tracking and ground truth bounding box respectively. If $score$ is larger than 0.5 in one frame, the tracking result is considered as a success. The other one is the Center Location Error (CLE), which is the *Euclidean distance* from the tracking location center to the ground truth center at each frame. The SR and CLE are showed in Table 1a, b. It can be easily seen that our method has less average CLE and high SR to obtain the tracking results.

Table 1 The performance of our tracking method

Sequences	L_1	MIL	CT	Our
(a) Success rate (%)				
Girl	31	**49**	32	48
Deer	4	14	9	**90**
Jumping	8	46	22	**48**
Occluded face2	81	72	72	**91**
CarDark	60	8	**95**	94
Average SR	37	38	46	**74**
(b) Center location error				
Girl	62	32	41	**31**
Deer	171	66	219	**10**
Jumping	92	**10**	46	**10**
Occluded face2	11	14	16	**10**
CarDark	33	43	**5**	5
Average CLE	74	32	65	**13**

Bold fonts indicate the best performance
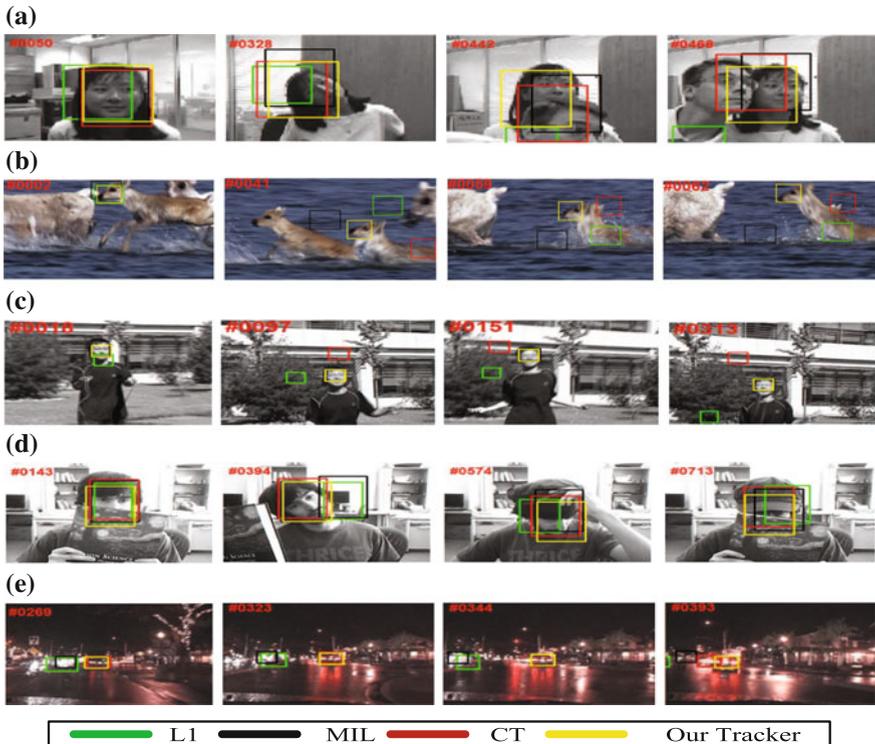
Fig. 1 Screenshots of tracking result. **a** Girl. **b** Deer. **c** Jumping. **d** Occluded face2. **e** CarDark

Occlusion, Scale and Pose Variation The target in *girl* sequence in Fig. 1a undergoes heavily scale variation, occlusion and posed variation, only our method as well as the *MIL Tracker* performs well on this sequence (e.g. frame #442, #468). The target *Occluded face2* sequence in Fig. 1d is almost fully occluded by other objects. In some frames (e.g. #394, #713), the ℓ_1 and *MIL Tracker* drift away, but our method always has a better performance.

Fast Motion and Motion Blur For the *Deer* and *Jumping* sequences shown in Fig. 1b, c, the objects undergo both fast motion and drastic motion blur. *CT* and *MIL Tracker* finally drift away, though our method always can keep best robustness.

Illumination Variation and Background Clutters The target object in *CarDark* sequence in Fig. 1e changes gradually in illumination and the surrounding background has similar texture. The ℓ_1 and *MIL Tracker* are difficult to keep tracking of the object correctly, whereas both *CT* and our tracker successfully track the correct one.

According to the experimental results are showed in Table 1a, b and Fig. 2a, b, it can be observed that the performance in terms of effectiveness and robustness in tracking considerably improves in our research, compared with other algorithms.
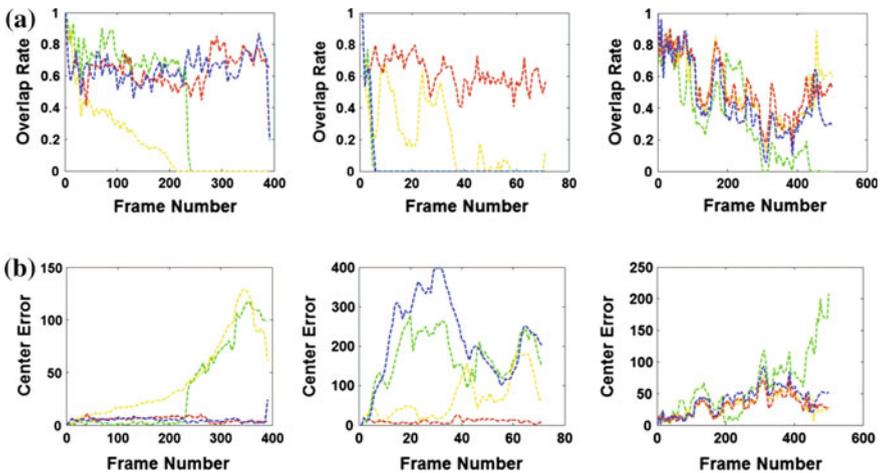
Fig. 2 The performance of tracking methods (*green line* L$_1$, *yellow line* MIL, *blue line* CT, *red line* Our). **a** Overlap rate (sequences from *left* to *right*: CarDark, Deer, Girl). **b** Center location error (sequences from *left* to *right*: CarDark, Deer, Girl)

5   Conclusion

In this paper, a simple yet effective and efficient compressive tracking algorithm based on a weighted classifier boosted by feature selection is presented. The compressed features from current and previous frames are efficiently extracted by the random measurement matrix in compressed sensing domain. Then the selected compressed features that have a high discrimination are used to train the classifier. Finally, the weighted classifier that combines two sub-classifiers, which are trained by previous and current samples bags respectively, is used to separate the target object from the background. Numerous experiments demonstrated that our method performs favorably against state-of-the-art adaptive tracking algorithms on challenging sequences in terms of efficiency, accuracy and robustness. Our future work will focus on scale variation in object tracking.

Ethical approval All applicable institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Zhang, K., Zhang, L., Yang, M.H.: Real-time compressive tracking. In: ECCV 2012. LNCS, vol. 7583, pp. 864–877. Springer, Heidelberg (2012)
2. Zhang, K., Zhang, L., Yang, M.H.: Fast compressive tracking. IEEE Trans. Pattern Anal. Mach. Intell. **36**(10), 2002–2015 (2014)
3. Zhang, J., Ma, S., Sclaroff, S.: MEEM: robust tracking via multiple experts using entropy minimization. In: ECCV 2014. LNCS, vol. 8689, pp. 188–203. Springer, Heidelberg (2014)
4. Zhang, K., Zhang, L., Liu, Q., Zhang, D., Yang, M.H.: Fast visual tracking via dense spatio-temporal context learning. In: ECCV 2014. LNCS, vol. 8689, pp. 127–141. Springer, Heidelberg (2014)
5. Danelljan, M., Khan, F.S., Felsberg, M., Joost, V.D.W.: Adaptive color attributes for real-time visual tracking. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1090–1097. IEEE Computer Society, Washington DC (2014)
6. Mei, X., Ling, H.: Robust visual tracking using l1 minimization. In: IEEE 12th International Conference on Computer Vision (ICCV), pp. 1436–1443. IEEE Computer Society, Washington DC (2009)

7. Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. IEEE Trans. Pattern Anal. Mach. Intell. **33**(8), 1619–1632 (2011)
8. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
9. Wu, Y., Lim, J., Yang, M.H.: Online object tracking: a benchmark. In: IEEE Conference on Computer vision and pattern recognition (CVPR), pp. 2411–2418. IEEE Computer Society, Washington DC (2013)

Deep Representation Based on Multilayer Extreme Learning Machine

Ya-Li Qi and Ye-Li Li

Abstract Here, we propose a fast deep learning architecture for feature representation. The target of deep learning in our model is to capture the relevant higher-level abstraction from disentangling input features, which is possible due to the speed of the extreme learning machine (ELM). We use ELM auto encoder (ELM-AE) to add a regularization term into ELM for improving generalization performance. To demonstrate our model with a high performance for deep representation, we conduct experiments on the MNIST database and compare the proposed method with state-of-the-art deep representation methods. Experimental results show the proposed method is competitive for deep representation and reduces amount of time needed for training.

Keywords Deep networks · Extreme learning machine · Deep representation · Auto encoder

1 Introduction

The essence of deep representation is transitioning from simpler elementary features to more abstract advanced features based on deep learning models. Restricted Boltzmann machine (RBM) [1] and auto-encoders [2] can be used to train multiple-layer neural networks, or deep networks. For the RBM, one is the deep belief network (DBN) [1], the another is deep Boltzmann machine (DBM) [3]. For the auto-encoder, one is the stacked auto-encoder (SAE) [2], there is another type auto-encoder model called the stacked de-noising auto-encoder (SDAE) [3]. Deep networks sometimes perform better than single-layer feed-forward neural networks

Y.-L. Qi (✉) · Y.-L. Li
Department of Computer Science, Beijing Institute of Graphic Communication,
Beijing, China
e-mail: qyl@bigc.edu.cn

Y.-L. Li
e-mail: liyl@bigc.edu.cn

(SLFNs) and traditional multilayer neural networks for big data, but they usually have slow learning speed [4]. Training the deep networks requires fine-tuning of the system weights and bias in reverse, which consumes a lot of time, thus reducing the learning speed [5–7]. To solve this problem, we proposed a method based on extreme learning machine (ELM).

Huang proposed ELM, which has a fast learning speed and good generalization [8]. Lekamalage et al. [4] propose a multilayer extreme learning machine that performs layer-by-layer unsupervised learning. It is similar to deep networks but has significantly faster speed. Yu et al. propose a deep representation ELM (DrELM) [9]. It utilizes ELM as stacking elements to construct a stacked framework.

In the work presented here, we propose an ELM-AE multilayer model for classification. The remainder of this paper is organized as following: Sect. 2 proposes the multilayer model based on ELM-AE; the experimental results are presented in Sects. 3 and 4 concludes this study and mentions the directions for future works.

2 Multilayer Model Based on ELM-AE

2.1 ELM-AE

The hidden nodes of the ELM can be randomly generated between its input and output layers, and they can be shown as a weight matrix and calculated by least-square method. The ELM can be described as following.

For N training samples $(x_i, t_i)_{i=1}^{N} \, | x_i \in R^d, t_i \in R^c, i = 1, 2 \ldots m,$

$$H\beta = T \tag{1}$$

where

$$H = \begin{bmatrix} h(x_1) \\ \ldots \\ h(x_m) \end{bmatrix} = \begin{bmatrix} G(w_1 \cdot x_1 + b_1) & \ldots & G(w_L \cdot x_1 + b_L) \\ \ldots & \ldots & \ldots \\ G(w_1 \cdot x_N + b_1) & \ldots & G(w_L \cdot x_N + b_L) \end{bmatrix}, \tag{2}$$

$$\beta = [\beta_1, \ldots \beta_L]_{L \times c}^{T} \tag{3}$$

and

$$T = [t_1, \ldots t_m]_{m \times c}^{T} \tag{4}$$

where H is the hidden layer output matrix. $G(W_i, b_i, X)$ is the hidden nodes output mapping function of the ith hidden node. $W = (w_1, w_2, \ldots w_L)$ is the hidden node

weight, and $B = (b_1, b_2,...b_L)$ is the bias. T is the target labels. β is the output weight matrix between the hidden nodes and output nodes. Then we can calculate β from

$$\beta = H^\dagger T \tag{5}$$

where $H^\dagger$ is the Moore-Penrose generalized inverse of the matrix H.

To improve generalization performance, we add the regularization parameter into the model. The regularization adds the plus constraint minimize the empirical error function, which can improve the generalization of the system [10]. In this work we add a regularization parameter as in [4, 11].

$$\beta = \left(\frac{I}{C} + H^T H\right)^{-1} H^T X \tag{6}$$

where C is a user-specified parameter to pursue a good generalization performance, and $X = [x_1, x_2,...x_3]$ is the input data.

The output of a single ELM network is $f(x) = h(x) \beta$.

The extreme learning machine auto encoder (ELM-AE) represents features via singular values. The ELM-AE usually represents the input feature in three different representations: compressed, sparse and equal dimension, (see Fig. 1). The detail of the ELM-AE are extensively covered in Lekamalage et al. [4].

For the compressed or sparse ELM-AE representation, we calculate output weights β following formula (6).

For the equal ELM-AE representation, we calculate output weights β as follows:

$$\begin{aligned} \beta &= H^{-1}T \\ \beta^T \beta &= I \end{aligned} \tag{7}$$

Fig. 1 The three types of ELM-AE representation

Input nodes

Feature mapping

Output nodes

d>L:Compressed representation

d=L:Equal dimension representation

d<L:Sparse representation
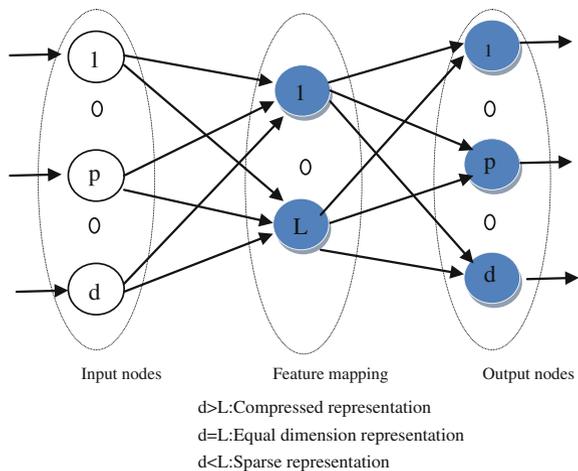
Then the ELM-AE can be summarized in Algorithm 1.

Algorithm 1
Input: Training set $X = \{(x_i, t_i) | x_i \in R^d, t_i \in R^c, i = 1, 2, \ldots m\}$, the number of hidden node is L, hidden node transfer function $G(\cdot), j = 1, 2, \ldots L$.

Output: The decision function of ELM-AE.

- Randomly generate hidden node weight parameter matrix $W \in R^{d \times L}$ and bias $B = (b_1, b_2, \ldots b_L)$;
- Compute the hidden layer output matrix $H = G(W, X)$;
- Compute output weight vector, where $\beta = (I/C + H^T H)^{-1} H^T X$ for the compressed or sparse ELM-AE representation, and $\beta = H^{-1} T$ for equal ELM-AE representation;
- Compute the decision function for classification: $f(x) = h(x) \beta$.

In the Algorithm 1, the parameter matrix W and bias B are randomly generated for the hidden node parameter matrix, so fine-tuning is not needed, unlike DBN, DBM, SAE and the like. The fine-tuning process for W and B consumes a lot of time [9, 12–14].

2.2 Multilayer Architecture

The goal of deep learning in our model is to train deep representation networks for classification problem. We propose a deep learning model similar to DrELM [9]. Our proposed architecture differs from DrELM in two ways. First, we add a regularization term to improve generalization performance and the robustness of solution. Second, we use ELM-AE in each layer and use different representation strategies in different layers. The algorithm of the multilayer architecture is showed in Algorithm 2.

The training set $(X, T) | (x^{(i)}, t^{(i)})$, $i = 1, 2 \ldots m$, is the feature vector, $t^{(i)}$ is the corresponding representation to the input $x^{(i)}$. For the ith layer in a multilayer architecture with k layers, the input is X_i, the output O_i. Then the input of the $(i + 1)$th layer is O_i.

Algorithm 2
Input: Training set $X = [x^{(1)}, x^{(2)}, \ldots x^{(m)}]^T$, and $T = [t^{(1)}, t^{(2)}, \ldots t^{(m)}]^T$ corresponding to each training instance, $x^{(i)} \in R^d, t^{(i)} \in R^c$, and hidden node number L.

Output: The decision function of multilayer ELM-AE.

- Choose the depth of multilayer architecture k;
- Initial $X_1 = X$;
- For i from 1 to k
- Generate hidden node parameters matrix W_i and B_i randomly.
- Compute hidden layer output $H_i = X_i W_i$;

Table 1 Performance comparison of our multilayer model with some other deep networks

Deep networks	Our multilayer architecture	ELM with Gaussian kernel	Deep brief network (DBN)	Deep Boltzmann machine (DBM)
Classification accuracy (%)	98.94	98.61	98.89	99.03
Training time (s)	503.64	840.87	26850	69962

- Compute output weight vector β_i, where $\beta_i = (I/C + H_i^T H_i)^{-1} H_i^T X_i$ for the compressed or sparse ELM-AE representation, and $\beta_i = H_i^{-1} T$ for equal ELM-AE representation;
- Compute the classification results $O_i = H_i \beta_i$, and $X_{i+1} = O_i$;
- End
- Compute the decision function for classification: $f_k(X) = X_k W_{1k} \beta_k$.

3 Experiments

The MNIST is often used for evaluating multi-layer or deep network performance. It includes handwritten digits images, each of which is a 28×28 pixel gray image. The experiments in this section are performed on a Lenovo workstation with an Intel Core 2.5 GHz CPU running Matlab 2010a.

The structure of proposed multilayer architecture is 784-700-700-700-12000-10, with a sigmoidal hidden layer activation function. It compares with DBNs with network structure 784-500-500-2000-10, DBM with network structure 784-500-1000-10 respectively. In addition, we compare to ELM with Gaussian kernel. The result is the average result of 20 times, which is shown in Table 1.

From Table 1, we can see the comparison of two aspects: classification accuracy and training time. For the classification accuracy, deep Boltzmann machine (DBM) is the optimal, which can get 99.03 %. The accuracy of our model can get 98.94. For the time required for training, our proposed model is the optimal, which only needs 503.64 s and has the least amount of time consumption for training. Though our model is not the optimal model for classification accuracy, it has high speed to train and is competitive to the other deep network in general.

4 Conclusion

The existing deep learning networks require a lot of time to fine-tune the weight parameter matrix and bias. In our proposed multilayer ELM model, the weight parameter matrix and bias are generated randomly and doesn't require fine-tuning. In addition, we add a regularization parameter to improve generalization

performance of the system, and use ELM-AE with different representation strategy in different layers. The experiment shows that the proposed multilayer ELM-AE has faster speed of training than the other method, though its accuracy is less the DBM.

References

1. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**, 504–507 (2006)
2. Vincent, P., Larochelle, H., Lajoie, I., et al.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. Mach. Learn. Res. **11**, 3371–3408 (2010)
3. Salakhutdinov, R., Larochelle, H.: Efficient learning of deep Boltzmann machines. Mach. Learn. Res. **9**, 693–700 (2010)
4. Lekamalage, L., Kasun, C., Zhou, H., Huang, G.B.: Representational learning with ELMs for big data. IEEE Intell. Syst. **11**, 31–34 (2013)
5. Hinton, G.E., Osindero, S., Teh, Y.W.: A Fast learning algorithm for deep belief nets. Neural Comput. **18**, 1527–1554 (2006)
6. Salakhutdinov, R., Larochelle, H.: Efficient learning of deep Boltzmann machines. J. Mach. Learn. Res. **9**, 693–700 (2010)
7. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new prespectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1798–1828 (2013)
8. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. Neurocomputing **70**, 489–501 (2006)
9. Yu, W., Zhuang, F., He, Q., Shi, Z.: Learning deep representations via extreme learning machines. Neurocomputing **149**, 308–315 (2015)
10. Wan, L., Zeiler, M., Zhang, S., LeCun, Y., Fergus, R.: Regularization of neural network using DropConnect. In: International Conference on Machine Learning (ICML), Vol. 28, No. 3, pp. 1058–1066. JMLR W&CP (2013)
11. Huang, G.B.: Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. **42**, 513–529 (2012)
12. Gu, Y., Chen, Y., Liu, J.: Semi-supervised deep extreme learning machine for Wi-Fi based localization. Neurocomputing **166**, 282–293 (2015)
13. Migel, D.T., Mark, D.M.: Deep extreme learning machines: supervised auto-encoding architecture for classification. Neurocomputing (2015 in press)
14. Han, H.G., Wang, L.D., Qiao, J.F.: Hierarchical extreme learning machine for feed-forward neural network. Neurocomputing **128**, 128–135 (2014)

Design and Implementation of ZLL-DALI Gateway for Home Lighting

Shu-Yan Sun, Jian-Guo Shi and Jun-Sheng Yu

Abstract The developing trends in home lighting include networks, intelligence and wireless capabilities. However, coexisting mainstream wire protocols such as Digital Addressable Lighting Interface (DALI) still play a prominent role in the home lighting market, which limits the development of future home lighting systems. In this case, the realization of protocol exchange between wireless protocols and wire protocols is a popular topic. To realize the lighting protocol exchange between wireless protocol ZigBee Light Link (ZLL) and DALI protocol, this paper systematically investigates the ZLL, applies a CC2530 single-chip solution and then designs the ZLL-DALI gateway, which was successfully verified by a self-built small-scale ZLL-DALI lighting system. The results indicate that wireless nodes can successfully visit the DALI cable control network through the ZLL–DALI gateway.

Keywords DALI · Protocol conversion · ZLL · CC2530 · Gateway

S.-Y. Sun · J.-S. Yu
State Key Laboratory of Electronic Film with Integrated Devices,
University of Electronic Science and Technology of China, Chengdu, China
e-mail: sysun121200@163.com

J.-S. Yu
e-mail: jsyu@uestc.edu.cn

S.-Y. Sun · J.-G. Shi (✉)
State Key Laboratory of Electronic Film with Integrated Devices,
Zhongshan Department, University of Electronic Science
and Technology of China, Zhongshan Institute, Zhongshan, China
e-mail: mcu@tom.com

1   Introduction

Home lighting composed of various connection modes plays an important role in the lighting market as well as in daily life [1]. The connection modes of home lighting control networks can be divided into two types: i.e. the wired and the wireless. The inflexible layout of cable way causes high installation costs; alternatively, most wireless protocols are private which also leads to high product and maintenance costs. Under these circumstances, the ZigBee alliance issues ZLL standard [2], which includes many advantages of ZigBee Pro, such as low cost, low power consumption, easy installation, easy extensibility, etc. Meanwhile, the ZigBee cluster library (ZCL) has been extended. For instance, the addition of the Touchlink function with coordinatorless system configurations can establish a ZLL network with the use of only one button.

In this work, in order to promote the ZLL protocol, DALI was selected as an expanding control object with which to design a ZLL-DALI gateway, which exhibits the characteristics of low development cost, open protocol and easy extensibility. Additionally, a self-built small-scale ZLL-DALI lighting system was realized.

2   Format of ZLL and DALI Data Frame

Based on ZigBee PRO, ZLL extends and standardizes the network, security and application layers. Nodes of coordinatorless ZLL systems communicate with one another via certain clusters provided in the ZCL. For instance, the ZLL commissioning cluster is used for the basic commissioning of a new network or adding a new node to an existing network; the on/off cluster is used to establish certain nodes in the "on" or "off" states, or toggle between the two states. The format of the ZLL data frame is depicted in Fig. 1. Compared to the ZigBee PRO data frame format, the APS (Application Support Sublayer) of ZLL is divided into ZCL headers and ZCL payload [2].

DALI uses a Manchester encoded unidirectional serial protocol with a transmission rate of 1.2 kHz. There are two types of data frames. The frame transmitted from the main control unit to the DALI slave unit is a forward information frame; the alternative is a backward information frame. The former is composed of 19 binary bits, including one start bit, 8 address bits, 8 data bits and two stop bits; the latter is composed of 11 bits, including one start bit, 8 data bits and two stop bits [3, 4].

PHY header	MAC header	NWK header	APS header	ZCL header	ZCL payload	FCS
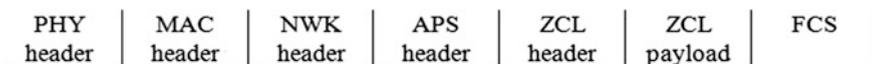
Fig. 1   Format of the ZLL data frame

3   Design of ZLL-DALI Gateway Hardware

Considering the comprehensive cost and performance, the proposed design abandons the traditional dual chip solution [5–7], which realizes wireless transceiver functions and MCU (Microcontroller Unit) architecture, respectively, with two separate chips, and then adopts the CC2530 single chip solution. CC2530 incorporates an IEEE802.15.4 RF transceiver, enhanced 8051 CPU and 8 KB SRAM, etc. Therefore, CC2530 is an ideal SoC controller to realize related applications of IEEE802.15.4/ZigBee. The hardware diagram of the ZLL-DALI gateway is shown in Fig. 2.

The power supply module provides dual voltages: 12 V for DALI bus, and 3.3 V for MCU. The CC2530 core board schematic diagram refers to the resolution of TI [8]. The built-in RF transceiver will convert ZLL frames received from the air into DALI frames, then transmit them to the DALI device via the pin of IO-TX (Input/Output Transmit). Additionally, it can send the ZLL frame transferred from the DALI feedback data frame to ZLL equipment.

To achieve the mutual conversion between the TTL and DALI electrical levels, the DALI interface circuit was adopted, as shown in Fig. 3 [9, 10]. When IO-TX is 0, Q2 will be in the conduction state and Q1 in the closed state, so that the DALI bus is characterized at a low level; when IO-TX is 1, Q2 is in a closing state, Q1 is in a conduction state, and the DALI bus is characterized at a high level. Q3 limits the maximum current of the DALI bus, because when the bus current exceeds 250 ma, the partial pressure of R5 can force Q3 to conduct and Q1 to close. R3, R4 and LM311 compose the hysteresis comparator. Compared with a single-limit comparator, a hysteresis comparator has a strong anti-interference ability. The cooperation of the hysteresis comparator and inverter can successfully transfer the DALI electrical level to the TTL (Transistor-Transistor Logic) level.

Fig. 2 Structure of hardware layer
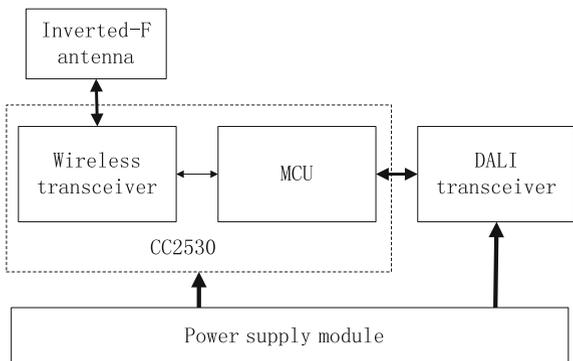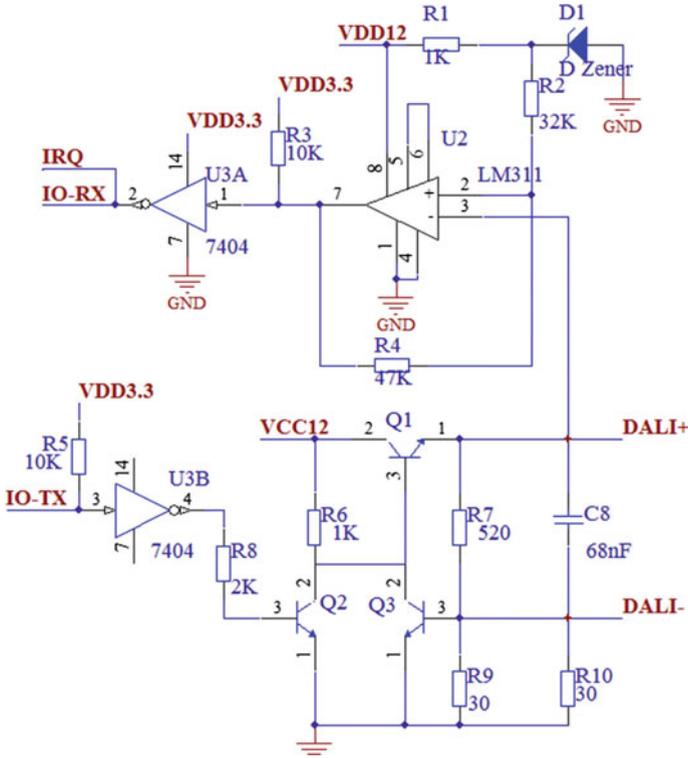
Fig. 3 Interface circuit of DALI

4 Design of Gateway ZLL-DALI Software

The software structure of the ZLL-DALI gateway is shown in Fig. 4. The programming employed is IAR Embeddded Workbench. This software was designed based on the API functions in Z-Stack Lighting 1.0.2. Such a gateway joined the network by using zllTarget_PermitJoin(), called zcl_SendCommand(), to submit messages to the ZLL node. Meanwhile, the HAL_ISR_FUNCTION() function was also used to submit messages to the DALI node.

The following three points should be taken into consideration in the software design process.

1. Due to frame format differences, the encoding method and baud rate of the two protocols, and the lack of a hardware transceiver for the built-in DALI CC2530, the software simulation method was adopted to realize the conversion between the two types of protocol. To reduce communication conflict caused by data error, following avoidance mechanism was adopted: dali_rx_status and
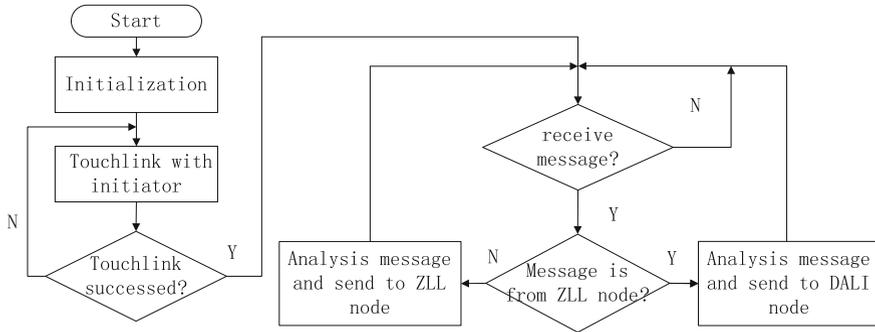
Fig. 4 Structure of software layer

dali_tx_status were used to switch the variable to identify the sending and receiving state of DALI at different times.

2. The ZLL-DALI gateway is a derivative node of the ZLL network as well as a device of the DALI system. The APS address based on the ZLL protocol stack and DALI slave device are different. Hence, an address adaptation function was established in APS to match them.

3. To address the method by which the DALI slave device can adopt odd or even values, an addressing mode was employed. When the value of the address bit is even, a data bit represents the voltage level (0–255); otherwise, a data bit represents the voltage level or the commands of the scene controlling, group controlling and status query, etc. This paper utilizes the odd value method of address.

5 Validation Test

To verify the feasibility of the ZLL-DALI gateway, a small ZLL-DALI lighting system was built, which included a ZLL remote controller, a ZLL-DALI gateway and a DALI slave module. According to the test results, the controller was able to successfully send control information to the DALI slave module through the ZLL-DALI gateway, which could control RGB tricolor LEDs with the methods of separate control, group control and scene control and could query the communication status of the DALI device. Figure 5 shows the signal of the DALI bus when the controller sent the communication status query command to the DALI slave. According to DALI protocol, the device, of which the address value is 0x07, is operating at normal communication status.

Fig. 5 Waveform of communication status query
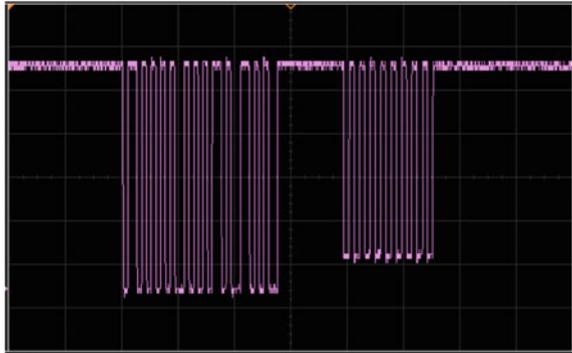
6   Conclusion

With the development of communication technology, the demand of intelligent and wireless home lighting is ever increasing. This work demonstrated the high practical value of the ZLL-DALI gateway design, which not only expands the topology of the lighting network, but also paves the way for the scalable fabrication of low-cost home lighting products.

References

1. Lee, U.S., Park, T.J., Park, M.G.: Implementation of integrated interface based on wire and wireless dual network for ensuring the reliability of intelligent LED lighting system. Korean Inst. Electr. Eng. **63**, 306–312 (2014)
2. ZigBee Alliance. ZigBee Light Link Standard. http://www.zigbee.org
3. Bellido-outeirino, F.J., Flores-Arias, J.M., Domingo-Perez, F., Gil-de-Castro, A., Moreno-Munoz, A.: Building lighting automation through the intergration of DALI with wireless sensor networks. IEEE Trans. Consum. Electron. **1**, 47–52 (2012)
4. National Electrical Manufacturers Association, Digital Addressable Lighting Interface (DALI) Control Devices Protocol. www.nema.org
5. Jahnavi, V.S., Ahamed, S.F.: Smart wireless sensor network for automated greenhouse. IETE J. Res. **2**, 180–185 (2015)
6. Wang, Y.N., Si, L.L., Liang, S.: The design of wireless sensor network gateway based on Zig Bee and GPRS. Int. J. Future Gener. Commun. Networking **7**, 47–56 (2014)
7. Hong, S.H., Kim, S.H., Kim, G.M., Kim, H.L.: Experimental evaluation of BZ-GW (BACnet-Zig Bee smart gridgateway) for demand response in buildings. Energy **65**, 62–70 (2014)

8. Texas Instruments. CC2530 Development Kit User's Guide. http://www.ti.com.cn
9. Zhang, W.C., Wang, J.C., Yuan, X.P.: Design of intelligent light control module based on DALI protocol. Comput. Modernization **1**, 154–158 (2011)
10. Hong, S.I., Song, S.Y., Lin, C.H.: An efficient multi-protocol gateway system design on the zigbee. In: 2015 17th International Conference on Advanced Communication Technology (ICACT), pp. 525–528. IEEE, Seoul (2015)

Design of a Gysel Power Divider/Combiner with Harmonic Suppression Using Cross-Shaped Transmission Line

Jin Guan, Yong Zhu, Xian-Rong Zhang and Ran Zhang

Abstract A novel modified Gysel power divider/combiner (PDC) is presented in this work. The proposed PDC consists of embedding the equivalent cross-shaped open stubs cells, and features suppression at harmonic frequencies, and area miniaturization. Also, by analyzing the ABCD matrix of the equivalent cross-shaped open stubs cell, closed-form formulas of the cell are derived. Measurement results of the proposed PDC, which are collected from the network analyzer, conform to the theoretical predictions.

Keywords Gysel power divider/combiner · Harmonic suppression · Cross-shaped transmission line · Miniaturization

1 Introduction

Power divider/combiner (PDC) is commonly used in microwave circuits. They have various applications, such as combining the output power of high power amplifiers, [1–4]. The Wilkinson and Gysel structures are the most popular structures [5–8]. However, a major drawback of the conventional PDC is the presence of spurious response due to the adoption of quarter-wavelength TL [1–4].

Recently, PDC, hybrid and coupler with harmonic suppression performance or size reduction have been studied in some papers. Researchers Guan et al. and Wu et al. studied the PDCs with even-odd mode or ABCD matrix analysis method.

J. Guan (✉) · Y. Zhu · X.-R. Zhang · R. Zhang
Southwest China Institute of Electronic Technology, Chengdu 610036, China
e-mail: 281179517@qq.com

Y. Zhu
e-mail: huyong73@126.com

X.-R. Zhang
e-mail: zxrhappy205@126.com

R. Zhang
e-mail: brian8982@126.com

Some studies [9, 10] use electromagnetic band gap (EBG) or defected ground structure (DGS) cells to make the suppression of harmonic frequency or size reduction.

In this paper, the design of equivalent cross-shaped TL cell is proposed to increase the harmonic suppression of modified Gysel PDC. In Sect. 2, the equivalent cross-shaped TL cell is analyzed by the ABCD matrix model. In Sect. 3, we present a simple process to design the proposed Gysel PDC with harmonic suppression characteristic by replacing the original TL in the Gysel PDC with equivalent cross-shaped TL. And in Sect. 4, for illustration, an example of Gysel PDC is fabricated and tested. In Sect. 5, we present the conclusion.

2 Design Cross-Shaped TL Cell for Harmonic Suppression

Figure 1 shows the original TL, equivalent cross-shaped TL and T-shaped [11] TL. In order to achieve the transmission zero at harmonic frequency f_s [11], the electrical length θ_s of the open-stub in the equivalent TL cell can be calculate as below:

$$\theta_s = \left(\frac{\pi}{2}\right)\left(\frac{f_0}{f_s}\right) \tag{1}$$

The ABCD matrix of the original TL, the equivalent cross-shaped and T-shaped TL are as below:

$$M_0 = \begin{bmatrix} \cos(\theta_0) & jZ_0\sin(\theta_0) \\ jY_0\sin(\theta_0) & \cos(\theta_0) \end{bmatrix} \tag{2}$$

$$M_c = \begin{bmatrix} \cos(\theta_{c1}) & jZ_{c1}\sin(\theta_{c1}) \\ jY_{c1}\sin(\theta_{c1}) & \cos(\theta_{c1}) \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ jY_{c3}\tan(\theta_{c3}) & 1 \end{bmatrix}$$
$$* \begin{bmatrix} 1 & 0 \\ jY_{c4}\tan(\theta_{c4}) & 1 \end{bmatrix} * \begin{bmatrix} \cos(\theta_{c2}) & jZ_{c2}\sin(\theta_{c2}) \\ jY_{c2}\sin(\theta_{c2}) & \cos(\theta_{c2}) \end{bmatrix} \tag{3}$$

$$M_t = \begin{bmatrix} \cos(\theta_{t1}) & jZ_{t1}\sin(\theta_{t1}) \\ jY_{t1}\sin(\theta_{c1}) & \cos(\theta_{t1}) \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ jY_{t3}\tan(\theta_{t3}) & 1 \end{bmatrix}$$
$$* \begin{bmatrix} \cos(\theta_{t2}) & jZ_{t2}\sin(\theta_{t2}) \\ jY_{t2}\sin(\theta_{t2}) & \cos(\theta_{t2}) \end{bmatrix} \tag{4}$$

Wu et al. [11] sets the parameters of the T-shaped TL, as shown in as Eqs. (5a, b) and (6).
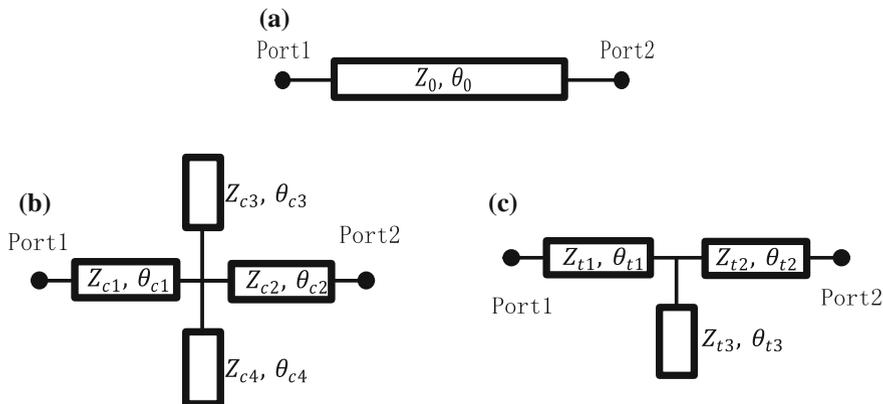
Fig. 1 **a** The original TL, **b** cross-shaped TL, **c** T-shaped TL

$$Z_{t1} = Z_{t2} = Z_0 \tan(\theta_0/2)\cot(\theta_{t1}) \tag{5a}$$

$$Z_{t3} = Z_0\tan(\theta_0/2)\tan(\theta_{t3})\frac{\cos^2(\theta_{t1})}{\cos(2\theta_{t1}) - \cos(\theta_0)} \tag{5b}$$

In order to get the parameters of the equivalent cross-shaped TL, we make $M_c = M_t$, which means that $Z_{c1} = Z_{c2} = Z_{t1}$, $\theta_{c1} = \theta_{c2} = \theta_{t1}$, and the two open stubs in the cross-shaped TL equal to the one open stub in the T-shaped TL @f_0. Then we will get Eq. (6) as below:

$$Y_{c3}\tan(\theta_{c3}) + Y_{c4}\tan(\theta_{c4}) = Y_{t3}\tan(\theta_{t3}) = Y_0\frac{\cos(2\theta_{c1}) - \cos(\theta_0)}{\tan(\theta_0/2)\cos^2(\theta_{c1})} \tag{6}$$

To suppress 2nd and 3rd harmonics, we define $\theta_{c3} = 30°$ and $\theta_{c4} = 45°$. We define the limit range of Z_{ci} (general microstrip manufacture 25–135 Ω) and θ_{ci} ($0° < \theta_{ci} \le 45°$). From Eq. (6), we can get that the values of $Z_{c1,\ c2}$, Z_{c3} and Z_{c4} depend on Z_0, θ_{c1} and θ_{c0}. The original TLs in the Gysel PDC which need to be replaced are all 90°. Then, we will get that

$$Z_{c1} = Z_{c2} = Z_0\cot(\theta_{c1}) \tag{7a}$$

$$\frac{\sqrt{3}}{3}Y_{c3} + Y_{c4} = Y_0\frac{\cos(2\theta_{c1})}{\cos^2(\theta_{c1})} \tag{7b}$$

Figure 2 shows that the values of $Z_{c1,\ c2}$, Z_{c3} and Z_{c4} depend on θ_{c1} when $Z_0 = 50\ \Omega$, $\theta_0 = 90°$, $\theta_{c3} = 30°$ and $\theta_{c4} = 45°$. From the Fig. 2, we can see that the higher θ_{c1}, the lower $Z_{c1,\ c2}$. It also shows that the higher is the value of θ_{c1}, the higher is the value of Z_{c4} when making Z_{c3} be a fixed value. So we should choose suitable values to fabricate
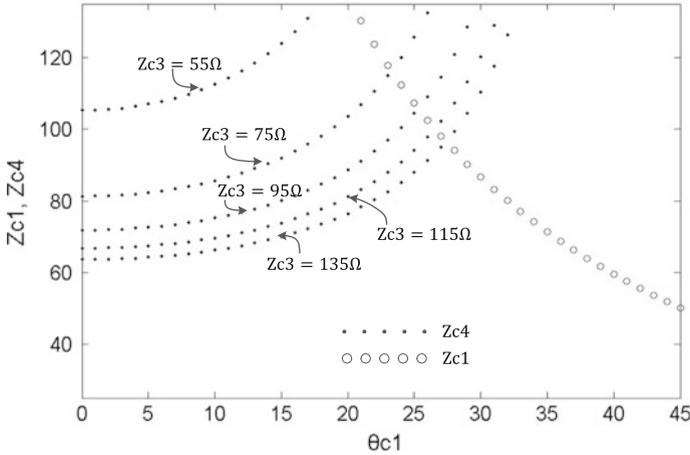
Fig. 2 The value of $Z_{c1,\ c2}$, Z_{c3} and Z_{c4} depend on θ_{c1}

Table 1 Ideal parameter values of the equivalent cross-shaped TL	Original TL		Cross-shaped TL	
	Z_0	50 Ω	$Z_{c1,\ c2}$	122.3 Ω
			Z_{c3}	88.3 Ω
			Z_{c4}	109.8 Ω
	θ_0	90°	$\theta_{c1,\ c2}$	22.3°
			θ_{c3}	30°
			θ_{c4}	45°

TLs. Table 1 gives an example of the equivalent cross-shaped TL with ideal parameter values. The simulation result in Fig. 3 shows that the cross-shaped TL not only has similar S-parameter with the original TL at around fundamental frequency, but also the bandstop filter character at the harmonic frequencies.

3 Design Procedure of Proposed PDC

In this section, we introduce the design procedure of the proposed Gysel PDC. Generally, the impedances of TLs are $Z_1 = 70.7$ Ω, $Z_2 = 50$ Ω, $Z_3 = 35$ Ω, and $Z_L = R_L = 50$ Ω, [1–4]. Then we calculate and choose the suitable parameters of the equivalent cross-shaped TL depending on Eq. (7a, b) and Fig. 2.

By replacing the TLs in the GyselPDC with the equivalent cross-shaped TL cells, the modified Gysel PDC does not change the S-parameter characters nor other characters at the fundamental frequency f_0. The modified Gysel PDC provides
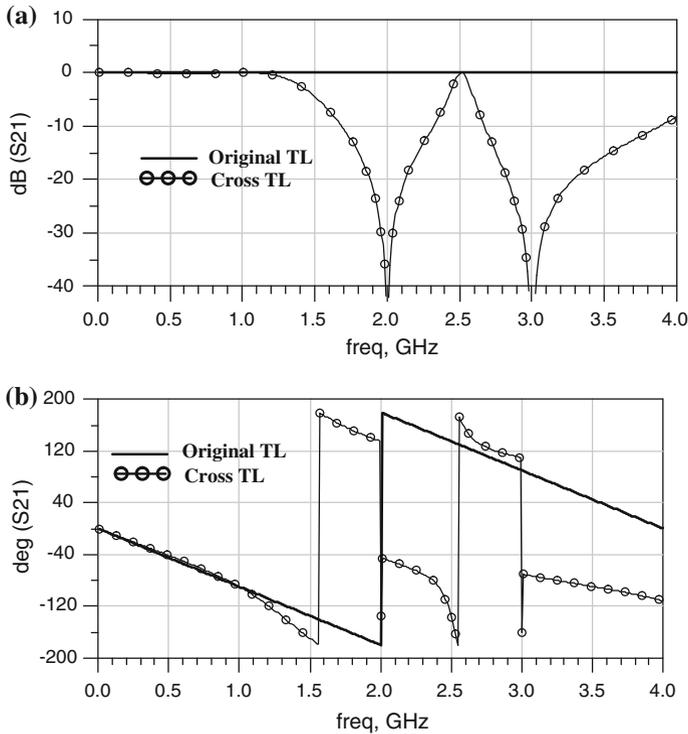
Fig. 3 Simulated results of the original TL and equivalent cross-shaped TL (f_0 = 1 GHz, Z_0 = 50 Ω, θ_0 = 90°)

transmission zeros at higher frequency. To make the design simple, the design procedure of the proposed PDC sets as follows:

1. Design a conventional Gysel PDC with the parameters of Z_i, θ_i, Z_L and R_L.
2. Calculate and choose the suitable parameters of the equivalent cross-shaped TL by formula (7a, b) and Fig. 3.
3. And then use the equivalent TL cells to replace the Z_i TL in the Gysel PDC.

4 Simulation and Measurement

In order to verify the proposed structure, a proposed Gysel PDC (f_0 = 1 GHz) has been fabricated using microstrip TL, which is shown in Fig. 4.

The measurement results are collected by a network analyzer. The simulated and measured S-parameter performance shows in Fig. 5. Inside the fundamental band, the PDC exhibits an insertion loss lower than 3.4 dB, a minimum input return loss
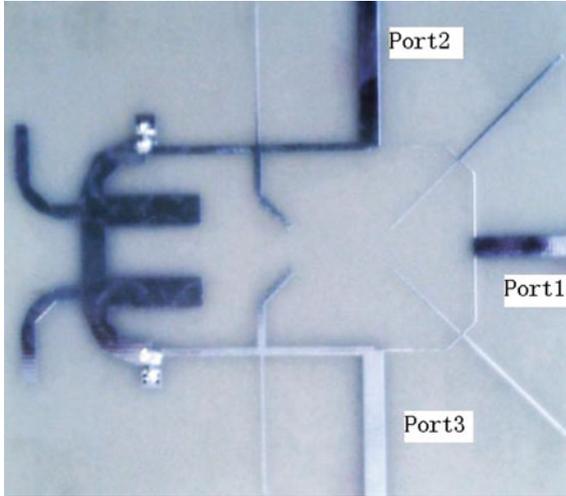
Fig. 4 The proposed Gysel PDC
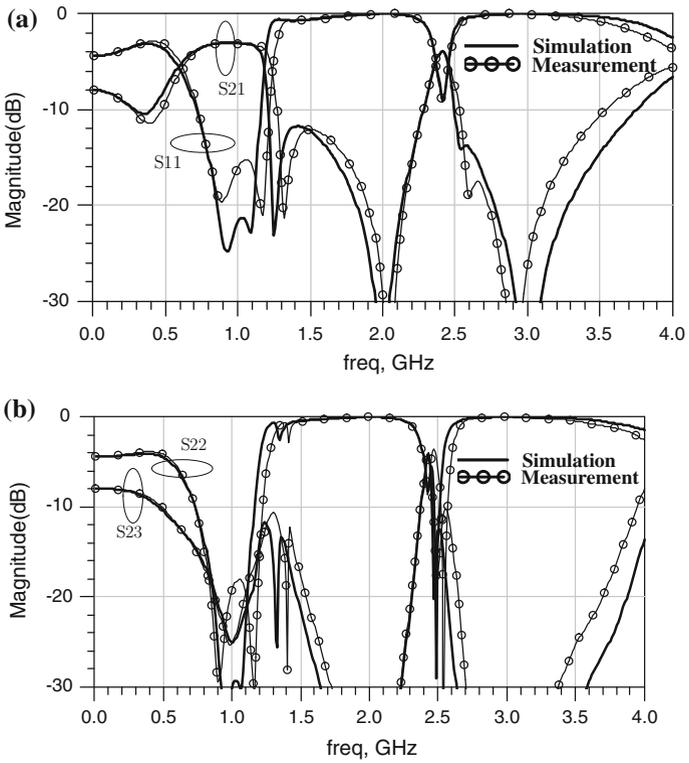
Fig. 5 The simulated and the measured S-parameter performance of Gysel PDC

and ports isolation higher than 15 dB covering a fractional bandwidth of about 30 %. From the simulation and measurement results, the proposed Gysel PDC shows good suppression at harmonic frequencies and good matching at fundamental frequency. It is also shown in Table 1 that Z_0 with $\theta_0 = 90°$ is substituted by equivalent TL structures which have smaller electrical length.

5 Conclusion

In this paper, the design, fabrication and measurements of modified two-way Gysel PDC with equivalent cross-shaped TL cells are reported. Measurement results of the proposed PDC conform to the theoretical predictions. This presented PDC features suppression at the hoped harmonic frequencies with area miniaturization. This presented Gysel PDC could be applied to microwave circuits and systems which need PDC with not only harmonic suppression performance but also high power handling capability.

References

1. Gysel, U.H.: A new N-way power divider/combiner suitable for high-power applications. In: International Microwave Symposium, pp. 116–118. IEEE, Piscataway (1975)
2. Wilkinson, E.J.: An N-way hybrid power divider. IRE Trans. Microwave Theor. Tech. **8**, 116–118 (1960)
3. Guan, J., Zhang, L., Sun, Z., Leng, Y., Peng, Y.: Designing power divider by combining Wilkinson and Gysel structure. Electron. Lett. **48**(13), 769–770 (2012)
4. Guan, J., Zhang, L., Sun, Z., Leng, Y., Peng, Y., Yan, Y.: Modified gysel power divider with harmonic suppression performance. Prog. Electromagnet. Res. C. **31**, 255–269 (2012)
5. Malakooti, S.A., Siahkamari, H., Siahkamari, P.: Compact and unequal gysel power divider with dual harmonic rejection and simple structure. Int. J. Electron. Lett. **3**(1), 58–67 (2015)
6. Tas, V., Atalar, A. A performance enhanced power divider structure. In: Microwave Symposium (IMS), 2014 IEEE MTT-S International, pp. 1–4. IEEE, Piscataway (2014)
7. Ahmadzadeh, M., Rasekh, P., Safian, R., Askari, G., Mirmohammad-Sadeghi, H.: Broadband rectangular high power divider/combiner. IET Microwaves Antennas Propag. **9**(1), 58–63 (2014)
8. Kumar, K.V.P., Karthikeyan, S.S.: A compact 1:4 lossless T-junction power divider using open complementary split ring resonator. Radioengineering **24**, 717–721 (2015)
9. Ko, Y.J., Park, J.Y., Bu, J.U.: Fully integrated unequal Wilkinson power divider with EBG CPW. IEEE Microw. Wireless Compon. Lett. **13**(7), 276–278 (2003)
10. Duk-Jae, W., Taek-Kyung, L.: Suppression of harmonics in Wilkinson power divider using dual-band rejection by asymmetric DGS. IEEE Trans. Microw. Theory Tech. **53**, 2139–2144 (2005)
11. Wu, Y., Liu, Y., Li, S.: An unequal dual-frequency Wilkinson power divider with optional isolation structure. Prog. Electromag. Res. **91**(4), 393–411 (2009)

Design of a High Performance Low Voltage Differential Signal Receiver

Kai-Xin Fan, Guang-Hui Xu, Yong Xu, Kai-Li Zhang and Ying-Cai Xiao

Abstract This paper proposes a Low Voltage Differential Signal (LVDS) transmission receiver chip design, which is fully compatible with the IEEE Std. 1596.3-1996 standard. The proposed design utilizes a new rail-to-rail fold cascode pre-amplifier to expand the receiving range, with an independent current source circuit to provide bias for the system. The chip is fabricated with CSMC (a semiconductor manufacturing corporation) 0.5 μm technology, with DC analysis, AC analysis, and transient analysis conducted on the receiver chip. Simulation results show that the chip can meet the design specifications required, within the ±1 V range of the common-mode voltage, and can achieve the hysteresis of 100 mV, with a maximum data transfer rate of 200 Mbps.

Keywords LVDS · Receiver · Rail-to-rail

1 Introduction

With the development of semiconductor process technology, the clock frequency of IC chips has been greatly improved, as has the speed of high-performance microprocessors, which already exceeds 4 Gbps. However, for conventional CMOS

K.-X. Fan · G.-H. Xu · Y. Xu (✉) · K.-L. Zhang · Y.-C. Xiao
College of Communications Engineering, The PLA University of Science and Technology, Nanjing 210007, Jiangsu, China
e-mail: xu_yong99@163.com

K.-X. Fan
e-mail: fkx123@163.com

G.-H. Xu
e-mail: xu_gh@163.com

K.-L. Zhang
e-mail: zhangkl@163.com

Y.-C. Xiao
e-mail: xiaoyc@163.com
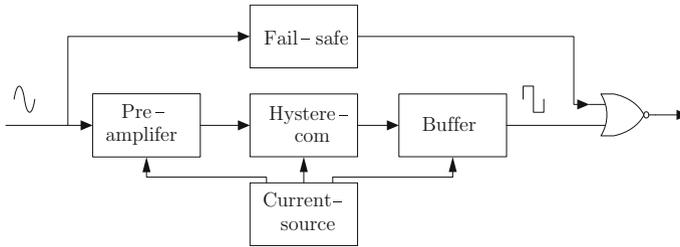
Fig. 1 Topology of receiver

(Complementary Metal Oxide Semiconductor) and TTL (transistor transistor logic) circuits, it is difficult for signals to transmit between chips at a rate of 200 Mbps or greater due to their own limitations and signal characteristics. In order to realize high-speed transmission of digital signals, Low Voltage Differential Signaling (LVDS) is proposed to replace traditional TTL and CMOS circuits [1]. The core of LVDS is to utilize the low-voltage and low-swing I/O interface to transmit differential data, which can provide advantages such as high speed, low cost and small electromagnetic interference. It is now a mainstream technology, widely-applied in optical fiber communications and data transfer between chips.

2   Topology of the LVDS Receiver

Since the LVDS receiver belongs to the peripheral interface, its role is to convert the differential signal received from the LVDS driver into a CMOS logic signal. Since there is a long distance for the signal to travel from the transmitter to the receiver, to avoid the influence of noise during signal transmission, the IEEE Std. 1596.3-1996 specifies that the input common-mode voltage of the receiver is allowed a 1 V swing and with a minimum of 25 mV hysteresis [2]. Thus, based on analysis, the receiver proposed in this paper can be divided into five modules, as shown in Fig. 1.

3   Design of the LVDS Receiver

3.1   Design of Pre-amplifier

In order to ensure that the LVDS receiver has a wide common-mode voltage range and high transmission rate, the pre-amplifier utilizes a rail-to-rail folded cascode structure [3], shown in Fig. 2, in which (M_1, M_2) and (M_1, M_2) represent the two

Fig. 2 Pre-amplifier

differential input pairs which constitute complementary differential inputs in parallel structure. Since the NMOS (N-channel metal oxide semiconductor FET) common-mode input range can reach VDD (the supply voltage of the LVDS receiver) and PMOS (P-channel metal oxide semiconductor FET) can reach GND (the ground voltage of the LVDS receiver), the two input differential pairs in parallel achieve a rail-to-rail input, to realize a wider range of common-mode input voltage.

The folded cascode circuit is of wide input common-mode range, high open-loop gain and high unity gain bandwidth, which makes it the best pre-amplifier option. (M_1, M_2, M_7, M_8) and $(M_3, M_4, M_{11}, M_{12})$ represent the two pairs of folded cascode tubes as an active load; (M_9, M_{10}) and (M_{13}, M_{14}) provide the bias current.

3.2 Design of Hysteresis Comparator

The impedance mismatch caused by the long distance will produce differential mode noise. If the noise falls in the vicinity just prior to the comparator threshold voltage, it will cause an incorrect comparator flip. In order to suppress the influence of noise, the comparator requires a hysteresis of at least 25 mV [4].

As shown in Fig. 3, a pair of MOSFET (M_5, M_6) is added in a two-stage operational amplifier, forming a parallel voltage positive feedback mechanism to generate threshold hysteresis. Considering that the circuit itself is a differential amplifier of symmetrical structure, it can be assumed that $\beta_{M1} = \beta_{M2} = \beta_1$, $\beta_{M3} = \beta_{M4} = \beta_2$, $\beta_{M5} = \beta_{M6} = \beta_3$, Using the saturation region current formula, one obtains:

Fig. 3 Hysteresis comparator

$$V_{TRP}^{+} = \sqrt{\frac{2I_{\text{bias}}}{\beta_1}\left(\frac{\sqrt{\beta_2} - \sqrt{\beta_3}}{\sqrt{\beta_2} + \beta_3}\right)} \tag{1}$$

$$V_{TRP}^{-} = \sqrt{\frac{2I_{\text{bias}}}{\beta_1}\left(\frac{\sqrt{\beta_3} - \sqrt{\beta_2}}{\sqrt{\beta_2} + \beta_3}\right)} \tag{2}$$

As shown by the above equation, the hysteresis voltage value is determined by the I bias and *W/L* of each tube. The hysteresis comparator is in the output stage, which can provide reasonable output voltage swing and output resistance, and can realize the conversion of a differential signal to a single signal. By properly utilizing the bias current source and the *W/L* of related tubes properly, an ideal hysteresis voltage can be obtained.

By adjusting the relevant parameters, the circuit generates the threshold hysteresis of 100 mV shown in Fig. 4, which meets design requirements.

3.3 Design of Fail-Safe Circuit

LVDS receivers are widely applied in multiple-bus systems and often appear idle; it may occur that the voltage on the bus cannot be determined, and may lead to

Fig. 4 Hysteresis voltage measurements
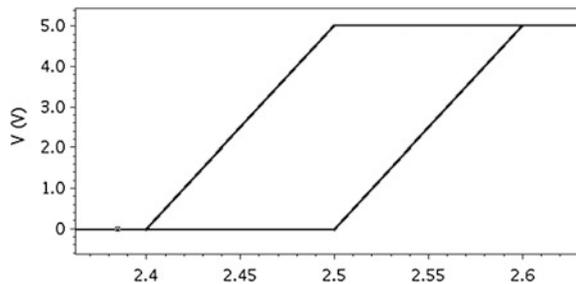
incorrect output due to the interference of noise with the input signal. Therefore, a fail-safe detection circuit can effectively avoid this by setting the possible output to a fixed level.

As shown in Fig. 5 which depicts a fail-safe circuit of a voltage comparator, when there is no input signal on the bus, the output of the fail-safe module is high, indicating that the chip is working properly. By selecting the NOR gate, the correct result can be obtained.

3.4 Design of Bias Current Source

A system bias current source circuit is designed to provide a stable bias current for the pre-amplifier circuit, the hysteresis comparator circuit and the fail-safe circuit. The bias current source used in the current study is shown in Fig. 6.

Fig. 5 Fail-safe circuit
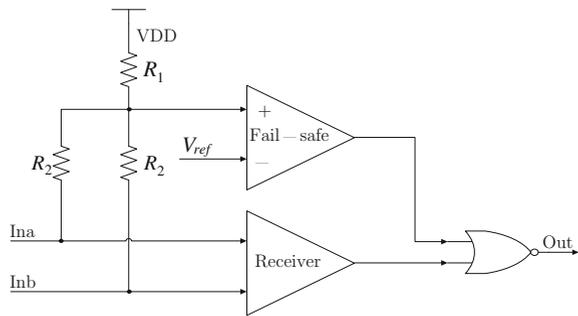
Fig. 6 Current source circuit
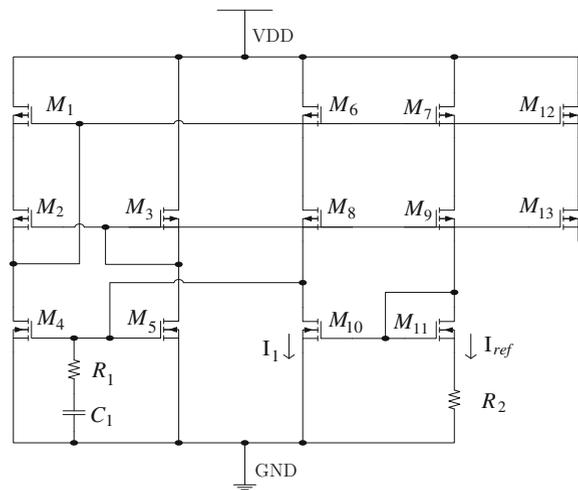
In Fig. 6, (M_8, M_9, R_2) constitutes the core of micro-current source circuit [5]. With a small voltage difference U_{gs} between the two tubes M_8 and M_9 to control output current, the following can be obtained:

$$I_1 = I_{ref} = \frac{U_{gs10} - U_{gs11}}{R_2} \tag{3}$$

The cascode current mirror is composed of M_6–M_9, while (M_{10}, M_{11}) forms a closed negative feedback loop; its high gain characteristics further enhance the stability of the reference current. (M_{12}, M_{13}) constitutes a cascode current mirror output branch with high output impedance; more branches can be expanded in order to meet system demands [6].

4 Simulation Results and Layout

The receiver chip is processed according to CSMC 0.25 μm technology. Simulation results are shown in Fig. 7. Figure 8 depicts an eye diagram of the output values when a random signal is sent to the receiver.
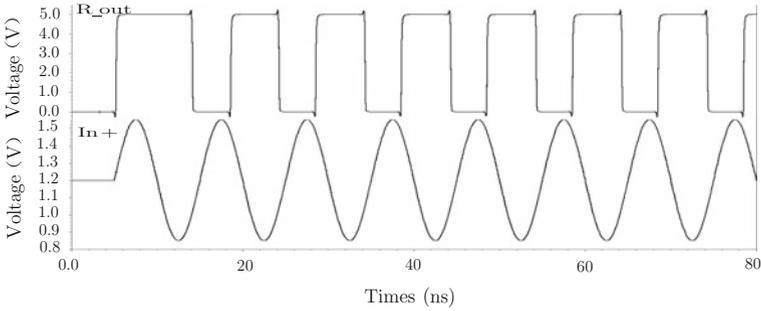
Fig. 7 Simulation result
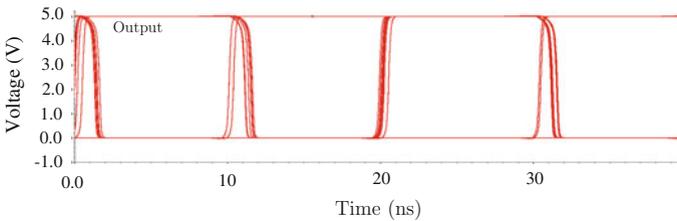
Fig. 8 Eye diagram of output
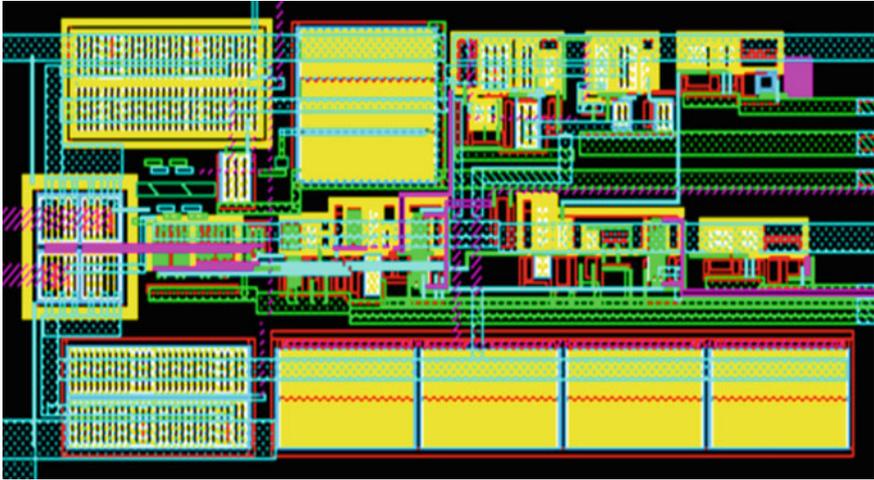
Fig. 9 Chip layout

Table 1 Receiver indicators

Parameter	Result
Power supply voltage	5 ± 0.5 V
Common mode input voltage	$0.2 - 2.2$ V
Differential input voltage	$200 - 400$ mV
Power dissipation	3.3 mW

A chip layout is shown in Fig. 9. While planning the layout of this chip, the symmetry of the differential tubes, differential signal lines and the noise shield were fully taken into account [7, 8]. Analysis of the simulation results indicates that the receiver circuit can achieve a 0.2–2.2 V common-mode range, at a rate of 200 Mbps. Specific technical indicators are listed in Table 1.

5 Summary

This paper presents a high-performance LVDS receiver chip that is fully compatible with IEEE Std 1596.3-1996. The rail-to-rail cascode circuit can realize a wide common-mode voltage range and higher bandwidth, the independent current source circuit ensures the stability of the working status of the chip, and the fail-safe circuit ensures that the chip in the idle state does not produce incorrect output results. This LVDS receiver chip can be widely-used in high-speed data transfer systems.

References

1. Li, Z., Chen, D.P., Zhao, J.Z.: Design of a TYPE-I M-LVDS receiver IC based on TIA/EIA-899 standard. Computer Engineering & Science. **37**, 452–456 (2015)
2. Alqarni, S., Kamal, A.: LVDS receiver with 7 mW consumption at 1.5 Gbps. In: International Conference on Microelectronic, vol. 26, pp. 204–207 (2014)
3. Sun, J.Z., Xie, F.Y.: Design of a LVDS Receiver for high speed data communication IC. China Integr. Circ. **157**, 39–43 (2012)
4. Matig-a, G.A., Yuce, M.R., Redoute, J.M.: An integrated LVDS transmitter in 0.18 μm CMOS technology with high immunity to EMI. IEEE Trans. Electromagn. Compat. **57**, 128–134 (2015)
5. Huang, Y., Lin, Y., Xu, Y.: The design of self-staring bias circuit chip. J. Mil. Commun. Technol. **1**, 64–68 (2014)
6. Kim, K.M., Hwang, S.: An 11.2-Gb/s LVDS receiver with a wide input range comparator. In: IEEE Transactions on VLSI Systems, vol. 22, pp. 2156–2163 (2014)
7. Xu, Y., Sun, T.Q., Zhao, F.: A Full-integrated LVDS Transceiver in 0.5 μm CMOS technology. In: Industrial Electronics and Applications, vol. 9, pp. 1672–1675 (2014)
8. Satlawa, T., Drozd, A., Kmon, P.: Design of the ultrafast LVDS I/O interface in 40 nm CMOS process. Mixed Design Integr. Circ. Syst. **21**, 200–204 (2014)

Design of a Multifunctional CAN Bus Controller for a Ground Source Heat Pump System

Hui Li, Pei-Yong Duan and Chen-Guang Ning

Abstract In order to realize the data acquisition of field instruments, device control and data communication for GSHP (ground source heat pump) systems, a multi-functional CAN (Controller Area Network) bus controller is designed with a single-chip STC12C5A60S2 microcomputer as the core. The controller makes CAN communication, wireless communication and communication conversion from RS-485 to CAN possible. Moreover, data acquisition and control modules are designed, which include an A/D (Analog/Digital) converter module, a D/A (Digital/Analog) converter module and a DI/DO (Digital Input/Digital Output) module. Testing of the controller indicates that it has strong adaptability and flexibility. Therefore, this controller meets the different data acquisition and control needs for GSHP systems.

Keywords Controller · CAN bus · GSHP · Design

1 Introduction

Since the GSHP system has the advantage of energy conservation, the technology of GSHP has developed rapidly in China in recent years. Now, the application scale of the GSHP system is ranked first in the world [1]. In order to ensure the effective operation of the GSHP system, it is necessary to monitor and control the GSHP

H. Li (✉)
Key Laboratory of Renewable Energy Utilization Technologies in Buildings
of the National Education Ministry, Shandong Jianzhu University, Jinan, China
e-mail: lhh@sdjzu.edu.cn

P.-Y. Duan · C.-G. Ning
Shandong Key Laboratory of Intelligent Building Technique,
Shandong Jianzhu University, Jinan, China
e-mail: duanpeiyong@sdjzu.edu.cn

C.-G. Ning
e-mail: chenguangning89@163.com

system in real-time. At present, the monitoring technology of GSHP systems mostly adopts ready-made devices produced by automation companies. Zhao et al. [2] developed an integrated experiment platform for the solar/geothermal energy/heat pump system. This platform adopted the Siemens S7200 Programmable Logic Controller (PLC) and Wincc configuration software. Zhang et al. [3] completed the design of a remote monitoring system of GSHP using WebAccess software and an ADAM 5510 field controller. Chen [4] designed a remote wireless monitoring system based on GPRS technology to allow remote transmission of field data for the GSHP system. Liao et al. [5] designed a remote real-time monitoring system based on Industrial Ethernet for the GSHP system. CAN bus is one of the most widely used field buses in the world, and it has been widely used in many fields [6–9]. In order to reduce the hardware cost of monitoring devices and improve the flexibility of application, a multi-functional CAN bus controller is designed in this paper, which is one node of the CAN bus system. This multi-functional CAN bus controller has the advantages of small volume, multiple functions and strong flexibility. It can not only connect with the field instruments with standard analog signals and digital signals, but also connect with the intelligent instruments with RS-485 communication. Moreover, the controller can accomplish wireless communication through a wireless communication module.

The body of this paper is organized as follows. Section 2 presents the design of a multifunctional CAN bus controller, which includes CAN, RS-485 and RS-232 communication modules, a wireless communication module, a node Identifier (ID) setting and a display circuit. Section 3 presents the design of the data acquisition and control modules. The installation and testing of these models will be outlined in Sect. 4. Section 5 summarizes the work.

2 Design of Multifunctional CAN Bus Controller

The structure of the proposed multifunctional CAN bus controller is shown in Fig. 1. The microprocessor adopts a STC12C5A60S2 chip, which has the advantages of high speed, low power consumption and strong anti-interference. The SJA1000 chip is selected as the CAN controller that implements the conversion of CAN protocol. The PCA82C250 chip is selected as the CAN transceiver, which is an interface of an actual physical bus. The RS-485 communication interface adopts a MAX485 chip, which can communicate with intelligent instruments. The RS-232 communication interface adopts a MAX232 chip, which can communicate with the upper computer. The wireless communication interface adopts a CC1100 RF chip. The EEPROM (Electrically Erasable Programmable Read-Only Memory) memory adopts an AT24C02 chip to save important data under the condition of power down. The ID setting circuit of the controller can modify the node ID via dial switches without modifying the software, which can greatly improve the flexibility
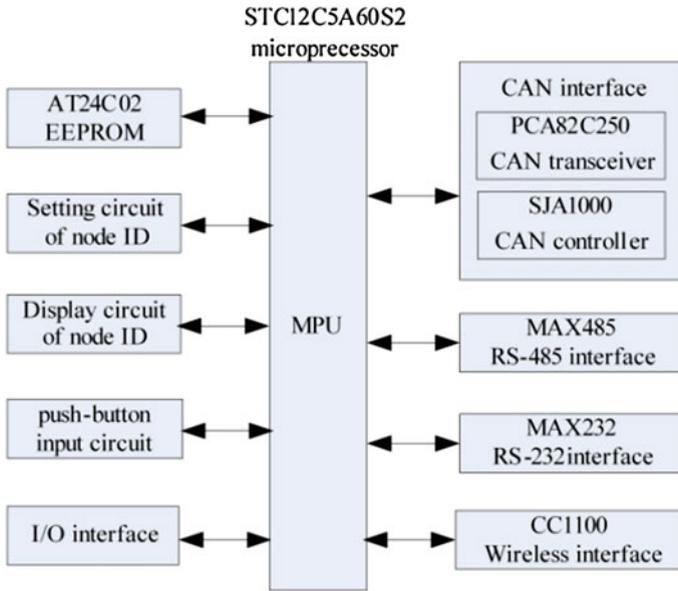
Fig. 1 Structure of multifunctional CAN bus controller

of the node ID setting. The display circuit can display the node ID of the controller. The push-button input circuit can be used to choose the communication interfaces, CAN interface or wireless interface. The I/O interface is used to connect the data acquisition and control modules. Figure 2 shows the physical CAN bus controller. The design is outlined in the following section.

2.1 Design of CAN Communication Module

The design of the CAN communication module is shown in Fig. 3. An SJA1000 chip is selected as the CAN protocol controller. A PCA82C250 chip is selected as the CAN transceiver. The pin links between SJA1000 and STC12C5A60S2 are as follows. The data signal lines AD0-AD7 of SJA1000 connect with the P0 port of STC12C5A60S2. The chip selected signal $\overline{CS}$ of SJA1000 links to the P2.7 pin of STC12C5A60S2. The read and write signals of SJA1000 link to the corresponding pins of STC12C5A60S2. The interrupt signal of SJA1000 links to the $\overline{INT0}$ pin of STC12C5A60S2. In order to improve the anti-interference of the communication circuit and realize the electrical isolation between the CAN controllers and the CAN transceiver, the high speed optocouplers 6N137 are chosen to connect the TX0 and RX0 pins of SJA1000 with the TXD (Transmit(tx) Data) and RXD (Receive(rx) Data) pins of PCA82C250. In order to improve the electromagnetic compatibility of
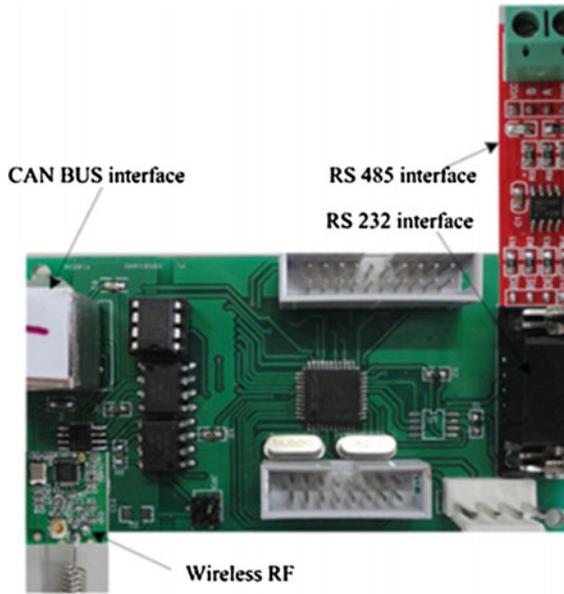
Fig. 2 Physical CAN bus controller
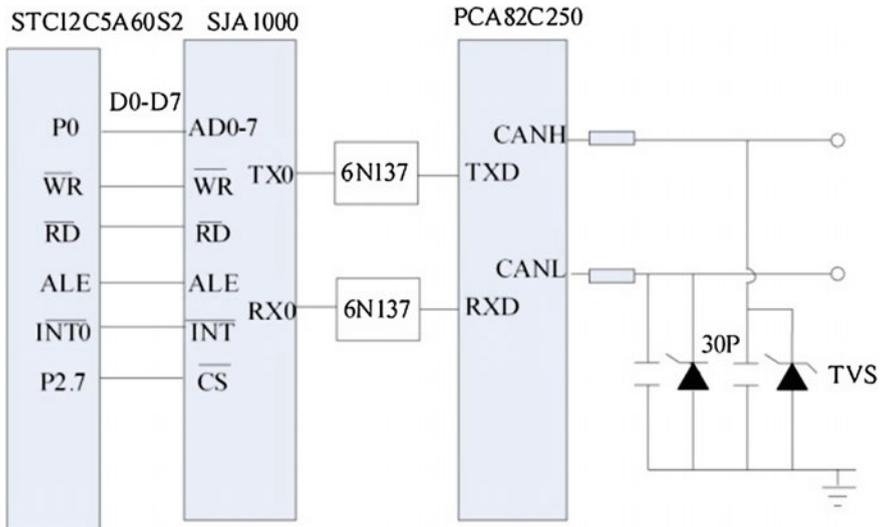
Fig. 3 CAN communication module

the CAN controller, two TVSs (Transient Voltage Suppressor) and two 30pF capacitors are chosen in the output circuit of PCA82C250. The TVS can effectively inhibit the emergence of large transient currents and prevent damage to the instruments.

2.2 Design of RS-485 and RS-232 Communication Modules

At present, most intelligent instruments adopt RS-485 communication with MODBUS protocol. The CAN bus controller can communicate with the intelligent instruments directly by RS-485 interface. The MAX485 chip is used to receive and send the RS-485 data with intelligent instruments. The P3.4 pin of the microprocessor links to the $\overline{\text{RE}}$ and DE pins of MAX485 to control the direction of RS-485 communication.

In order to realize communication between the CAN bus system and the upper computer, the RS-232 [10] communication interface is designed to incorporate a MAX232 chip. Because the RS-232 communication interface and the RS-485 communication interface use the same serial port of microprocessor, a J1 jumper is designed to select the correct communication interface. Linking pin 1 to pin 5 and pin 2 to pin 6 of J1, the RS-485 communication interface is selected. Linking pin 1 to pin 7 and pin 2 to pin 6 of J1, the RS-232 communication interface is selected. The resulting design is shown in Fig. 4.
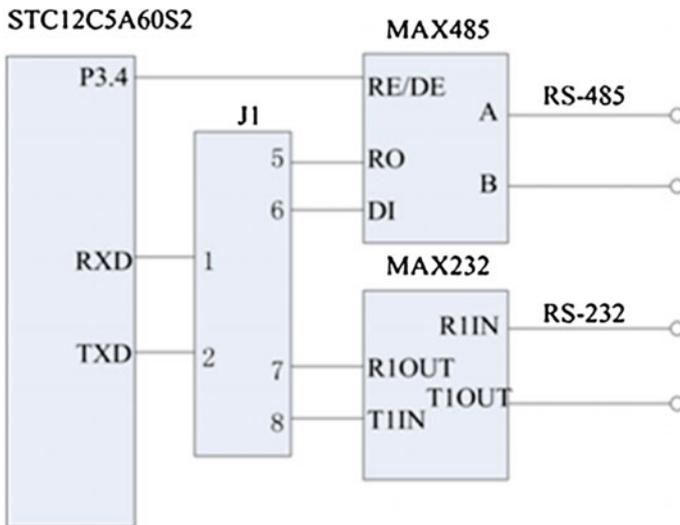
Fig. 4 RS-485 and RS-232 communication modules

2.3 Design of Wireless Communication Module

The CC1100 chip produced by Texas Instrument (TI) company in America is selected for the design of the wireless communication interface. It is a single-chip Ultra High Frequency (UHF) transceiver [11] that has the advantages of low cost and low power consumption. The CC1100 chip can work in different frequency bands, including 315, 433, 868 and 915 MHz. The designed controller adopts the 433 MHz frequency band. Because the supply voltage of the STC12C5A60S2 microprocessor is 5 V and the supply voltage of CC1100 is 3.3 V, a voltage converting circuit is designed, as shown in Fig. 5. The LM1117 chip can accomplish voltage conversion from 5 to 3.3 V. The N-Metal-Oxide-Semiconductor (NMOS) is used to accomplish bidirectional voltage conversion between 3.3 and 5 V. It is simple and effective if the communication rate is not very high. Compared with a specialized voltage level conversion chip, it can save cost. The data communication between CC1100 and microprocessor adopts the SPI (Serial Peripheral Interface) communication interface.

2.4 Design of Node ID Setting and Display Circuit

The setting circuit of the node ID shown in Fig. 6 can modify the node ID through the dial switch at any time without modifying the software, which can improve the convenience and flexibility of CAN node application. The input of bus driver 74LS244 links with the pull-up resistor array and dial switch. The output of bus driver 74LS244 links with the P0 port of the microprocessor.

Serial port 2 of the microprocessor is used to output the node ID. The A and B pins of 74LS164 link with the P1.2 pin, or the RxD2 of serial port 2 of the microprocessor. The clock (CLK) pin of 74LS164 links with the P1.3 pin, or the TxD2 of serial port 2 of the microprocessor. When the serial port 2 is set on working mode "0", the data can be sent from P1.2 when the data are sent to the serial data buffer (SBUF) register. The shift clock is sent by P1.3.

Fig. 5 Voltage converting circuit
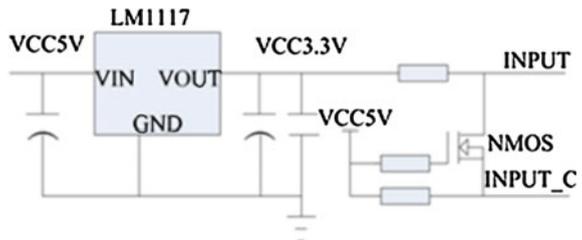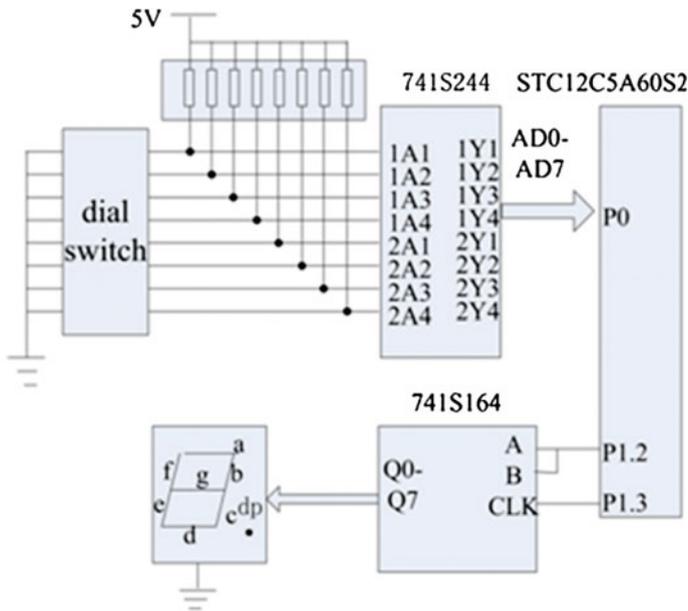
Fig. 6 Setting and display circuit of node ID

3 Design of Data Acquisition and Control Modules

The data acquisition and control modules include the A/D module, the D/A module and the DI/DO module. The 12 bit serial TLC2543 chip is selected as the A/D converter of the A/D module, which has the characteristics of high resolution, fast transformation speed, and simple circuit design. The 4–20 mA standard current signal is converted to a 0–5 V voltage signal by a RCV420 chip, which contains an advanced operational amplifier, an on-chip precision resistor network and a precision 10 V reference voltage.

The 12 bit TLV5616 chip is selected as the D/A converter of the D/A module. It has the characteristics of power down mode, low power consumption, and high input impedance. It also has 4 analog output channels.

The DI/DO module has 4 inputs and 4 outputs. The microprocessor accesses the DI/DO module through the external memory mode. When the module needs to output a DO signal, the microprocessor orderly sends an external address signal and write signal, which are the inputs of 74LS32. The output of 74LS32 is the clock signal of 74LS374. The output of 74LS374 links to the input of ULN2003A. The output of ULN2003A can drive the relay components. The ULN2003A is Darlington transistor, which has the advantages of large output current and high voltage [12]. When the module needs to input DI signal, the microprocessor orderly

Fig. 7 Installation of controllers

sends an external address signal and read signal, which are the inputs of the 'OR' gate. The output of the 'OR' gate is the enable signal of 74LS244, which can latch the inputs and send them to microprocessor.

In order to improve the anti-interference performance and realize electrical isolation between data acquisition and control modules and the CAN bus controller, high speed optocoupler 6N137 chips are used to link the external signals of data acquisition and control modules to the STC12C5A60S microprocessor.

4 Installation and Testing

An actual GSHP system in an office building located in Jinan was used to test the multifunctional CAN bus controllers. The GSHP system has two GSHP units, with two circulating pumps in the ground source side and two circulating pumps in the load side. The circulating pumps adopt variable frequency controls to save energy. In order to realize the monitoring of the GSHP system, 4 CAN bus controllers were installed, including a host node and three slave nodes, shown in Fig. 7.

The left of Fig. 7 shows the control cabinets. The top of Fig. 7 shows the Acorn RISC Machine (ARM) computer, which is based on the Linux operating system. The host node is on the right of the ARM computer. This node is responsible for communication with the ARM computer. The bottom of the figure shows the three slave nodes, which are responsible for field data acquisition and control. After continuous operation of the CAN bus control system, the controllers were stable, and data communication was normal. Therefore, this design can meet the monitoring and control needs of the GSHP system.

5   Conclusion

This paper designs a multifunction CAN bus controller for the GSHP system, which can accomplish data acquisition, device control and data communication. The communication interfaces of the controller include CAN communication, wireless communication, RS-485 communication and RS-232 communication interfaces for different applications. Moreover, the node ID setting and display circuits are designed for modifying the node ID by dial switches. The data acquisition and control modules include A/D, D/A and DI/DO modules, which can be integrated with the CAN bus controller according to different needs. The results of the testing suggest that the controller can accomplish data acquisition and control of the GSHP system. Data communication was correct and stable. The design has low cost and strong flexibility. Therefore, it is a viable candidate as a GSHP system controller.

References

1. Chen, B.X.: Six fields and prospects of the maximum potential of energy saving of building of our country. City Dev. Res. **17**(5), 1–6 (2010) (in Chinese)
2. Zhao, Y.J., Lou, C.Z., An, N.: Development of integrated experiment platform for solar/geothermal energy/heat pump system. Acta Energiae Solaris Sinica **36**(6), 1306–1311 (2015) (in Chinese)
3. Zhang, X.L., Lian, X.Q., Yu, C.C., Gao, J.: Research and implementation of ground source heat pump remote monitor system control based on WebAccess. Measur. Control Technol. **28**(6), 51–54 (2009) (in Chinese)
4. Chen, X.L.: GPRS-based ground source heat pump water supply of wireless monitoring parameters. Ind. Control Comput. **24**(2), 53–54 (2011) (in Chinese)
5. Liao, S.P., She, Q.Z., Zhu, H.Y.: Design and experimental analysis of the remote monitoring and control system for ground source heat pump. Build. Energy Conserv. **41**(9), 53–55 (2013) (in Chinese)
6. Hafizah, M., Konstantinos, M., Keith, M.: CAN bus risk analysis revisit. Lect. Notes Comput. Sci. **8501**, 170–179 (2014)
7. Matthew, D.: CAN bus technology enables advanced machinery management. Eng. Technol. Sustain. World **19**(5), 10–11 (2012)
8. Aydogmus, H.U., Akbas, A., Delibasoglu, I.: Designing a CAN BUS based measurement system for forklift simulator. In: 2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering, pp. 498–501 (2013)
9. Rezha, F.P., Saputra, O.D., Shin, S.Y.: Extending CAN bus with ISA100.11a wireless network. Comput. Stand. Interfaces **42**, 32–41 (2015)
10. Pavluchenko, A., Kukla, A., Lozovoy, S.: Simple and robust multipoint data acquisition bus built on top of the standard RS232 interface. Sens. Transducers **148**(1), 72–82 (2015)
11. CC1100 User's Manual Website, http://www.ti.com.cn/
12. ULN2003A Website, http://wenku.baidu.com/view/c72ffd2e7375a417866f8f19.html

Detecting Communities Based on Edge-Fitness and Node-Similarity in Social Networks

Cheng Yang and Wencai Du

Abstract Recent years have seen the rapid development of online social networks. Many algorithms have been proposed to assign each node to more than a single community. The traditional approaches have focused on the node community, while some recent studies have shown the great advantage of edge community detection methods. This paper presents a novel algorithm used to discover local communities in networks. A local edge community can be detected by maximizing a local edge fitness function from a seed edge which was previously ranked. Meanwhile, the method can effectively control the scale and scope of the local community based on the boundary node identification, so as to obtain complete structural information of the local community. The algorithm has been tested on both synthetic and real-world networks, and has been compared to other community detection algorithms. The experimental results show significant improvement in the detection of community structures.

Keywords Local community · Edge fitness · Node similarity

1   Introduction

With the continuous development and popularization of the internet, there constantly emerge various platforms of electronic applications under open network environments, providing rich and colorful electronic technology and virtual interactive environments for communications between people, and producing a great socio-economic effect, from traditional BBS to blogs and Twitter in recent

C. Yang · W. Du (✉)
College of Information Science and Technology, Hainan University,
Haikou, Hainan, China
e-mail: wencai@hainu.edu.cn

C. Yang
e-mail: bityangc@qq.com

W. Du
City University of Macau, Taipa, Macau, China

years, and represented by WeChat, which is a mobile social application with hundreds of millions of users. With this background, social network analysis has gradually become an important issue for both industry and academia. People hope to reveal the information, rules and knowledge hidden in the network data through quantitative and effective social network analysis. Community is an important index to describe the structure of social networks, while community discovery is one of the basic research problems in social network analysis.

This paper proposes an edge community discovery algorithm based on boundary node identification. The main contributions are as follows: (1) Breaking from the traditional community detection methods based on both nodes and edges; (2) Improving the other edge community detection methods, which are insufficient for the detection of local communities, by proposing an expansion algorithm based on edge fitness for community detection; (3) Presenting a sorting method based on edge clustering coefficients to choose the initial seed edge which can better control the distribution of local communities and optimize the speed of community detection; (4) Controlling community clustering by boundary node recognition, thus greatly reducing deviation in the size and scope of the community caused by the uncertainty of input parameters.

The remainder of this paper is structured as follows: Sect. 2 introduces the detailed process of the algorithm and its experimental environments. Results and analyses are given in Sect. 3, and a summary and proposals for future study are presented in Sect. 4.

2   Community Detection Algorithm

In this paper, the local community detection algorithm primarily consists of two steps. The first step is to select an initial edge using the edge clustering coefficient, expanding from the feed edge by maximizing the adaptation function (edge fitness). The second step is to identify boundary nodes for the local edge community acquired in step 1, eventually obtaining the complete structure of the local community. This section introduces the complete steps of the algorithm.

2.1   Selection of the Initial Seed Edge

This paper selects an edge as the initial feed. First, it will sort all edges in the network so as to select edges in an area with a large density. With this consideration, the clustering coefficient is used for sorting. In graph theory, a clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. For an edge (i, j), its clustering coefficient is defined as follows:

$$C_{i,j}^{(3)} = \frac{z_{i,j}^{(3)} + 1}{s_{i,j}^{(3)}} = \frac{z_{i,j}^{(3)} + 1}{\min(k_i - 1, k_j - 1)} \tag{1}$$

$Z_{i,j}^{(3)}$ is the number of a triangle which contain edge (i, j). After sorting all the edges in the network from big to small according to the edge-clustering coefficient, the first edge in the queue connected to its surrounding neighbor edges is closer than other edges to a certain extent. Therefore, this edge is more likely to be one of the core members in the relative community.

2.2 The Edge Fitness Function

Specific to node communities, a representative fitness function was proposed by Lancichinetti et al. [1]. Similarly, the fitness of an edge community and the fitness of edge can be obtained. For a community S, its fitness lf_s is defined as follows:

$$lf_S = \frac{m_{in}^S}{(m_{in}^S + m_{out}^S)^\alpha} \tag{2}$$

where α is an experienced parameter of a positive real number which can control the size of the community, m_{in}^S is the number of edges inside subgraph S, and m_{out}^S is the number of edges which are connected to the outside f or a community S and a neighbor edge (i, j). The fitness of this edge $lf_s^{(i,j)}$ for the community is defined as follows:

$$lf_S^{(i,j)} = \left(C_{i,j}^{(3)} + 2 \right) \left(lf_{s + (i,j)} - lf_S \right) \tag{3}$$

2.3 The Modularity Function

During boundary node identification, all the nodes in the neighbor node set of the current community must be traversed. The modularity function is introduced based on the similarity of nodes [2, 3].

2.4 The Process Based on Edge Community and Boundary Node Identification Based on Node Community

The overall process of the algorithm is described as follows:

(1) First, sort all edges in the network to the queue Q based on clustering coefficient; (2) Set the first edge in Q as the feed edge, then add this edge to community C; (3) Add new edges to the community beginning from the feed, and update the queue Q;

(4) Identify boundary nodes based on the modularity function; (5) Repeat steps 3 and 4 until the structure of the community is completely stable; this step can be ignored in most cases, but if the input parameter α is too small for the network, this step can be used to ensure the size of the community. The feature of the community tends to the modularity function at this time. (6) Set the next edge in Q as a new feed, and repeat step 3, 4, and 5 until all edges in Q have been attributed to in a community.

Algorithm 1 describes the edge sorting in step 1. Algorithm 2 describes how to extend to a community from the seed edge, and algorithm 3 describes the process of boundary node identification.

Algorithm 1: Rank Edges

```
Input: a network G = <V, E>
Output: the queue Q of edg-
es
Priority Queue Q;
For all e in E
  Node i = e.node 1;
  Node j = e.node 2;
  z3 = i.neighbors ∩
j.neighbors;
  minK = min (i.degree-1,
j.degree -1);
  c3 = (z3 + 1)/minK;
  e.c3 = c3;
  Q.insert(e);
End for
```

Algorithm 3: Identify Boundary Nodes

```
Input: a community and its
neighbor nodes
Output: boundary nodes of
the community
Boundary Nodes B
Neighbor Nodes N
For all nodes in B
  if   ΔQ_S(node) > 0
    add this node to bounda-
ry nodes;
  else
    confirm this node is
just a neighbor;
  update N;
End for
```

Algorithm 2: Find A Local Community

```
Input: A seed edge
Output: A community
Candidate Priority Queue Q_C;
Community C;
add seed to C;
for all e in seed.neighbors
  if fitness(e) > 0
    add e to Q_C;
  end if
end for
A: if Q_C is not empty
  e = Q_C.GetFront();
  add e to C;
  recalculate fitness in Q_C;
end if
else
  for all e in C
    if fitness(e) > 0
      add e to Q_C;
 end if
  end for
  if Q_C is empty
    break;
  end if
  goto A;
end else
```

For a given network G, input the parameter α which controls the size of the community structure; the proposed algorithm will obtain the feed edge by sorting, then extend to a community and identify the boundary nodes until a group of complete communities is obtained. The time complexity of the edge sorting in step 1 is O(m); m is the number of edges in the network. The time complexity of step 2 depends on the size of the local community. For a certain parameter α, the time complexity of building a local community with edges is approximately $O(e^2)$. The time complexity of the entire algorithm is $O(c * e^2)$. The parameter c is the number of communities. The most extreme situation occurs when the entire network is a community, for which the time complexity would be $O(m^2)$. Generally, such extremes will not appear. The algorithm runs quickly in most cases and will incur linear time complexity when the size of the community is small.

3　Experiment and Analysis of Results

In order to test the proposed algorithm quantitatively, the algorithm was tested both on a computer-generated network and a real network. The experimental environment consists of a PC with Intel(R) Core(TM) i3 CPU 550 @3.20 GHz, 2 GB memory and Windows 7 OS. The application environment is Java 1.8.0_45.

4　Evaluation Index

For a network without prior knowledge, metadata and label, there is no general measure such as Newman's Q-value [4] to evaluate the edge community. In this paper, typical indicators in the field of local community detection of complex networks were adopted such as precision [5], recall [6] and the harmonic index F-measure to evaluate the performance of the algorithm.

Because the community structures of computer-generated networks and portions of real networks are known, their standard data sets are easy to acquire. For other real networks whose structures are unknown, this paper takes the results of Louvain community detection algorithm whose high accuracy is currently well-known as the measuring standard [7, 8]. In order to analyze the performance of the proposed algorithm, the experiment is based on both computer-generated networks and real networks. The results obtained using the real network are compared with the performance as presented in literature [9], literature [10] and literature [11] which were known as common community detection algorithms [12].

4.1 Computer-Networks

Newman-benchmark is a computer-generated network which contains 128 nodes and 4 communities of the same size; each community contains 32 nodes. The average degree of nodes K = 16. For all nodes, the average number of edges which are linked to the outside of the community and the number of edges which are linked to the inside is $Z_{out} + Z_i = 16$. For each edge, the probability of being in a community $P_{in} = Z_{in}/(Z_{out} + Z_{in})$ while the probability of being between communities $P_{out} = Z_{out}/(Z_{out} + Z_{in})$. Therefore, the random generation of edge will fall within one community or two communities.

The proposed algorithm has two adjustable parameters: α, which represents the fitness function, and polygon parameter g which represents the clustering coefficient. The former is used to control the size of the local community while the latter is used to obtain different initial seed edges. As Z_{out} increases, the community structures of the Newman-benchmark network become increasingly fuzzy. The proposed algorithm demonstrates good performance when $Z_{out} = 3$.

4.2 Real Networks

In this section, the proposed algorithm will be tested in several real networks. These networks include: a university-based karate club [13], Dolphins' social network [14], books about U.S. politics [15], and the American College football league [4]. Detailed descriptions are provided in Table 1.

The local communities from the above real networks are detected respectively, obtaining results r, p and F. Analysis focuses on the football league and karate club networks. Figures 1 and 2 show the performance of these algorithms. Literature [9] and Literature [11] both reported low p-values and high r-values, revealing great disparity between the two indices. Literature [10] reported a greater balance between the p-value and r-value, but its r-value is low compared to that obtained by other algorithms. The performance of the proposed algorithm in this paper has balanced p- and r-values and maintains a high F-value, which proves its accuracy and effectiveness.

Table 1 The basic attributes of real networks

Network	Nodes	Edges	Description
Karate	34	78	An university-based karate club [13]
Dolphins	62	160	Dolphins' social network [14]
Books	105	441	Books about US politics [15]
Football	115	613	Network of American football games between division IA colleges during regular season fall 2000 [4]

Fig. 1 Performance of four algorithms on football league network

Football league
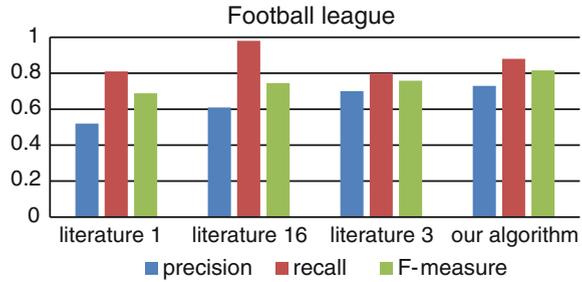
Fig. 2 Performance of four algorithms on the karate club network

Karate club
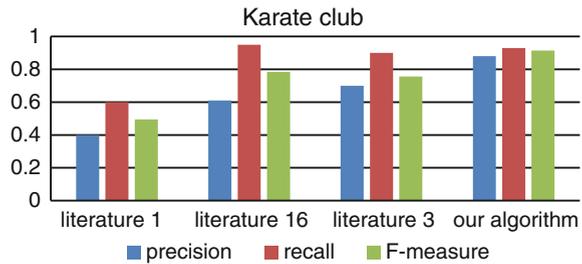
Table 2 F-measure in four algorithms

Networks	Literature [9]	Literature [11]	Literature [10]	Proposed algorithm
Karate	0.6267	0.7623	0.7633	0.7815
Dolphins	0.5433	0.8111	0.8025	0.8603
Books	0.4950	0.7833	0.8257	0.9145
Football	0.6882	0.7452	0.7584	0.8160

Designing experiments on computer-generated networks and real networks, we found that the algorithm based on edge-fitness and node-similarity demonstrates improved precision and recall to varying degrees. Although it cannot achieve optimization of both precision and recall, the two indices are balanced and are maintained at a high level, which maintains its performance advantage as a whole. This conclusion is also fully validated in the comprehensive harmonic index, as shown in Table 2.

5 Conclusion and Future Work

This paper described an algorithm of community detection which is based on edge-fitness and node-similarity. The method differs from the maximizing index of local community structure. An extended clustering coefficient and edge fitness were used as measures of local community detection, and boundary node identification

was used to control scale and scope, so as to address problems of low stability and presetting thresholds in the existing algorithms. Experimental analysis indicates that the algorithm of edge community detection based on boundary node identification demonstrates high accuracy and stability and can involve the real local community structure of a given network. On the premise of guaranteeing accuracy and stability, our next research work will investigate how to further apply this algorithm to large-scale networks and dynamic sequential networks.

References

1. Lancichinetti, A., Fortunato, S., Kertész J.: Detecting the overlapping and hierarchical community structure in complex networks. New J. Phys. **11**(3), 033–015 (2009)
2. Li, S.H., Lou, H., Jiang, W., Tang, J.H.: Detecting community structure via synchronous label propagation. Neurocomputing **151**(3), 1063–1075 (2015)
3. Feng, Z., Xu, X., Yuruk, N., Thomas, A.J.: A novel similarity-based modularity function for graph partitioning. In: Data Warehousing and Knowledge Discovery, pp. 385–396, 4654. Berlin Springer, Heidelberg (2007)
4. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**, 026–113 (2004)
5. Kim, P., Kim, S.: Detecting overlapping and hierarchical communities in complex network using interaction-based edge clustering. Physica A **417**, 46–56 (2015)
6. Aidong, Z.: Protein Interaction Networks, pp. 44–47. Cambridge University Press, New York (2009)
7. Krista, R.Ž., Borut, Ž.: A local multiresolution algorithm for detecting communities of unbalanced structures. Physica A **407**, 380–393 (2014)
8. Ling, C., Qiang, Y., Bolun, C.: Anti-modularity and anti-community detecting in complex networks. Inf. Sci. **275**, 293–313 (2014)
9. Clauset, A.: Finding local community structure in networks. Phys. Rev. E **72**(2), 254–271 (2005)
10. Chen, Q., Wu, T.T., Fang, M.: Detecting local community structures in complex networks based on local degree central nodes. Phys. A **392**(3), 529–537 (2013)
11. Luo, F., Wang, J.Z., Promislow, E.: Exploring local community structures in large networks. Web Intell. Agent Syst. **6**(4), 387–400 (2008)
12. Cui, Y.Z., Wang, X.Y., Eustace, J.: Detecting community structure via the maximal sub-graphs and belonging degrees in complex networks. Physica A **416**, 198–207 (2014)
13. Zachary, W.: An information flow model for conflict and fission in small groups. Anthropol. Res. **33**, 452–473 (1977)
14. Newman, M.E.J.: Modularity and community structure in networks. Proc. National Acad. Sci. **103**(23), 8577–8582 (2006)
15. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. **103** (23), 8577–8582 (2006)

Energy-Efficient Algorithm for CDMA Uplink Based on Nash Bargaining Solution

Chuan-Chao Wang, Jin-He Zhou and Yuan Zhang

Abstract To improve wireless terminal energy efficiency, code division multiple access (CDMA)-based systems, which are still widely-used in existing 3G networks, are chosen as the scenario. Through the establishment of the terminal energy efficiency model, a Nash Bargaining Solution (NBS) based on a cooperative game is applied to uplink resource allocation, and a distributed algorithm is designed to obtain the optimal solution. The algorithm satisfies each user's power and rate constraints, reflects fairness of resource allocation, and maximizes the overall energy efficiency in uplink wireless communication systems, so as to realize energy efficiency.

1 Introduction

In order to solve the problem of high energy consumption in wireless communications, scholars have proposed many energy-efficient strategies. The optimal packet transmission strategies of energy efficiency were derived in order to meet single packet delay constraints [1–3]; the problem of selecting the optimal transmission rate based on the minimum leave curve by using the network calculus method has been visually proved [4] and a corresponding algorithm was proposed.

C.-C. Wang · J.-H. Zhou (✉) · Y. Zhang
Beijing Information Science and Technology University, Beijing, China
e-mail: zhoujinhe@bistu.edu.cn

C.-C. Wang
e-mail: woaiwchch@126.com

Y. Zhang
e-mail: tttqzyzy@163.com

However, it is not optimal for the independent energy efficiency transmission control strategy of a single node. An energy efficient non-cooperative game algorithm was applied based on a CDMA system [5, 6], but simply introducing a cost function does not reflect the fairness of the algorithm, and increased the complexity, which may not obtain the Pareto optimal solution [7–9], due to the selfishness of a non-cooperative game.

To address the above problems, we innovatively introduce the cooperative game theory in CDMA uplink, and propose an energy efficient algorithm of joint power and rate allocation based on NBS. The algorithm obtains the Pareto optimal solution, reflects fairness, maximizes the uplink's overall energy efficiency, and meets each terminal's maximum and minimum constrains of power and rate.

2 System Model

Consider a 3G cellular network with a single base station (BS) and N terminals. The SINR corresponding to a given terminal i is represented as follows:

$$\gamma_i(r_i, p_i) = \frac{W}{r_i} \cdot \frac{h_i p_i}{\sum_{j=1, j \neq i}^{N} h_j p_j + \sigma^2} \tag{1}$$

where p_i and r_i are the respective transmission power and transmission rate, W is the spread-spectrum bandwidth. h_i denotes the link gain, and σ^2 is the Gaussian white noise. An energy-efficient function with power and rate [6] is expressed as follows:

$$u_i = \frac{L r_i f(\gamma_i)}{M p_i} \tag{2}$$

where $f(\gamma_i) = (1 - 2P_e)^M$ is the efficiency function and $0 \leq f(\gamma_i) \leq 1$, which denotes the probability of correctly-transmitted data packets. P_e is the bit error rate (BER), L represents the number of information bits of M length in the packet. Considering the use of non-coherent FSK demodulation, the bit error rate is $P_e = \frac{1}{2} e^{-\gamma/2}$.

3 Cooperation Game Model and Algorithm Design

3.1 NBS Model

NBS focuses on the allocation of revenue when people cooperate together. NBS is the equilibrium solution after the game participants' bargaining, and must meet the Nash axioms [10]. Let $u^{\min} = \left[u_1^{\min}, u_2^{\min}, \ldots, u_i^{\min}\right]$ which denotes the set of the

minimum energy efficiency demands in a cooperative game. Therefore, the control model based on NBS is expressed as follows:

$$(p, r) \in \arg \max_{p_i, r_i} \prod_{i=1}^{N} \left(u_i - u_i^{\min} \right)$$

$$s.t. \quad a(1) . r_i^{\max} \geq r_i \geq 0, i = 1, 2, \ldots, N \tag{3}$$

$$a(2) . p_i^{\max} \geq p_i \geq 0, i = 1, 2, \ldots, N$$

where $r = (r_1, r_2, r_3, \ldots, r_N)$ and $p = (p_1, p_2, p_3, \ldots, p_N)$ are vectors, and $r_i^{\max}$ and $p_i^{\max}$ respectively denote the maximum rate and power. The equivalent logarithmic formation of Eq. (3) is described as follows:

$$(p, r) \in \arg \max_{p_i, r_i} \sum_{i=1}^{N} \ln \left(u_i - u_i^{\min} \right), \quad s.t \ a(1), a(2) \tag{4}$$

The objective function in Eq. (4) is a concave function [7], and its inequality constraint functions satisfy the convex features; therefore, this problem has an optimal solution. We construct the Lagrange function as follows:

$$U(r, p, \mu, \lambda) = \sum_{i=1}^{N} \ln \left(u_i - u_i^{\min} \right) + \sum_{i=1}^{N} \mu_i (r_i^{\max} - r_i) + \sum_{i=1}^{N} \lambda_i (p_i^{\max} - p_i)$$

$$= \sum_{i=1}^{N} \left[\ln \left(u_i - u_i^{\min} \right) - u_i r_i - \lambda_i p_i \right] + \sum_{i=1}^{N} \left(u_i r_i^{\max} + \lambda_i p_i^{\max} \right) \tag{5}$$

where the vector $\mu(\mu_i \in \mu)$ and $\lambda(\lambda_i \in \lambda)$ are the Lagrange multipliers.

Since the objective function and constraint functions satisfy KKT conditions, the dual problem is expressed as follows:

$$d^* = \min_{\mu, \lambda} \max_{p, r} U(r, p, \mu, \lambda)$$

$$s.t. \quad \mu_i \geq 0, \ \lambda_i \geq 0, \quad i = 1, 2, \ldots, N \tag{6}$$

By solving the dual problem, we can obtain the optimal solution of the original problem, and reduce the computational complexity.

3.2 Design of Distributed Algorithms

The simplified gradient projection method is adopted to obtain the optimal solution μ^*, λ^* of the dual problem. The iterative distributed algorithm for the optimal solution is as follows:

1. At the initial moment, each user terminal randomly selects their initial transmission power $p_i(t)$ and rate $r_i(t)$ within the scope of power and rate. The values of multipliers are $\mu_i(t) = 1/r_i$, $\lambda_i(t) = 1/p_i$.
2. The terminals recognize each uplink interference noise via the BS, then update their rate and power according to the following optimization problem:

$$\arg \max_{r_i, p_i}[\ln(u_i - u_i^{\min}) - u_i r_i - \lambda_i p_i]$$

$$r_i = \min\{r_i, r_i^{\max}\}, \quad p_i = \min\{p_i, p_i^{\max}\}$$

3. Update the Lagrange multipliers as follows:

$$\mu_i(t+1) = \mu_i(t) - a(t)(r_i^{\max} - r_i)$$
$$\lambda_i(t+1) = \lambda_i(t) - a(t)(p_i^{\max} - p_i)$$

4. If $\|p_i(t+1) - p_i(t)\| \leq \varepsilon$, the iteration ends; otherwise returns to step (2) until the algorithm converges. Let ε be a small number.

4   Simulation and Analysis

We used established system model parameters [9] to investigate the effectiveness of the proposed algorithm on a single cell system with 11 stationary terminals. The terminals are considered at distances d = [50, 100, 150, 200, 250, 300, 350, 400, 450, 450, 500], the maximum rate of each terminal is 96 Kbps, the maximum power is 0.2 W, the link gain is $h_i = c/d_i^4$ and $c = 0.097$, the variance of received noise is

Fig. 1 Rate comparison
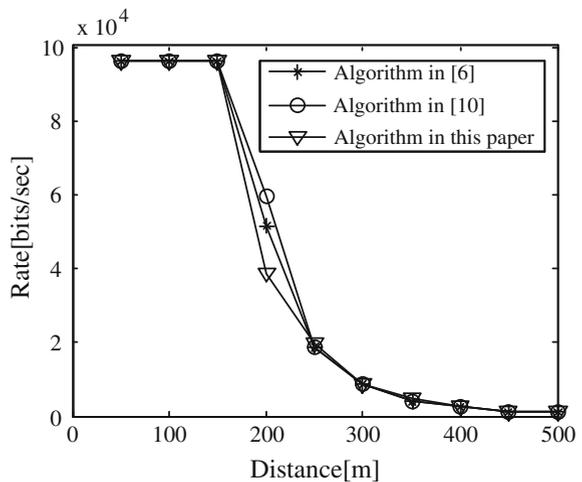
Fig. 2 Power comparison
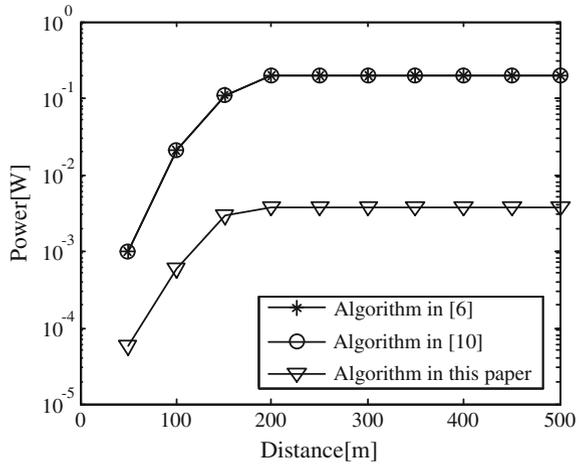
$\sigma^2 = 5 \times 10^{-15}$ W. Let $L = 64$ bits, $M = 80$ bits, and $u_i^{\min} = 0$ bits/J. Assume that the characteristics of all =wireless terminals are identical. The simulation results compare with previous proposed methods [6, 9].

As shown in Figs. 1, 2 and 3, the rate and SINR results are almost identical to those of the other two algorithms, and the proposed algorithm has a lower convergence power and higher energy efficiency. Meanwhile, the proposed algorithm effectively solves the near-far effect. The majority of SINRs converge to 12.42, indicating that the terminal can transmit data with a very low bit error rate.

Figure 4 shows that the proposed algorithm is able to achieve a higher energy efficiency, which is closely related to the lower transmission power displayed in Fig. 2, because the reasonable power allocation reduces the interference between one another. In this scene, two terminals at an identical distance of 450 m are able to

Fig. 3 SINR comparison
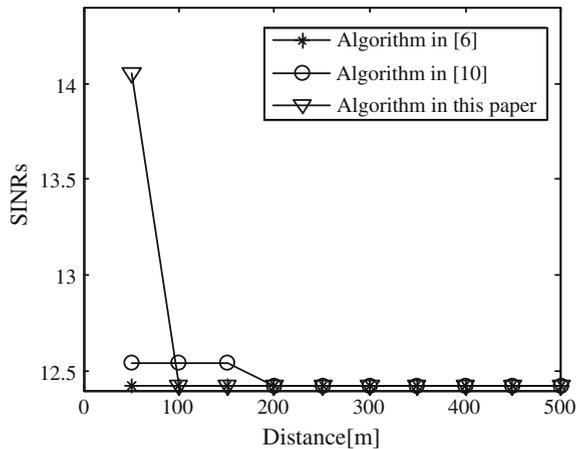
Fig. 4 Energy efficiency comparison
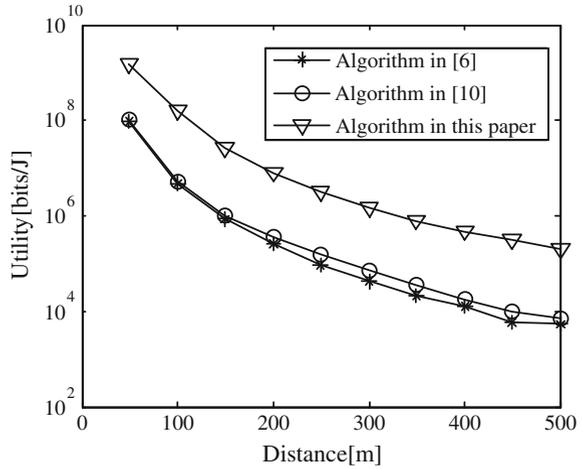
converge to the same energy efficiency, thus reflecting the fairness of allocation. The algorithm also achieves the Pareto optimal solution, improves each terminal's energy efficiency, reflects the individual rationality of NBS and ensures fairness.

5   Conclusions

Due to the fact that current CDMA-based 3G networks are still widely-used, it is still of great significance to study the energy efficiency of CDMA-based networks. In this paper, joint power and rate control problems are solved to maximize overall energy efficiency of wireless terminals by using the cooperative game theory in uplink CDMA systems. By equivalent transformation of the NBS model, we decompose the dual optimization problem, and finally propose a distributed algorithm in which power and rate are adaptive in order to be reasonably allocated. The algorithm satisfies each terminal's power and rate constraints, reflects fairness, and maximizes the overall energy efficiency.

References

1. Jin, Y., Xu, J., Qiu, L.: Energy-efficient scheduling with individual packet delay constraints and non-ideal circuit power. J. Commun. Netw. **16**(1), 36–44 (2014)
2. Darabkh, K.A., Awad, A.M., Ala'F, K.: Efficient PFD-based networking and buffering models for improving video quality over congested links. Wireless Pers. Commun. **79**(1), 293–320 (2014)

3. Nordio, A., Chiasserini, C.F., Tarable, A.: Bounds to fair rate allocation and communication strategies in source/relay wireless networks. IEEE Trans. Wireless Commun. **13**(1), 320–329 (2014)
4. Zafer, M., Modiano, E.: Optimal rate control for delay-constrained data transmission over a wireless channel. IEEE Trans. Inf. Theory **54**(9), 4020–4039 (2008)
5. Daoud, A., Kesidis, G., Liebeherr, J.: Zero-determinant strategies: a game-theoretic approach for sharing licensed spectrum bands. IEEE J. Sel. Areas Commun. **32**(11), 2297–2308 (2014)
6. Musku, M.R., Chronopoulos, A.T., Popescu, D.C.: A game-theoretic approach to joint rate and power control for uplink CDMA communications. IEEE Trans. Commun. **58**(3), 923–932 (2010)
7. Tsiropoulou, E.E., Vamvakas, P., Papavassiliou, S.: Joint utility-based uplink power and rate allocation in wireless networks: a non-cooperative game theoretic framework. Phys. Commun. **9**, 299–307 (2013)
8. Pandremmenou, K., Kondi, L.P., Parsopoulos, K.E., Bentley, E.S.: Game-theoretic solutions through intelligent optimization for efficient resource management in wireless visual sensor networks. Sig. Process. Image Commun. **29**(4), 472–493 (2014)
9. Zhao, W., Lu, M.: Distributed rate and power control for CDMA uplink. In: Wireless Telecommunications Symposium, pp. 9–14. IEEE Press, New York (2004)
10. Nash, J.F.: The bargaining problem. Econometrica: J. Econometr. Soc. **18**, 155–162 (1950)

Enhanced Simulation-Based Verification and Validation of Automotive Electronic Control Units

Thomas Herpel, Thomas Hoiss and Jan Schroeder

Abstract Modern cars comprise of increasingly complex electronic devices. Especially for electronic control units (ECUs) in safety-critical application areas, testing must be effective and efficient in the potentially conflicting area of increasing system complexity and shorter development cycles. As a significant enhancement to well-known development efforts according to the V-Model, the use of ECU simulation models at the early stages of development were proposed. These models closely resembled the behavior and functionality of a real ECU in a state-based implementation compliant to modeling standards like UML (Unified Modeling Language). The approach was seamlessly embedded in automotive hardware in loop-based ECU integration and system testing, which allowed for a thorough verification and validation of basic system requirements and test case implementations.

Keywords Automotive electronics · Verification and validation · Simulation · System modeling · State charts · Hardware-in-the-Loop testing

1 Introduction

Modern luxury vehicles are equipped with up to one hundred electronic control units (ECUs) serving various purposes like driver assistance, occupant protection, and on-board entertainment. Thus, the more software functions implemented in these

T. Herpel (✉) · T. Hoiss
Automotive Safety Technologies, Gaimersheim, Germany
e-mail: thomas.herpel@astech-auto.de

T. Hoiss
e-mail: thomas.hoiss@astech-auto.de

J. Schroeder
Department of Computer Science and Engineering,
University of Gothenburg, Gothenburg, Sweden
e-mail: jan.schroder@gu.se

ECUs, the more complex the hardware becomes. Furthermore, the broad variety of vehicle options, shorter production cycles, and faster times-to-market yield challenges during the development of the automotive electronic systems. Especially for safety-critical systems, like airbag control or emergency braking systems, safe and reliable development must be assured without compromise. Electronic development in the automotive domain typically follows a process depicted in the "V-Model", as shown in Fig. 1, starting from required specifications (upper left) to hardware/software implementation (lower tip) to final application tests (upper right). The key advantage of this developmental process model is a direct connection of each development phase on the left hand side to its counterpart validation activity at the corresponding layer on the right hand side. However, this means that all development and test activities have to be performed consecutively, one after the other. This can sometimes lead to significant delays when tests at a certain level are performed and systematic faults, if they exist, are discovered. To overcome this, this study proposed that the information encoded in the system requirements be used to derive simulation models of the under-development ECUs at the very early stages. The proposed simulation component is called "Software-ECU" (Soft-ECU) and resembles the functional behavior of a real ECU according to the system architecture, e.g. controller area network (CAN), FlexRay bus communication, failure handling, fault tolerance, and diagnosis. However, non-functional aspects, such as memory consumption, should still be validated using real ECU hardware samples in order to obtain realistic and unbiased results. By applying a suitable simulation environment, the Soft-ECU can be exposed to test cases that are derived from the same system requirements for later use in the verification and validation process steps. Hence, effectiveness and efficiency in ECU development are enhanced on two fronts. First, test cases can reveal basic systematic faults at the very beginning of hardware development without the need for a physically present hardware sample. Second, test
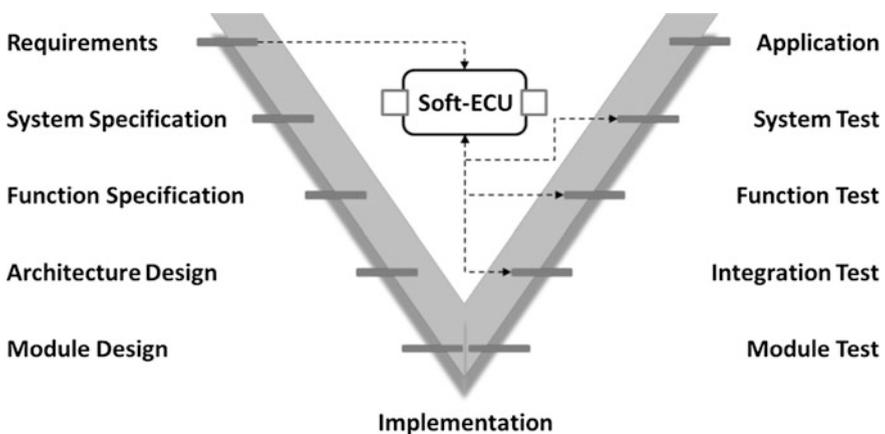
Fig. 1 Electronics development according to V-Model with use of Soft-ECU

case validation is achieved without obstructing test system resources like hardware-in-the-loop simulators (HiL), i.e. the correct implementation of the system tests can be validated. In Fig. 1, these two aspects are resembled by the block "Soft-ECU" in the middle portion connected to both the requirements on the left and the different test activities on the right.

Because design and production of the first prototype hardware samples and the validation of up to a few thousand test cases are typically very time-consuming during developmental stages, significant benefits at the early stages are expected by front-loading these activities.

This paper is organized as follows: Sect. 2 presents related work in the said area of research. In Sect. 3, the most common testing approaches for automotive electronics are presented. The proposed process enhancements are introduced in Sect. 4, together with some first analytical results from the Soft-ECU modeling in a real vehicle development project. Finally, Sect. 5 summarizes the paper and gives an outlook on future work.

2 Related Work

Herpel et al. [1] have shown how to improve ECU testing by applying combined electric/electronic and functional HiL simulator testing. Challenges in integration testing of automotive electronics were identified and strategies for sound validation and verification were proposed by Schroeder et al. [2]. Suh et al. and Schuette et al. [3, 4] proposed a HiL- or PC-based simulation environment for hardware testing, while requiring a real hardware sample for validation purposes. Himmler [5] introduced virtualization in ECU testing from the field of aircraft system development. Model-based development of ECU software and validation were employed by Vora et al. [6] at various stages of the development of powertrain control strategies. Witter et al. [7] combined a HiL test of ECUs with a virtual environment for the generation of driving scenarios, which required a real ECU. Caraceni et al. [8] proposed to deploy a real-time model of an engine ECU in software and HiL tests to shift as much of the development as possible to off-vehicle testing. However, for HiL testing, a prototype sample of the ECU is needed. Andrianarison and Piques [9] employed the systems modeling language (SysML) for modeling the system capabilities and system interaction on various levels of architecture.

3 Testing Approaches for Automotive Electronics

In order to put safe and reliable systems to market, thorough testing activities for both hardware and software are performed during automotive ECU development. From the system requirements, which encode the functional and non-functional

specifications and are typically denoted in a semi-formal or formal manner, test cases are derived. Depending on system complexity, this yields up to thousands of test cases to cover all the specification aspects. The V-Model encodes a stepwise testing process for various consecutive stages, which is well known and commonly applied with respect to developmental standards like ISO 26262 [10].

3.1 Module Tests

A module test is the very first testing activity after implementation of ECU software. It is performed either as model-in-the-loop (MiL), where interfaces of model-based ECU software components are stimulated to validate correct implementation of requirements, or as software-in-the-loop (SiL), where ECU software is compiled according to the target hardware platform, and MiL test cases are applied to the compilation for purposes of back-to-back testing. Additionally, processor-in-the-loop (PiL) testing aims to assess the compiled software integrated into the target hardware platform.

All approaches—MiL, SiL, and PiL—serve to validate fulfillment of the functional requirements. While MiL and SiL testing are typically performed on a desktop PC and cover no ECU hardware aspects like memory consumption or runtimes, PiL testing is carried out on a target hardware platform processor boards.

3.2 Integration Tests

The aim of integration testing is to prove successful hardware and software integration for an ECU sample. To this end, the ECU sample is connected to a HiL simulator, which is capable of stimulating the ECU and tracking the corresponding reaction. It is typically a black-box test, where automated test cases trigger the HiL simulator. The focus is on electrical and electronic aspects, like bus communication, diagnosis data or handling of peaks, leakages, or toggling of the power supply [1].

3.3 Function Tests

A proposal for effective validation of safety-critical functionality of an airbag ECU was shown by Herpel et al. [1] using an enhanced HiL simulator, which allows for direct stimulation of the ECU-internal sensor values. In general, functional testing aims to estimate the functional performance of an ECU in terms of both correct and on-time reaction to a given set of input values.
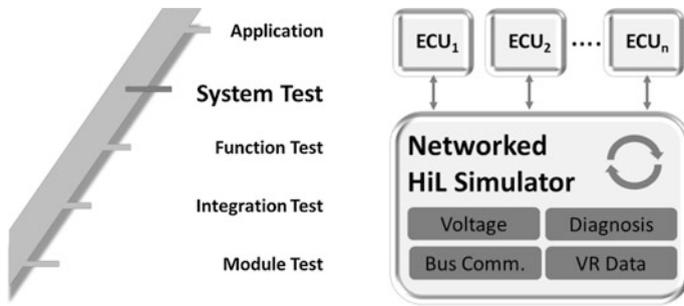
Fig. 2 Networked HiL simulator at ECU system test level

3.4 System Tests

System tests focus on the validation of plausible behavior of an ECU when interacting with sensors, other ECUs, or actuators inside a car. A suitable test system for that purpose is a networked HiL simulator (Net-HiL). As depicted in Fig. 2, the Net-HiL can connect all ECUs in a car and control data exchange on the communication bus level. Hence, each connected ECU is connected to the detailed in-car electronic infrastructure and performs on the level of vehicle function.

3.5 Application

Application is the final step of the verification and validation. ECUs in a series-production status are integrated into a real car and tested on the vehicle function level in dedicated test campaigns or driving maneuvers on proving grounds. With respect to the observed behavior, ECU function parameters can be applied towards an optimal system performance. Of course, these testing activities are both costly and time consuming, and reproducibility is difficult, especially for complex driving scenarios.

4 Test Process Enhancements

ECU testing depends heavily on the availability of real ECU hardware components. Any delay in hardware development inevitably causes delays in test system operations. On one hand, test cases cannot be validated with respect to correctness, coverage, plausibility, or completeness. This holds for testing activities at all stages of verification and validation. On the other hand, system testing with Net-HiL simulators suffers from late hardware availability, as joint in-car functionality relying on data from several ECUs cannot be validated in cases where single ECUs

are not available. Actually, it can be assumed that test case validation must not be done as part of a test campaign, since valuable time of test system availability can be used more effectively once real ECU hardware is available. Likewise, the Net-HiL must not be kept in a waiting mode until a real hardware sample has been delivered for each connected ECU.

4.1 Simulation Approach

It is suggested that Soft-ECUs be employed for validation in order to decrease idle-times of test systems, which will increase test system availability for actual testing.

Soft-ECU: A Soft-ECU mimics a real ECU as a software component. Interfaces, structure, and functionality are derived from system requirements, allowing for the simulation of both physical and functional behaviors. Examples of ECU requirements to be captured by a Soft-ECU are "Message XY is sent to FlexRay every 10 ms" or "Detection of a voltage drop is written to error memory." For implementation of a Soft-ECU with communication ports, internal states, and function blocks, as depicted in Fig. 3, a model-based approach was proposed using Matlab and Simulink.

Simulation Framework: Soft-ECUs cannot be operated alone, which means a higher-level entity is required for timing, data generation, and scheduling, as shown in Fig. 4. Simulation must either be capable of test case validation by the application of HiL test cases to Soft-ECUs, or of enabling system testing by integrating Soft-ECUs in a given hardware environment with the Net-HiL.

4.2 Concept for Test Case Validation

HiL test cases for ECUs are derived from the ECU requirement specifications. To achieve a high degree of automation, various modeling paradigms are available in

Fig. 3 Structural model of a Soft-ECU
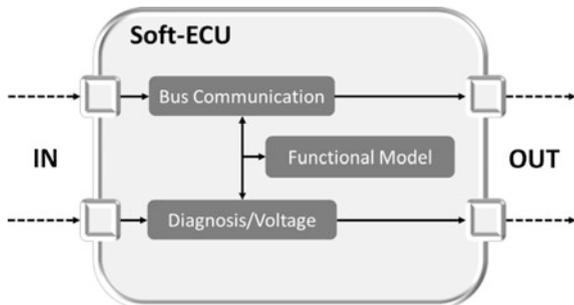
Fig. 4 Simulation framework
for Soft-ECU operation
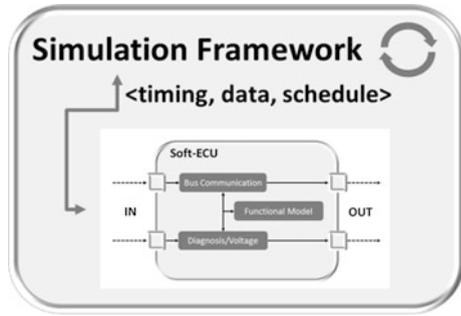
order to obtain a machine-readable test case implementation, e.g. based on Python scripts, sequence diagrams, and state charts according to the unified modeling language (UML, http://www.uml.org/). The set of test cases has to be validated with respect to completeness and correct implementation of the test goals as defined by the requirements. Instead of using valuable resources of the HiL simulator for that purpose, an "offline validation" approach is proposed, which employs the capabilities of the Soft-ECU. This process is depicted in Fig. 5.

Both test cases and Soft-ECUs are derived from ECU requirements. Applying the test cases to the Soft-ECU in the simulation framework allows for the validation of the correctness of test case implementation such as for time outs, system values, parameters, or error-handling strategies. The scope of the test cases applicable to the Soft-ECU depends on the level of detail in the Soft-ECU model. It can range from simple request/response state charts in bus communication to complex physical models that resemble thermal behaviors of real ECUs. To avoid obstructing the use
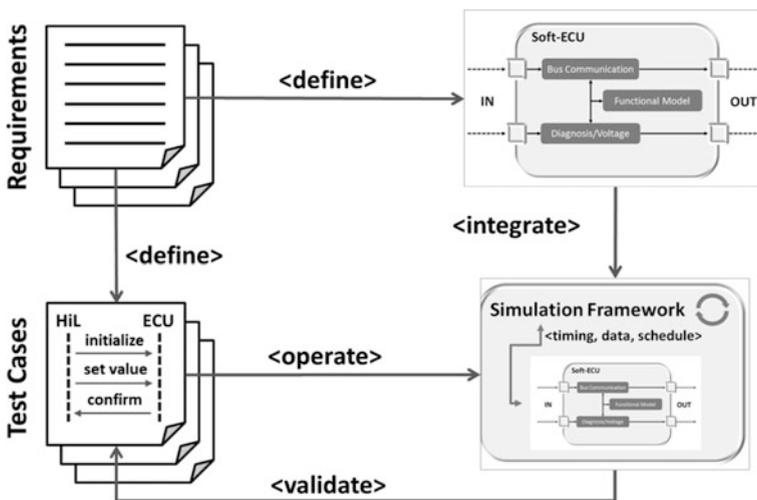
Fig. 5 Process of test case implementation and validation using Soft-ECU

of the HiL simulator for test case validation, a HiL-independent implementation for the simulation framework is currently being developed in a PhD project. Initial requirements and markets analyses have shown that a framework based on SysML appears to be appropriate due to standardized support of state-based artefacts and possible enhancements with respect to timing and scheduling towards real-time testing via profiles like MARTE (http://www.omgmarte.org/).

4.3 System Test with Networked HiL Simulator

Soft-ECUs can be employed with the Net-HiL for validation on system test levels at early stages of vehicle development, where ECUs are typically not yet available as real hardware samples. Each non-present ECU can be replaced by its counterpart Soft-ECU in order to complement the set of electronic peripherals of the Net-HiL and to enable a meaningful HiL test operation. This is illustrated in Fig. 6 for both early and late stages of vehicle development, with ECUs and Soft-ECUs shown grayed or highlighted according to availability and utilization, respectively. Since a Soft-ECU is capable of resembling functional behavior on a bus communication level, it is particularly suited to bridge the potential hardware gap at the Net-HiL.

According to the experiences of this study, control software packages of modern HiL simulators, such as ControlDesk (http://www.dspace.com/) or LabCar (http://www.etas.com/), show a high degree of compatibility with the Soft-ECUs modeled in Matlab and Simulink, and allow for a straightforward integration of Soft-ECUs into the simulator environment and test automation. Hence, HiL control software intrinsically serves as a simulation framework for Soft-ECUs, with no external standalone simulation framework at the Net-HiL required as in test case validation.
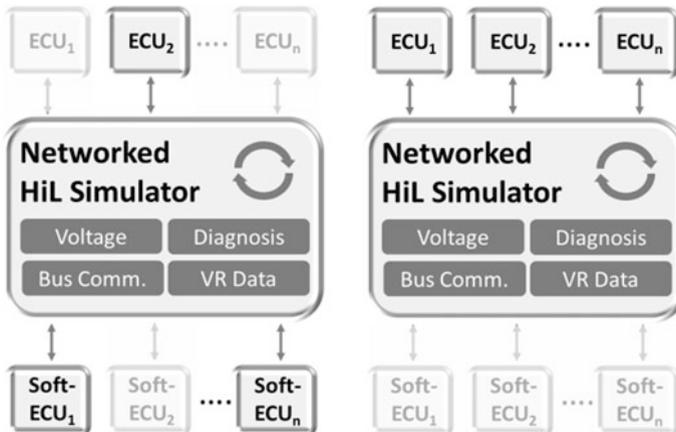
Fig. 6 ECU integration testing at networked HiL simulator using Soft-ECUs at early stage in development (*left*) and at late stage (*right*)

4.4 Analytical Results for Soft-ECU Development

During prototype ECU modeling activities for operation with a Net-HiL in a vehicle development project, the results of Soft-ECU modeling and model behavior along several revision cycles of Soft-ECUs implemented in Matlab and Simulink were determined, including the following performance measures:

- Block Count: Number of sub-blocks inside a Soft-ECU
- Structural Complexity (S): Connections to other Soft-ECUs (Coupling)
- Local Complexity (L): Work effort within one Soft-ECU (Cohesion)

The number of blocks inside a model was derived using a Matlab internal function for block counting. Structural complexity S and local complexity L were calculated according to the metrics following the approach as proposed by Card et al. [11] in Eqs. 1 and 2, where f_i remarks the fan out of module i, n is the number of modules on one layer and v_i are the I/O (input/output) variables of module i:

$$S = \frac{\sum f_i^2}{n} \tag{1}$$

$$L = \frac{\sum \frac{v_i}{f_i + 1}}{n} \tag{2}$$

Figures 7 and 8 illustrate these measures for two exemplary Soft-ECUs that were modeled for a real electronic stability control ECU (ESC) and a real engine ECU (Engine), respectively. The plots resembled modeling activities along 50 model
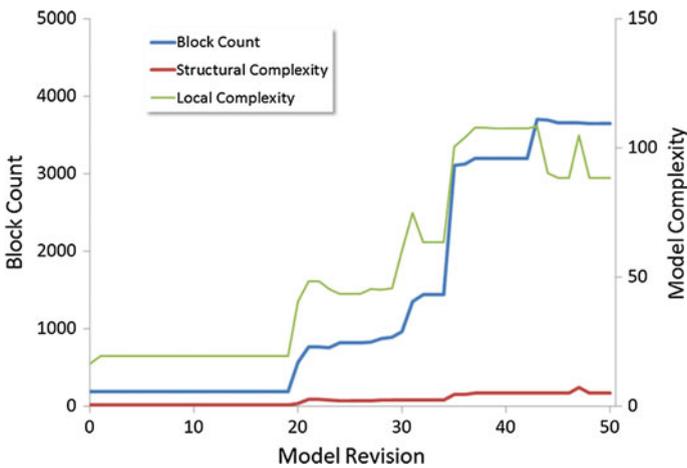
Fig. 7 Model size and complexity measures for ESC Soft-ECU along model revisions during vehicle development
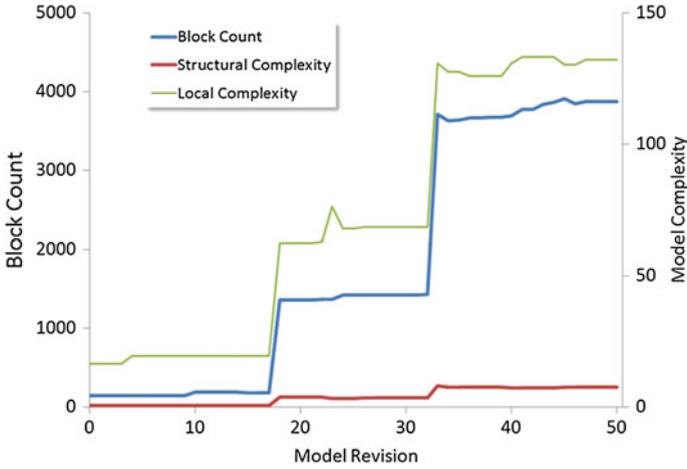
Fig. 8 Model size and complexity measures for engine Soft-ECU along model revisions during vehicle development

revision cycles and allow for an at-a-glance estimation of quantitative and qualitative parameters.

Obviously, both the extent and complexity of the models continuously increase during vehicle development, as more and more ECU system requirements are implemented in Soft-ECUs. Consequently, increased sizes yield increased requirements coverage with respect to the ECU system specifications, and increasing model complexity resembles the step-by-step extended functional scope and performance of respective Soft-ECUs. If the complexity measures exceed certain well-defined thresholds, a potential refactoring or redesign of a Soft-ECU can be initiated.

5 Conclusions and Outlook

This study proposed a simulation-based improvement for ECU testing during the early stages of vehicle development. With a model-based simulation component called Soft-ECU, the gap between rising electronics complexity and tight time schedules for verification and validation can be bridged with more effective and efficient HiL testing. Soft-ECUs were implemented in Matlab and Simulink for real ESC and engine ECUs, and quantitative and qualitative performance measures for model size and complexity were derived and presented. For test case validation, a simulator-independent simulation framework is being developed, in which a SysML-based approach shows potential for testing conformity to modeling standards and real-time capabilities. Consequently, the future work on this topic aims to accomplish implementation of the standalone simulation framework and integrate

the Soft-ECU-based validation of test cases into existing test processes. In addition, metrics, which allow for the numbering of test effort improvements, will be defined in a detailed case-study for the development of an airbag ECU. It is believed that virtualization is an indispensable part of verification and validation in the development of sound and secure automotive electronics.

References

1. Herpel, T., Hoiss, T., Sghair, A.: Hardware-in-the-loop-based systematic testing of automotive safety electronics. In: Proceedings of 4th International Conference on Electronics, Communications and Networks (CECNET2014), pp. 1425–1429. CRC Press, Boca Raton (2015)
2. Schroeder, J., Berger, C., Herpel, T.: Challenges from integration testing using interconnected hardware-in-the-loop test rigs at an automotive OEM. In: Presented at 1st International Workshop on Automotive Software Architectures, pp. 39–42, Montreal, Canada (2015)
3. Suh, M.W., Chung, J.H., Seok, C.S., Kim, Y.J.: Hardware-in-the-loop simulation for ABS based on PC. Int. J. Veh. Des. **24**(1), 157–170 (2000)
4. Schuette, H., Waeltermann, P.: Hardware-in-the-loop testing of vehicle dynamics controllers—a technical survey. Technical Report, SAE Technical Paper, 2005-01-1660 (2005)
5. Himmler, A.: From virtual testing to HIL testing—towards seamless testing. Technical Report. SAE Technical Paper, 2014-01-2165 (2014)
6. Vora, A., Wu, H., Wang, C., Qian, Y.: Development of a SIL, HIL and vehicle test-bench for model-based design and validation of hybrid powertrain control strategies. Technical Report, SAE Technical Paper, 2014-01-1906 (2014)
7. Witter, H., Heiden, M., Talwar, K.: ABS/ESP ECU testing with sophisticated HIL simulation methods. Technical Report, SAE Technical Paper, 2009-26-0079 (2009)
8. Caraceni, A., De Cristofaro, F., Ferrara, F. et al.: Benefits of using a real-time engine model during engine ECU development. Technical Report, SAE Technical Paper, 2003-01-1049 (2003)
9. Andrianarison, E., Piques, J.-D.: SysML for embedded automotive systems: a practical approach. In: 2010 Embedded Real Time Software and Systems, ERTS. http://web1.see.asso.fr/erts2010/Site/0ANDGY78/Fichier/PAPIERS%20ERTS%202010/ERTS2010_0057_final.pdf (2010)
10. International Organization for Standardization, ISO 26262: Road Vehicles—Functional Safety, October 2015
11. Card, D., Agresti, W.: Measuring software design complexity. J. Syst. Softw. **8**(3), 185–197 (1988)

Feature Extraction for Color Images

Zhen-Yu Han, Dong-Hua Gu and Qing-E Wu

Abstract The extraction of certain characteristics points such as color edge, inflection points, etc., is an imaging problem which requires urgent attention. This paper proposes a similar color segment algorithm. The algorithm is analyzed in different color distribution situations, and the extraction effect to the color is shown. Additionally, experimental analysis of the algorithms is provided. Experimental results indicate that the similar color segment algorithm performs better than existing algorithms in relation to a more obvious color edge, as it has better edge detection, stronger anti-noise ability, a faster processing speed and other advantages. Moreover, this paper compares the proposed algorithm to existing classical feature extraction algorithms.

Keywords Similar color segment · Edge detection · Corner point extraction

1 Introduction

In image retrieval, calibration, classification, and clustering, effective feature extraction from the image is vital. However, the color feature is one of the most widely-used visual features. Mahdi et al [1]. proposed a fuzzy recognition system of facial expressions, and conducted color facial expression Aydm et al [2]. Introduced an ant colony optimization method as a general color clustering analysis, and used the method to implement the classification and extraction of the area of a flower in a

Z.-Y. Han · D.-H. Gu · Q.-E. Wu (✉)
College of Electric and Information Engineering, Zhengzhou University
of Light Industry, Zhengzhou 450002, China
e-mail: wqe969699@163.com

Z.-Y. Han
e-mail: hzy196611@163.com

D.-H. Gu
e-mail: 646957326@qq.com

color image. Additionally, previous research presented an image segmentation method for an image of a flower. According to flame color and vibrational frequency analysis, Chen et al [3]. presented a digital image processing for the color of flame. For color image processing, Lissner, Gu, Kim et al [4–6]. implemented a unified color space based on perceptual image processing .

Since previous edge detection algorithms involve the operation to gradient, they are sensitive to noise and computationally intensive, among other disadvantages. In practice, this paper determined that the Susan algorithm [7] was based solely on the gray scale comparison of surrounding pixels and does not involve any gradient calculations, thus demonstrating a strong anti-noise ability and relatively small computation requirements. All of these algorithms are designed to conduct the image processing of gray scale in one dimension. However, during surgery and diagnosis [8], doctors must be able to pinpoint the color image; the classification of data must effectively process color pictures [9–13]. Feature extraction in color images requires the development of new extraction algorithms.

2 Similar Color Segment Algorithm

According to the given appropriate threshold values of edge and corner points, the extraction of characteristic points can be conducted, i.e., the extraction of edge and corner points. The edge and corner point detection algorithm is described as follows:

A point of interest or a characteristic point in a region of interest is chosen, called the core O. (1) If $R_0 > 0$ and the circle Θ O exists, in which point O is the center of the circle and R_0 is the radius, and which contains pixels of only one color, then the point O is not an edge point or a corner point, but represents a point in the flat region. (2) For any length $R > 0$, if the circle Θ O, of which the point O is the center of the circle and R is the radius, contains pixels of two or more colors, then this point O is not an edge point, a corner point, or a bifurcation point.

The hue value H, saturation value S and illuminance value I of each pixel within the circle O are compared with those of the core point O. If the corresponding differences between the hue value, saturation value and illuminance value of the pixels within the circle O and those of the core point O are all less than the corresponding given thresholds, then this point within the circle and the core point O are the same or similar. A group of pixels that satisfies such conditions and which constitutes a region are called a similar color segment (SCS).

According to the SCS detection algorithm, some detected characteristic points cannot be distinguished as an end point, a corner point or a bifurcation point. These characteristic points are assumed to be the center of circle and a series of concentric circle rings are expanded around the circle, respectively. The circle is viewed along the clockwise or counter-clockwise direction, so as to determine the number at which the image experiences a change of hue. If the center of that numbered circle

is greater than two, it represents a bifurcation point. The specific approach is described as follows. For some pixels in the neighborhood of point O, their red hue value, saturation value and illuminance value are the same or similar to those of point O; these pixels are labeled as (1). According to the same method, pixels with the same or similar blue hue are labeled as (2), the pixels with a similar green hue are labeled as (3), similar black pixels are labeled as 0, white pixels are labeled as 12, etc. The different colors are labeled by different numbers, respectively. From the beginning of a certain point on a circular ring, the distribution of pixels of any ring along the clockwise or counter-clockwise direction are analyzed, and a color change of pixels from "1" to "2", or from "2" to "3", or from "0" to "1" is recorded as a jump of color. The center of the circle in which the number of color jumps is equal to two is marked as an edge point or a corner point; however, the center of the circle that contains more than two color jumps is marked as a bifurcation point. Then, the center of the circle in which the number of color jumps is equal to n ($n > 2$) is marked as the n bifurcation point.

The specific algorithms are given as follows.

The hue value H, saturation value S and illuminance value I of each point within the circle O and those of the core point O are compared according to the following similar comparison function.

$$C_1(\vec{r}_0, \vec{r}) = \begin{cases} 1, & |H(\vec{r}) - H(\vec{r}_0)| \leq g_1 \\ 0, & |H(\vec{r}) - H(\vec{r}_0)| > g_1 \end{cases} \tag{1}$$

$$C_2(\vec{r}_0, \vec{r}) = \begin{cases} 1, & |S(\vec{r}) - S(\vec{r}_0)| \leq g_2 \\ 0, & |S(\vec{r}) - S(\vec{r}_0)| > g_2 \end{cases} \tag{2}$$

$$C_3(\vec{r}_0, \vec{r}) = \begin{cases} 1, & |I(\vec{r}) - I(\vec{r}_0)| \leq g_3 \\ 0, & |I(\vec{r}) - I(\vec{r}_0)| > g_3 \end{cases} \tag{3}$$

in which, g_1, g_2 and g_3 are the difference thresholds of the hue, saturation and illuminance, respectively, which are the threshold values used to determine the degree of similarity; the selection of g_i is determined by the comparative degree between the color and the background in an image. $\vec{r}_0$ is the position (x_0, y_0) of the current core point. $\vec{r}$ is the position (x, y) of other any point in the circle O. $H(\vec{r}_0)$ and $H(\vec{r})$, $S(\vec{r}_0)$ and $S(\vec{r})$, $I(\vec{r}_0)$ and $I(\vec{r})$ are the hue values, saturation values and illuminance values of the core point and other points in the circle O, respectively. $C_i(\vec{r}_0, \vec{r})$ is the discriminant function of pixels that belong to SCS in the circle O, and is an output value where $i = 1,2,3$.

By comparing the hue, the size of the SCS can be calculated according to the following equation:

$$n_1(\vec{r}_0) = \sum_{\vec{r} \in c(\vec{r}_0)} C_1(\vec{r}, \vec{r}_0) \tag{4}$$

Similarly, $n_2(\vec{r}_0)$ and $n_3(\vec{r}_0)$ can be obtained based on Formula (4) by comparing the saturation and the illuminance. Finally, the size of the SCS can be determined according to the following formula:

$$n(\vec{r}_0) = \min_i \{n_1(\vec{r}_0), n_2(\vec{r}_0), n_3(\vec{r}_0)\} \tag{5}$$

In the above formulas, $n(\vec{r}_0)$ is the size of the SCS of the core point $\vec{r}_0$, and $c(\vec{r}_0)$ is the circular region with $\vec{r}_0$ as the center of circle.

According to experimental analysis, the initial edge response is produced by the following formula:

$$R(\vec{r}_0) = \begin{cases} n_0 - n(\vec{r}_0) & \text{if } n(\vec{r}_0) < n_0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where n_0 is a threshold and $R(\vec{r}_0)$ is the response function.

The size of the initial edge response obtained by Formula (4) accords with a rule, i.e., that smaller SCS values produce greater initial edge responses.

3   Experimental Result Analysis and Comparison

3.1   Experiment

Here, the similar color segment algorithm is used to achieve feature extraction; the threshold values of the similarity comparison function used to detect edges and corner points are set to $g_1 = 9$, $g_2 = 11$ and $g_3 = 10$. The threshold of the initial edge response is set to $n_0 = 2382$. The threshold to detect the corner points is set to $n_{\text{angle}} = 1413$. The simulation result is shown in Fig. 1 b. Figure 1 a represents the original image.
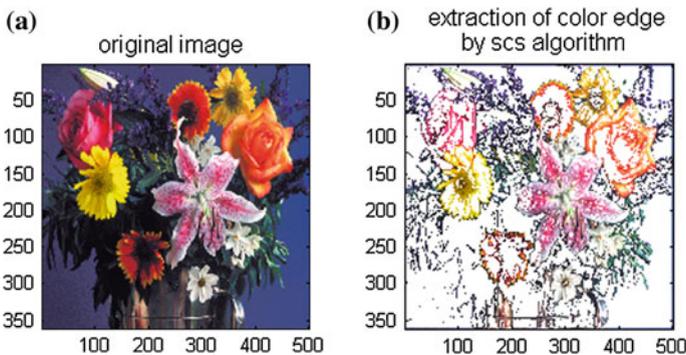
Fig. 1 Color feature extraction by similar color segment algorithm
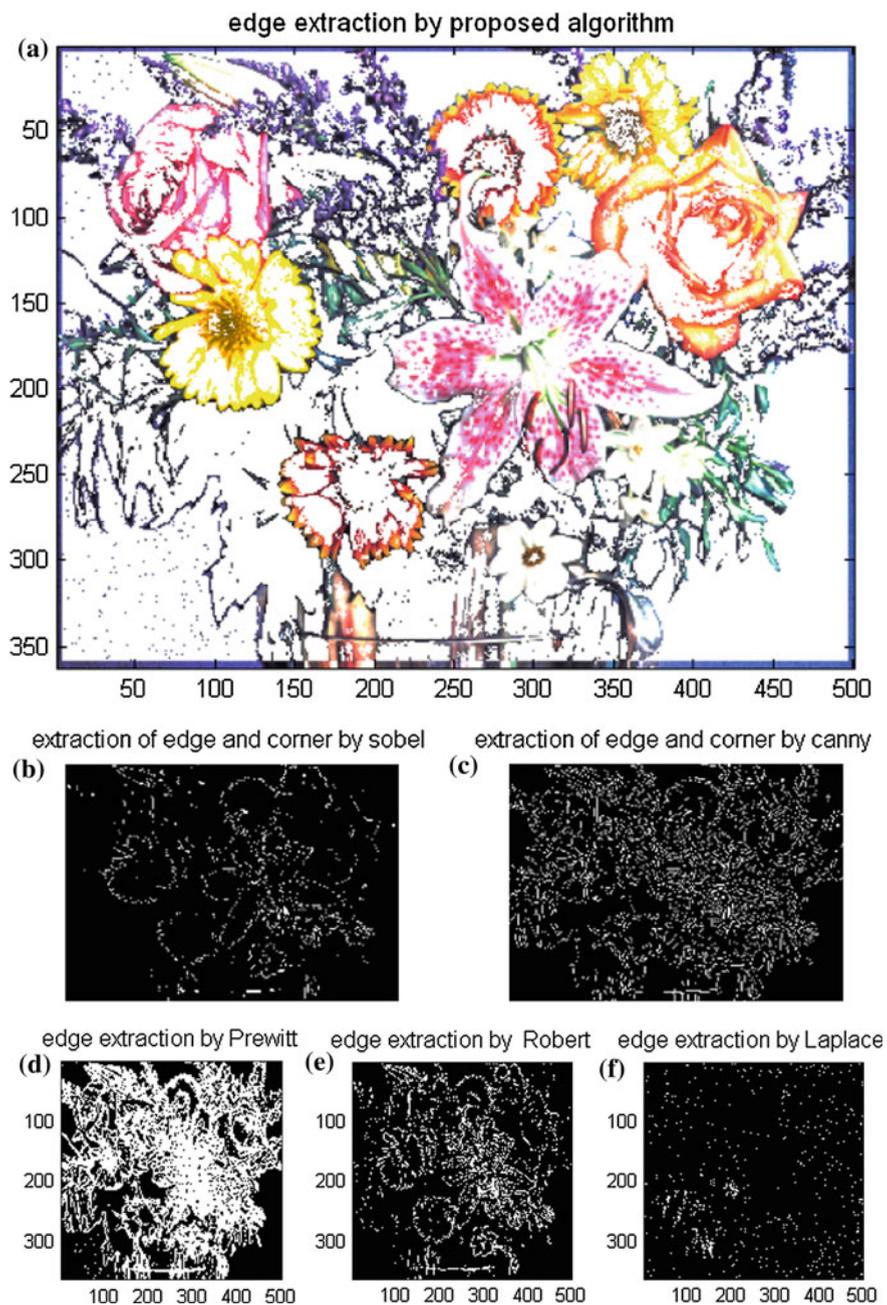
Fig. 2 Comparison of proposed and existing extraction algorithms

As shown in Fig. 1 a, b features such as the edge, corner points, color, etc., extracted by the similar color segment algorithm is more distinct, and noise filtering effect is good.

3.2 Comparison of the Proposed and Several Existing Algorithms

For the feature extraction of color images, the extraction results provided by the color feature extraction algorithm proposed in this paper and existing relevant extraction algorithms are compared. The simulation results indicate that the extraction effect of the similar color segment algorithm is not only good, but that it can also detect the edge direction information, as well as eliminate noise and exhibit a faster processing speed. For a color image with 362×500, its processing time is only 0.157 s. Moreover, this algorithm is flexible because some thresholds, such as the comparison function threshold, threshold of edge response, and threshold of corner response, can be set to different values based on the actual situation and experimental analysis. The experiment and simulation for the proposed algorithm is described in Sect. 3.1.

However, Robert, Sobel and Prewitt operators could not detect the edges of some straight lines, neither could they detect the part edges of circle. Though Gauss-Laplace and Canny operators could basically detect all the edges, their positioning effect was relatively poor, the edge pixels detected by them were wider, and the edges and corners extracted by them were all gray scale pixels. Their processing speed was also slightly slow. For a color image with 362×500, an average processing time required by the Robert, Laplace and Prewitt operators was approximately 13.969 s, that of Sobel was 0.75 s, and that of Canny was 21.61 s. Moreover, their noise elimination ability was weak. Simulation results are shown in Fig. 2 b–f; Fig. 2 a demonstrates edge extraction achieved by the proposed algorithm.

4   Conclusion

This paper studies the features of color images, presents a three-dimensntional feature extraction parameter for color images, describes the extraction algorithms of edges and corners, and provides comparison and experimental analysis of the proposed algorithm and existing extraction algorithms. Simulation results indicate that the proposed algorithm has a better extraction effect, stronger anti-noise ability, faster processing speed, more flexible use, and can detect edge direction information.

References

1. Ilbeygi, M., Shah-Hosseini, H.: A novel fuzzy facial expression recognition system based on facial feature extraction from color face images. Eng. Appl. Artif. Intell. **25**, 130–146 (2012)
2. Aydin, D., Uğur, A.: Extraction of flower regions in color images using ant colony optimization. Procedia Comput. Sci. **3**, 530–536 (2011)
3. Chen, J., Bao, Q.F.: Digital image processing based fire flame color and oscillation frequency analysis. Procedia Eng. **45**, 595–601 (2012)
4. Lissner, I., Urban, P.: Toward a unified color space for perception-based image processing. IEEE Trans. Image Process. **21**, 1153–1168 (2012)
5. Gu, J.P., Hua, L., Wu, X., Yang, H., Zhou, Z.T.: Color medical image enhancement based on adaptive equalization of intensity numbers matrix histogram. Int. J. Autom. Comput. **12**(5), 551–558 (2015)
6. Kim, H.T., Cho, K.Y., Kim, S.T., Kim, J., Jin, K.C.: Quick light mixing of multiple color sources for image acquisition using pattern search. Int. J. Precis. Eng. and Manuf. **16**(11), 2353–2358 (2015)
7. Smith, S.M.: Susan–a new approach to low level image processing. Int. J. Comput. Vis. **23**, 45–78 (1997)
8. Jukić, A., Kopriva, I., Cichocki, A.: Noninvasive diagnosis of melanoma with tensor decomposition-based feature extraction from clinical color image. Biomed. Signal Process. Control **8**, 755–763 (2013)
9. Xie, X.Z., Wu, J.T., Jing, M.G.: Fast two-stage segmentation via non-local active contours in multiscale texture feature space. Pattern Recogn. Lett. **34**, 1230–1239 (2013)
10. Atif, J., Hudelot, C., Bloch, I.: Explanatory reasoning for image understanding using formal concept analysis and description logics. IEEE Trans. Syst. Man Cybern. B Cybern. Syst. **12**, 1–19 (2013)
11. Hasanzadeh, R.P.R., Daneshvar, M.B.: A novel image noise reduction technique based on hysteresis processing. Int. J. for Light and Electron. Opt. **126**(21), 3039–3046 (2015)
12. Yuan, C., Qin, Y., Zhang, M., Zhang, H.F., Jiao, S.Y., Li, B.C.: A new method of testing and evaluating the quality of refined montan wax: digital color and gc fingerprint. Chromatographia. **78**, 1283–1292 (2015)
13. Kalra, G.S., Talwar, R., Sadawarti, H.: Adaptive digital image watermarking for color images in frequency domain. Multimed. Tools Appl. **74**, 6849–6869 (2015)

Human Gait Classification Using Doppler Motion Feature Analysis

Xi Wang, Jia-Wu He, Zhuo Chen and Gang Zhu

Abstract This paper proposes a distinguishing method based on Doppler feature analysis and extraction in human gait classification. The author first analyzes the characteristics of human gaits. Then, based on analysis of human Doppler features during walking and running, the paper extracts micro-Doppler parameters that can be used in motion identification. Finally, results are presented according to the classification of human gaits based on a support vector machine (SVM) classifier.

Keywords Feature extraction · Human gait · Doppler motion · SVM

1 Introduction

Due to the complicated battlefield circumstances and the requirement of safety protection, disaster rescue, and military medical science, the study of the radar feature of human Doppler motion is becoming increasingly more important [1]. Human body motion almost contains many types of micro-motion forms, due to the complexity of the human body structure and the strong coupling of each part of the

X. Wang (✉) · Z. Chen · G. Zhu
Department of Information Engineering, Academy of Armored Forces Engineering,
Beijing, China
e-mail: lycis@126.com

Z. Chen
e-mail: Jane2michelle@gmail.com

G. Zhu
e-mail: 18901078269@189.cn

J.-W. He
Department of Scientific Research, Academy of Armored Forces Engineering,
Beijing, China
e-mail: Jiawu_he@126.com

human; therefore, each motion pattern is a combination of many types of motion, so the characteristics of human gaits are very complex, and research regarding the classification of human gaits is of great significance.

2 Analysis of Characteristics of Human Gaits

The velocity of the body is described by four parameters [2–5]: the average speed of the human body, the speed of the trunk, the speed of the legs and the speed of the hands. The average speed of a person can be composed of an average speed s and average acceleration s, as follows:

$$v(t) = s + \dot{s}t \tag{1}$$

The speed of the trunk is slower than the average speed, described as follows:

$$v^T\left(t, z^T\right) \approx s + \dot{s}t + s^T \cos\left(2\pi * 2ft + 2\pi\varphi^T\right) \tag{2}$$

where S^T is the amplitude of sine, *and f* is the walking frequency. The trunk movement parameter can be expressed as follows:

$$z^T = \left(s^T, s, \dot{s}, f, \varphi^T\right) \tag{3}$$

Human body swing and translational velocity in the echo signal time frequency plane use different sine curves to describe all movement around the center of velocity, which are translated into a proportional relationship by the Doppler frequency axis and velocity. Arm movement is caused by the sine curve with the same frequency, but with a different amplitude and initial phase. The amplitude of the human body in different parts of the electromagnetic reflection intensity, and phase change is the representation of the movement with migration time [2]. The velocity change of the leg is similar to the sine law curve as is that of the arm, of which the amplitude is relatively large. Analysis has shown that the speed of motion of the body is not more than that of the foot. A lower leg produces the speed curve for the envelope of the entire speed of human motion, while other velocity components are located in the speed envelope curve. Therefore, the human body speed envelope curve generated by leg movement can be described as follows [1]:

$$v_U^L\left(t, z^L\right) \approx s + \dot{s}t + s^L\left|\cos\left(2\pi ft + 2\pi\varphi^L\right)\right| \tag{4}$$

$$v_L^L\left(t, z^L\right) \approx s + \dot{s}t - s^L\left|\cos\left(2\pi ft + 2\pi\varphi^L\right)\right| \tag{5}$$

where $z^L = (s, \dot{s}, f, \varphi^L)$ is the parameter vector for the lower limb.

According to the actual measurement data of the radar to the human and based on the time frequency transform, the micro-Doppler features of a human body

walking are obtained [3]. How to estimate the physical features of gait parameters during walking is shown as follows:

- The center velocity $C(t)$, which is caused by the body during walking;
- The maximum velocity $U(t)$ and minimum rate $L(t)$ represent the movement process in the swing of the lower limb and foot speed. Human body movement caused by other velocity components fall in between the maximum speed and minimum speed;
- Human movement cycle.

Micro-Doppler estimation of the target motion parameter requires that the micro-Doppler radar echo signal's instantaneous frequency is estimated with high precision, with focusing on high time-frequency resolution of time-frequency distribution and its efficient computation.

3   Micro-Doppler Parameter Extraction

Based on the above analysis, the estimation of the parameters of human limbs is key to extracting the extreme value of the micro-Doppler, and the extreme value of the micro-Doppler curve can be obtained by generalized random transform (GRT) [5–7]. The GRT is a peak detection problem in the image domain. As a method, it is insensitive to the noise effect. Therefore, it is especially suitable for the problem of extreme value estimation presented in this paper. Using GRT to detect the sine curve of the micro-motion in the frequency range, the Doppler frequency shift generated by the upper arm and the lower arm can be approximated as a sine curve, as follows:

$$c_{UA} = m_{UA} + \mu_{UA} \sin(\omega t + \varphi_{UA}) \tag{6}$$

$$c_{LA} = m_{LA} + \mu_{LA} \sin(\omega t + \varphi_{LA}) \tag{7}$$

The parameter space corresponding to the curve UA and LA is $\Gamma \triangleq \{m, \mu\}$, representing the mean and amplitude of the two sine curves corresponding to the mean and the amplitude of the signal component of the upper limb and the lower limb. The definition of GRT is as follows:

$$GRT(\Gamma) = \sum_{k=1}^{K} |S(k, \phi(k, \Gamma))| = \sum_{k=1}^{K} |S(k, d)| \tag{8}$$

The $\sum_{k=1}^{K} |S(k, \phi(k, \Gamma))|$ indicates the time-frequency distribution of the echo signal of the upper limb, $d = \phi(k, \Gamma)$ expresses the Doppler frequency determined by the parameters of the gamma curve at the K sampling time shift, $GRT(\Gamma)$ is realized in parameter space Γ determined by the numerical integral curve. When the

parameter of a point on the corresponding Γ curve in the echo time-frequency image is real, $GRT(\Gamma)$ reaches a local peak. $S(k, d)$ is the discrete data of the frequency domain. Therefore, in the calculation of $GRT(\Gamma)$ in the parameter space Γ of a point value, interpolation processing is necessary in the frequency domain, in which linear interpolation, spline interpolation and other numerical methods can be used. The simulation results indicate that the interpolation method for $GRT(\Gamma)$ calculation does not have a significant impact [5]. Based on the $GRT(\Gamma)$ on the plane of each peak detection, the corresponding parameters can be obtained by a sine curve.

4   Classification of Human Gaits Based on SVM

4.1   Estimation of Gait Parameters

Through the analysis of the above parameters, the estimation of the radar parameters can be obtained as follows [7–10].

- By filtering and removing the noise, the time frequency of the human body motion can be extracted.
- Physical features are obtained by feature extraction.
- The curves of the characteristic parameters are obtained by polynomial fitting.
- The gait parameters are obtained by estimating of the parameters of the curves.

4.2   The Process of Applying SVM Classifier

SVM was proposed by Vapnik in 1995. By maximizing the classification interval so that the structural risk is minimized, the optimal classification discriminant function can be obtained. It has a good generalizability in solving small-sample and high-dimensional nonlinear classification problems [11, 12].

In this paper, the application of the SVM classifier is described as follows :

- Prepare training datasets and test datasets for the SVM classifier;
- RBF function is used by the SVM classifier $K(x, x_i) = \exp(-\gamma \|x - x_i\|^2)$. After cross-validation, the training set is trained by the SVM classifier; C, γ and support vectors are then obtained. C is the penalty factor and γ is a parameter of the kernel function. The support vector is the sample data which directly contributes to the construction of the classification discriminant function;
- The optimal classification discriminant function is obtained according to the support vector;
- The recognition rate is obtained by testing the sample set.

4.3 Classification of Human Motion Gaits

Multi-sets of actual measurements of walking and running were captured by the human micro-motion measurement radar. This radar is a single-frequency continuous-wave system, the carrier is set to 10.48 GHz, and the sampling rate is 48 KHz. The radar range is approximately 50 m, and the radar position is fixed.

The human body is in the front of the radar, the direction of motion and the angle of the radar line of sight is not more than 30 degrees, and the pitch angle is 0 degrees. A total of 800 sets of data were captured from measured data of walking and running. A total of 600 sets of data were used as the training sets, and the remaining 200 sets were used as test sets. The feature parameters of the model are extracted from the training samples to train the SVM classifier.

Table 1 shows the characteristic values of the training sample data. Table 2 shows the accuracy of the SVM classifier in the classification of human gait motions.

Table 1 Ten sets of data in the training sample

No.	Walking				Running			
	μ_{UA}	μ_{LA}	ω	Mean	μ_{UA}	μ_{LA}	ω	Mean
1	185	143	1.21	12.1	220	190	1.34	21.5
2	201	150	1.30	12.6	259	203	1.67	23.5
3	178	139	1.22	11.0	257	210	1.52	24.2
4	150	110	0.87	10.2	234	198	1.10	23.2
5	148	107	0.95	10.9	256	220	1.32	22.0
6	156	123	1.11	11.2	230	192	1.09	20.3
7	145	123	1.15	11.8	243	197	1.34	21.0
8	160	133	1.07	10.8	265	210	1.78	23.5
9	162	128	1.09	12.3	232	187	1.40	22.6
10	167	132	1.13	12.9	240	192	1.47	21.4

Table 2 The accuracy of SVM classifier in classification of human gait motions

No.	Training set of data samples /test set of data samples	SVM classifier	
		Optimal parameters	Accuracy (%)
1	100/100	C = 1, γ = 0.0725	93
2	200/100	C = 1, γ = 0.0725	95
3	400/100	C = 1, γ = 0.0725	95
4	200/200	C = 1, γ = 0.1	93
5	400/200	C = 1, γ = 0.1	94.5
6	800/200	C = 1, γ = 0.1	95

5　Conclusions

Among walking and running gait patterns, the mean value of the frequency is the most obvious characteristic, but μ_{UA}, μ_{LA}, and ω are also effective supplement characteristics. By choosing the optimized parameters C and γ, a relatively high classification accuracy can be obtained. However, this method is not suitable for multiple people, primarily due to the non-rigidity of the body. There are many types of radar echo, which are affected and superimposed on one another. Thus, it is difficult to design a valid classification algorithm.

References

1. Lee, L., Crimson, W.E.L.: Gait analysis for recognition and classification. In: The 5th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 148–155. IEEE, Piscataway (2002)
2. Ren, L.Q., Liang, Y.H.: Preliminary studies on the basic factors of bionics. Sci. China Technol. Sci. **3**, 148–155 (2014)
3. Chen, V.C., Li, F.Y., Ho, S.S.: Micro-doppler effect in radar-phenomenon, model and simulation study. IEEE Trans. Aerosp. Electron. Syst. **42**(1), 2–21 (2006)
4. Lawrence, S., Marple, J.: Sharpening and bandwidth extrapolation techniques for radar micro-Doppler feature extraction. In: IEEE International Conference on Radar, pp. 166–170. IEEE, Piscataway, Adelaide, Australia (2003)
5. Greneker, E.F., Sylveste, V.B.: Use of the envelope detection method to detect micro-Doppler Orlando. In: The SPIE on Passive Millimeter-Wave Imaging Technology VI and Radar Sensor Technology VII.SPIE Proceedings. 5077, 167–174 (2003)
6. Zhang, X.J., Xu, C., Li, M., et al.: Sparse and low-rank coupling image segmentation model via nonconvex regularization. Int. J. Pattern Recognit Artif Intell. **23**, 234–237 (2015)
7. Albuquerque, P., Alfinito, S., Torres, C.V.: Support vector clustering for customer segmentation on mobile tv service. Commu. in Stat. – Simul. and Comput. **5**, 212–216 (2015)
8. Wu, T., Bae, M.H., Zhang, M., et al.: A prior feature SVM-MRF based method for mouse brain segmentation. Neuro Image. **23**, 234–237 (2011)
9. Kramer, K.A., Lawrence, O.H., Goldgof, D.B et al.: Fast support vector machines for continuous data. IEEE Trans. Syst. Man Cybern. Part B Cybern. Publ. IEEE Syst. Man Cybern. Soc. **4**, 478–492, (2009)
10. Luo, W.L., Moreira, L., Guedes Soares, C.: Manoeuvring simulation of catamaran by using implicit models based on support vector machines. Ocean Eng. **12**, 150–159 (2014)
11. Doumpos, M., Zopounidis, C., Golfinopoulou, V.: Additive support vector machines for pattern classification. IEEE Trans. Syst. Man Cybern. B Cybern. **775**, 1344–1350 (2015)
12. Lin, X.Q., Gao, Z., Lin, Y.M., Zhan, X.S.: Human Action Recognition Using Transformation and SVM. In : World Conference on Control, Electronics and Electrical Engineering(WCEE 2015) WIT Press, Shanghai, China (2015)

Large Predispersion for Reduction of Intrachannel Nonlinear Impairments in Strongly Dispersion-Managed Transmissions

Wen-Hua Cao

Abstract Predispersion for reduction of intrachannel nonlinear impairments in quasi-linear strongly dispersion-managed transmission system is analyzed in detail by numerical simulations. We show that for a moderate amount of predispersion, there is an optimal value at which reduction of the nonlinear impairments can be obtained, which is consistent with previous well-known predictions. However, we found improved transmission performance than that of the previous predictions can be obtained if predispersion is increased to some extent. For large predispersion, the nonlinear impairments reduce monotonically with increasing predispersion, and tend to become stable when predispersion is further increased. Thus, transmission performance can be efficiently improved by inserting a high-dispersive element, such as a chirped fiber bragg grating (CFBG), at the input end of the transmission link to broaden the signal pulses while, at the output end, using another CFBG with the opposite dispersion to recompress the signal.

Keywords Predispersion · Intrachannel nonlinear impairments · Strong dispersion management · Quasi-linear transmission

1 Introduction

Nonlinear effects such as intrachannel four-wave mixing (IFWM) and intrachannel cross-phase modulation (IXPM) are the dominant source of bit errors for high-speed long-haul quasi-linear fiber-optic transmission systems [1]. Strong dispersion management in combination with a return-to-zero (RZ) modulation format can reduce IXPM-induced timing jitter, but cannot reduce IFWM-induced "ghost

W.-H. Cao (✉)
College of Electronic Science and Technology, Shenzhen University,
Shenzhen 518060, Guangdong, China
e-mail: wcao@szu.edu.cn

pulses" in the signal 0's and amplitude jitter in the signal 1 s [2–4]. The existing techniques for reducing the impact of IFWM include: (1) Appropriate predispersion [3, 5–7], i.e., by optimizing the amount of predispersion, minimizing the path-averaged pulsewidth over the high power nonlinear sections of the system, overlapping of neighboring pulses is reduced and hence, distortion duo to intra-channel nonlinear effects are minimized; (2) Finding a proper dispersion mapping [8–10], which involves precise dispersion management and power control to achieve zero net dispersion and negligible nonlinear penalty at the end of the link, and hence, incurs massive effort in link construction and management; (3) Appropriate phase modulation [11–19]. The schemes rely on the fact that the phase of the IFWM-induced ghost pulse has a fixed relationship to the phases of the genuine 0's in the bit stream that enter the IFWM process. Thus, the strongest ghosts can be eliminated by tailoring the phase of the surrounding 1's to cause destructive interference between the various contributions to a ghost. However, for a random bit stream, no phase modulation format can achieve destructive inter-ference in several consecutive time slots; (4) Using constrained codes (also known as modulation codes or line codes) [20–25] in which the data are coded in such a way that the worst sequences, such as a 0 bit surrounded by many 1's, are simply avoided by bit stuffing. However, this will introduce an overhead; (5) Optical phase conjugation (OPC) [26–29], which allows nonlinearity cancellation, provided that the system's power profile and dispersion are symmetrical with respect to midspan. In real systems, these conditions can only be partially fulfilled; and (6) Coherent detection in combination with digital signal processing [30–33]. The drawback of the method is that reconstruction of the transmitted data from the received signal relies on computationally intensive techniques, and real-time digital signal pro-cessing is still a challenging problem. IFWM can also be suppressed by subchannel multiplexing [34], alternation of the polarization of consecutive pulses [35], introducing some polarization mode dispersion in transmitters [36], and using unequally spaced data pulses [37]. However, the realizations of these techniques are always complex and much expensive when the bit rate is up to 40 Gbps.

In this paper, we numerically show that intrachannel nonlinear effects such as IFWM and IXPM can be efficiently reduced by introducing a large predispersion to broaden the input pulses (signal 1's) through a high-dispersive device such as a chirped fiber bragg grating (CFBG) and then recompressing the pulses at the output end through another device with the opposite amount of dispersion. We note that Turitsyn et al. proposed the same idea for coherent communication systems with the aim of simplifying the following digital signal processing of nonlinear impairments [33], but they didn't consider IFWM in the model. Here, we apply the method to a return-to zero on-off keying (RZ-OOK) intensity-modulation direct-detection (IMDD) transmission system including intrachannel nonlinear effects such as IFWM, IXPM, and self-phase modulation (SPM) We should also stress that this method is not the standard well-known predispersion technique [3, 5–7], which will be analyzed subsequently through numerical simulations.
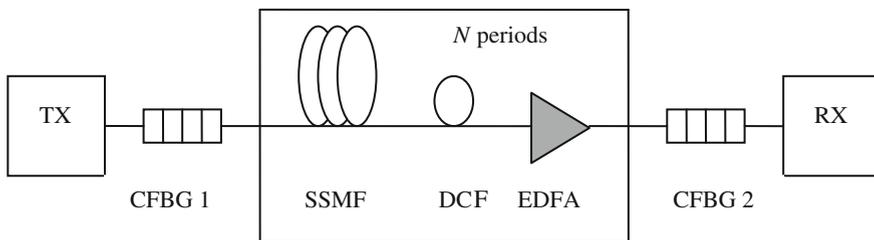
Fig. 1 Schematic diagram of quasi-linear strongly dispersion-managed transmission

2 System Configuration

Numerical simulation was performed for the transmission link schematically shown in Fig. 1. Ultrashort pulses (representing the signal 1's) from the optical transmitter were first stretched by CFBG1 with large anomalous dispersion and then entered a periodic transmission link. Each section of the link comprises a standard single-mode fiber (SSMF), a dispersion-compensating fiber (DCF), and an erbium-doped fiber amplifier (EDFA). The SSMF has the same sign of group-velocity dispersion (GVD) as that of CFBG1 so that compression of the stretched pulse in the SSMF could be prevented. We assume that the DCF can exactly compensate for chromatic dispersion of the SSMF so that the average dispersion of the transmission link is zero. The EDFA exactly compensates for energy loss produced by the SSMF and the DCF. After transmission, another CFBG2 (having the opposite dispersion of CFBG1) is used to recompress the broadened pulses.

3 Basic Equations

The simulation is based on the generalized nonlinear Schrödinger (NLS) equation which takes the form

$$i\frac{\partial u}{\partial \xi} \pm \frac{1}{2}\frac{\partial^2 u}{\partial \tau^2} + |u|^2 u = -\frac{i}{2}\Gamma u + \frac{i}{2}\mu u, \qquad (1)$$

where ξ, τ, and u are, respectively, the normalized distance, time, and field envelope in soliton units. The parameters Γ and μ account for, respectively, the fiber loss and the gain of the EDFA. The second term on the left side represents GVD where the sign "+" or "−" is chosen, respectively, when the field is transmitted in the SSMF (anomalous GVD) or the DCF (normal GVD). The third term on left side represents the Kerr nonlinearity.

Although a pseudo-random bit stream as an input is essential for accurate description of signal transmission in a realistic system, considerable physical insight could be gained with a limited number of input bits [3, 37] when we focus our attention solely on the suppression of the nonlinearities. Here, the input is assumed to be the sum of four bits in the form

$$
\begin{aligned}
u(0, \tau) &= u_1(0, \tau + 3q_0) + u_2(0, \tau + q_0) + u_3(0, \tau - q_0) + u_4(0, \tau - 3q_0) \\
&= A_1 \operatorname{sech}(0, \tau + 3q_0) + A_2 \operatorname{sech}(0, \tau + q_0) \\
&\quad + A_3 \operatorname{sech}(0, \tau - q_0) + A_4 \operatorname{sech}(0, \tau - 3q_0)
\end{aligned}
\qquad , \quad (2)
$$

where $2q_0$ represents the duration of the bit slot and A_j ($j = 1, 2, 3, 4$) represents the initial peak amplitude of the jth bit. We assume that all "1" bits have the same initial width and the same initial peak amplitude and, that the "0" bits have a much smaller peak amplitude than that of the "1" bits. Substituting the input into Eq. (1), we obtain the following set of four coupled equations:

$$
\begin{aligned}
&i\frac{\partial u_1}{\partial \xi} \pm \frac{1}{2}\frac{\partial^2 u_1}{\partial \tau^2} + \frac{i}{2}\Gamma u_1 - \frac{i}{2}\mu u_1 \\
&= -\left(|u_1|^2 + 2|u_2|^2 + 2|u_3|^2 + 2|u_4|^2\right) u_1 - u_2^2 u_3^* - 2u_2 u_3 u_4^*
\end{aligned}
\qquad , \quad (3)
$$

$$
\begin{aligned}
&i\frac{\partial u_2}{\partial \xi} \pm \frac{1}{2}\frac{\partial^2 u_2}{\partial \tau^2} + \frac{i}{2}\Gamma u_2 - \frac{i}{2}\mu u_2 \\
&= -\left(|u_2|^2 + 2|u_1|^2 + 2|u_3|^2 + 2|u_4|^2\right) u_2 - u_3^2 u_4^* - 2u_1 u_3 u_2^* - 2u_1 u_4 u_3^*
\end{aligned}
\qquad , \quad (4)
$$

$$
\begin{aligned}
&i\frac{\partial u_3}{\partial \xi} \pm \frac{1}{2}\frac{\partial^2 u_3}{\partial \tau^2} + \frac{i}{2}\Gamma u_3 - \frac{i}{2}\mu u_3 \\
&= -\left(|u_3|^2 + 2|u_1|^2 + 2|u_2|^2 + 2|u_4|^2\right) u_3 - u_2^2 u_1^* - 2u_1 u_4 u_2^* - 2u_2 u_4 u_3^*
\end{aligned}
\qquad , \quad (5)
$$

$$
\begin{aligned}
&i\frac{\partial u_4}{\partial \xi} \pm \frac{1}{2}\frac{\partial^2 u_4}{\partial \tau^2} + \frac{i}{2}\Gamma u_4 - \frac{i}{2}\mu u_4 \\
&= -\left(|u_4|^2 + 2|u_1|^2 + 2|u_2|^2 + 2|u_3|^2\right) u_4 - u_3^2 u_2^* - 2u_2 u_3 u_1^*
\end{aligned}
\qquad . \quad (6)
$$

In real parameters,

$$
\xi = \frac{z}{L_D} = \frac{z|\beta_2|}{T_0^2}, \ \tau = \frac{t - z/v_g}{T_0}, \ \Gamma = \alpha L_D = \frac{\alpha T_0^2}{|\beta_2|}, \ \mu = (g_0 - \alpha)L_D \qquad (7)
$$

where z, t, v_g represent, respectively, distance, time, and group velocity. T_0 is the half-width (at $1/e$-intensity point) of the input "1" bit, β_2 is the GVD coefficient, α is the attenuation constant of the fiber, g_0 is the unsaturated gain parameter of the EDFA, and $L_D = T_0^2/|\beta_2|$ is the dispersion length. We do not include the Raman

self-scattering (RSS) and self-steepening effects because for quasi-linear strongly dispersion-managed transmissions, the path-averaged bit width is very large and peak power is very low. Equations (3)–(6) and their modified versions can be used to describe bits transmission in the SSMF, DCF, EDFA and CFBG. The differences are: for transmission in the SSMF and the DCF, the parameter μ is zero; for transmission in the EDFA, the loss term and even the nonlinear terms can be neglected; and for transmission in the CFBG1 and CFBG2, only the GVD term is included. In real parameters, the relationship between the initial peak amplitude A_j of the jth bit in Eq. (2) and its real peak power is given by

$$A_j^2 = \frac{\gamma P_j T_0^2}{|\beta_2|} \tag{8}$$

where γ is the nonlinearity coefficient of the fiber.

4 Simulation Results and Discussion

4.1 Without Predispersion

First, let's see how intrachannel nonlinearities affect signal propagation without predispersion. We consider propagations of different bit patterns such as 0110, 1110, 1010 and 1001. In all cases, the transmission link is fixed and the same as shown by Fig. 1, except that CFBG1 and CFBG2 are removed. Each section of the link comprises a 72.77 km long SSMF with a GVD coefficient of $\beta_2 = -20$ ps^2/km near 1.55 μm, a 7.277 km long DCF with $(\beta_2)_{DCF} = 200$ ps^2/km, and an EDFA. The DCF can exactly compensate for the chromatic dispersion of the SSMF so that the average chromatic dispersion of the transmission link is zero. The EDFA exactly compensates for the energy loss caused by the SSMF and the DCF. The DCF is assumed to have the same attenuation constant and the same nonlinearity coefficient as those of the SSMF with $\alpha = 0.046$ km^{-1} and $\gamma = 1.3$ W^{-1} km^{-1}. Actually, since the DCF is much shorter than the SSMF, neglecting the difference of the parameter α or γ between the SSMF and the DCF will have a negligible influence on the simulation result.

Each input bit pattern is assumed to be in the form of Eq. (2), where all the "1" bits have a same initial width of $T_{FWHM} = 3$ ps ($T_0 = T_{FWHM}/1.763 \approx 1.7$ ps) and a constant initial peak amplitude of $A = 0.3$, which, according to Eq. (8), corresponds to a peak power of 478 mW. All the "0" bits have the same initial width and shape as those of the "1" bits, but with much smaller peak power of 4.78 mW ($A = 0.03$). The initial separation $2q_0$ between two adjacent bits is 12.5 ps, representing a bit rate of 80 Gb/s.

Figure 2 compares the output (transmitted) with the input in both shape (left column) and spectrum (right column) for bit patterns of 0110, 1110, 1010, and 1001. In all cases, the transmission distance is fixed at 160 km, which is equal to 2
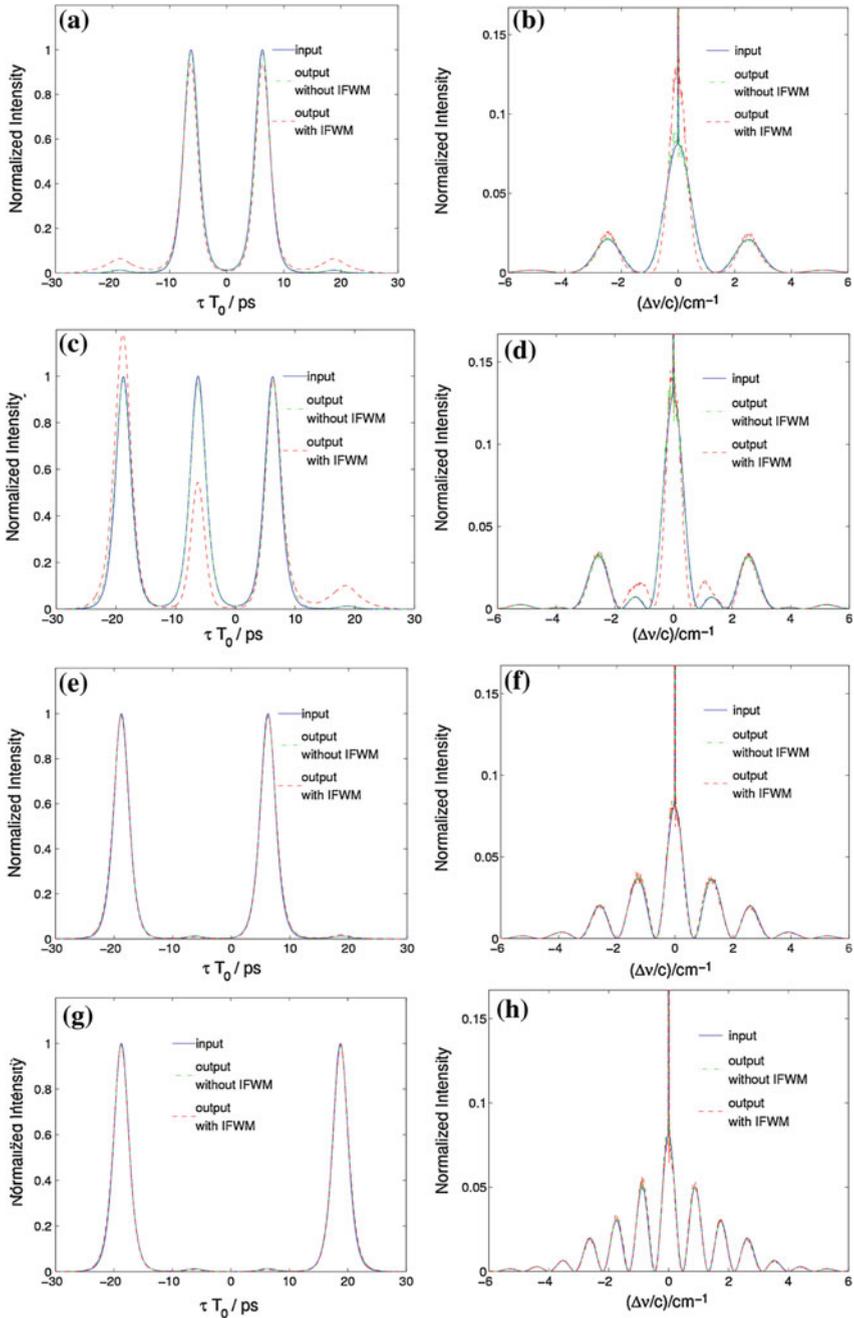
Fig. 2 Pulse shapes (left column) and spectra (right column) of four bit patterns. *Blue solid curve* represents the input, *green dashed-dotted curve* represents the output with IXPM, SPM, and GVD but without IFWM, and *red dashed curve* represents the output with IFWM IXPM, SPM, and GVD

transmission periods as shown by Fig. 1, where the blue solid curve represents the input, the green dashed-dotted curve represents the output with IXPM, SPM, and GVD but without IFWM, and the red dashed curve represents the output with all the mentioned effects. The intensities of the pulse shape and spectrum are normalized with respect to the input signal.

We see that the influence of nonlinearities depends on bit patterns. The patterns like 0110 and 1110 which contain consecutive "1s" manifest large distortion both in shape and spectrum as shown by Fig. 2a–d, while the distortion is very small in the case of non-consecutive "1s" such as 1010 and 1001 as shown by Fig. 2e–h. This is the so-called patterning effect, which is due to nonlinear interactions between neighbouring bits. In the cases of patterns 0110 and 1110, we clearly see the IFWM-induced "ghost pulses" in the "0" time slots, the IFWM-induced amplitude jitter, and the IXPM-induced timing jitter of the "1s". However, as compared to IXPM or SPM, IFWM has a much bigger effect on the transmission result as shown by Fig. 2a–d, where the green dashed-dotted curve represents the output with IXPM, SPM, and GVD but without IFWM, and the red dashed curve represents the output with all the mentioned effects. This means that IFWM is the dominant intrachannel nonlinear effect in quasi-linear strongly dispersion-managed transmission systems.

4.2 With Predispersion

Now let's see how large predispersion can reduce intrachannel nonlinearities. The idea is based on the fact that [38] IXPM and SPM are proportional to the peak power of the pulse, while IFWM is proportional to the third power of the peak power of the pulse. So, by introducing large predispersion, the peak power of the pulses can be efficiently reduced, and the intrachannel nonlinearities can be suppressed. Although large predispersion results in a wide range of overlapping of transmitted pulses, it is shown that [39] IFWM components induced between adjacent pulses are greatly larger than those between distant pulses. Consequently, the effects of the IFWM components induced between distant pulses can be neglected. Indeed, Turitsyn et al. [33] proposed the same idea of large predispersion for coherent communication systems with the aim of simplification of the following digital signal processing of nonlinear impairments, but they didn't consider IFWM in the model. Here, we apply the idea to an RZ-OOK IMDD transmission link. Our model includes intrachannel nonlinear effects such as IFWM, IXPM, and SPM. It should be emphasized that the large predispersion method used here is not the standard well-known predispersion technique [3, 5–7], which will be analyzed subsequently.

CFBG of several centimeters can provide very large normal or anomalous dispersion with GVD coefficient of β_2^g as large as 10^7 ps^2/km. In the transmission scheme as shown by Fig. 1, the CFBG1 and CFBG2 are assumed to have the same length, and the GVD coefficients of CFBG1 and CFBG2 are assumed to be,

Fig. 3 Comparison between the transmission results with (*green dashed-dotted* and *red dashed*) and without (*black dotted*) predispersion. The input (*blue solid*) and the transmission distance are identical to those of Fig. 2c, d
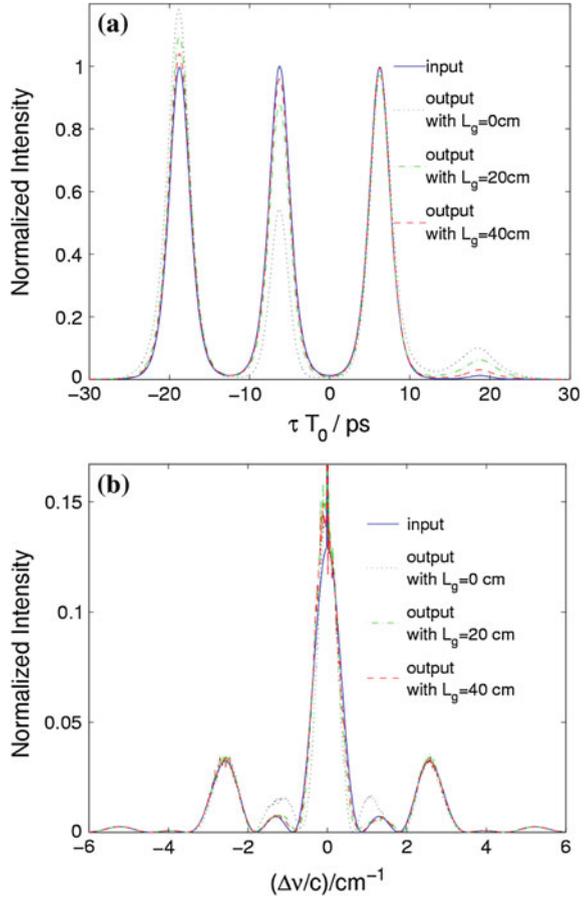
respectively, -1.82×10^6 and 1.82×10^6 ps^2/km. The transmission link here is fixed and is identical to that used for simulation of Fig. 2, except that CFBG1 and CFBG2 are used as shown in Fig. 1.

Figure 3 compares the transmission results when predispersion is introduced or not. In all cases, nonlinear effects such as IFWM, IXPM, and SPM are all considered. The blue solid line represents the input bit pattern 1110 which is identical to that of Fig. 2c, d. The black dotted line represents the output without predispersion, and the green dashed-dotted and red dashed lines show the outputs with two different predispersions corresponding to CFBG1 length L_g of 20 and 40 cm, respectively. Using the parameter β_2^g assumed above, the lengths L_g of 20 and 40 cm correspond to predispersions ($\beta_2^g L_g$) of -364 and -728 ps^2, respectively. It should be stressed that we chose anomalous dispersion for CFBG1 in order that both the CFBG1 and the transmission fiber SSMF have the same dispersion nature

Fig. 4 Peak intensity jitter of the "1" bits and the relative peak intensity of the ghost pulse as a function of the CFBG1 length. The input and the transmission distance are identical to those of Fig. 3
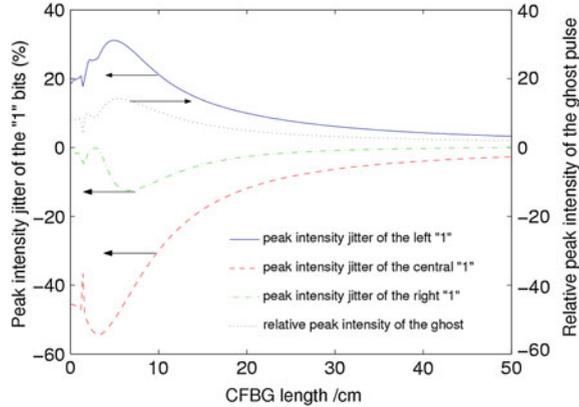
so that the pulses are broadened monotonically and pulse compression can be prevented in the SSMF.

Figure 3 shows that the longer the CFBG1 is, the better the transmission performance is, indicating that large predispersion can suppress intrachannel nonlinearities. Good performance can be achieved when $L_g = 40$ cm as shown by the red dashed line, where the peak intensity jitter (which will be defined later) of the "1" bits is less than 3 %, and the peak intensity of the ghost pulse relative to the input is less than 2 %.

Figure 4 shows in detail the peak intensity jitter of the "1" bits and the relative peak intensity of the ghost pulse as a function of the CFBG1 length, where the transmission link and the input bit pattern are identical to those of Fig. 3. The peak intensity jitter is defined as the difference between the peak intensity of the output "1" and that of the input "1", relative to that of the input "1". The relative peak intensity of the ghost pulse is defined as the peak intensity of the output ghost pulse relative to that of the input ghost pulse (we assumed earlier that the "0" bits have a much smaller peak amplitude than that of the "1" bits). Note that the peak intensity jitter of the left, central, and right "1" of the bit pattern 1110 is depicted by the blue solid, red dashed, and green dashed-dotted lines, respectively.

An interesting feature of Fig. 4 is that neither the peak intensity jitter of the "1" bits nor the relative peak intensity of the ghost pulse decreases monotonously, as expected with the increasing CFBG1 length (L_g), but reveals a minimum at L_g of about 1.4 cm and then reaches a maximum at L_g of about 5 cm. After the maximum, the peak intensity jitter and the relative peak intensity of the ghost pulse decreases monotonously with the increasing L_g. These features can be explained as follows.

In each section of the transmission link, although the DCF is assumed to exactly compensate for dispersion of the SSMF so that the average dispersion of the transmission link is zero, the local dispersion of the SSMF is large ($\beta_2 = -20$ ps^2/km) and the interplay between the nonlinearities and the local dispersion causes accumulated distortion of both the pulse shape and the spectrum when predispersion is not introduced (i.e., $L_g = 0$). However, if we insert a CFBG1 with an

appropriate amount of anomalous dispersion at the input end, the average dispersion of the transmission link (except for the CFBG2) is not zero but is anomalous, and it is this anomalous dispersion that tends to counteract the SPM effect, minimizing the path-averaged pulsewidth, reducing the overlap of neighboring pulses and suppressing intrachannel nonlinearities to an extent. This corresponds to the minimum point (at L_g of about 1.4 cm) of Fig. 4, which is just the idea of the previous well-known predispersion technique [3, 5–7].

When L_g is further increased, the overlap between the neighboring bits is enhanced as soon as the bits enter into the SSMF, and the nonlinear interactions among the adjacent bits are also enhanced. However, when L_g exceeds certain value, the bits are broadened so much that the peak powers of the "1" bits are very low when they enter the SSMF, which tends to suppress the nonlinearities because, as stated earlier, IXPM and SPM are proportional to the peak power of the bits, while IFWM is proportional to the third power of the peak power of the bits. So, on the one hand, a large-scale overlapping of neighboring bits tends to enhance the nonlinear interaction among the adjacent bits, but on the other hand, large predispersion decreases the peak power of the "1" bits and tends to suppress the nonlinearities, resulting in the maximums of Fig. 4.

When L_g reaches a certain extent, both the peak intensity jitter and the relative peak intensity of the ghost pulse tend to stabilize with increasing L_g, indicating that excessive predispersion doesn't help to suppress nonlinearities. This can be understood as follows: when predispersion is large enough, the energy is nearly well-distributed among all the bit slots; thus, transmission performance depends only on the average input power rather than the bit pattern.

5 Conclusion

Predispersion for reduction of intrachannel nonlinear impairments in quasi-linear strongly dispersion-managed transmission systems is analyzed in detail by numerical simulations. We have shown that in comparison with the previous well-known predispersion technique in which a small amount of predispersion is usually used, large predispersion can allow more effective suppression of intrachannel nonlinear effects such as IFWM and IXPM. When predispersion is increased to some extent, the nonlinear impairments of the signal reduce monotonically with increasing predispersion, then tend to be stabilized when predispersion is large enough. This will provide a simple and low-cost way to improve the transmission performance of high-speed long-haul quasi-linear fiber-optic transmission systems.

References

1. Agrawal, G.P.: Applications of Nonlinear Fiber Optics, 2nd edn. Academic Press, New York (2008)
2. Mamyshev, P.V., Mamysheva, N.A.: Pulse-overlapped dispersion-managed data transmission and intrachannel four-wave mixing. Opt. Lett. **24**, 1454–1456 (1999)
3. Johannisson, P., Anderson, D., Berntson, A., Martensson, J.: Generation and dynamics of ghost pulses in strongly dispersion-managed fiber-optic communication systems. Opt. Lett. **26**, 1227–1229 (2001)
4. Lefrançois, M., Barnasson, E., Charlet, G., Antona, J.C., Bigo, S.: Numerical discrimination of intrachannel cross-phase modulation and intrachannel four-wave mixing and their respective effect on 40 Gbit/s transmissions. Opt. Lett. **31**, 432–434 (2006)
5. Killey, R.I., Thiele, H.J., Mikhailov, V., Bayvel, P.: Reduction of intrachannel nonlinear distortion in 40-Gb/s WDM transmission over standard fiber. IEEE Photon. Technol. Lett. **12**, 1624–1626 (2000)
6. Mecozzi, A., Clausen, C.B., Shtaif, M.: System impact of intra-channel nonlinear effects in highly dispersed optical pulse transmission. IEEE Photon. Technol. Lett. **12**, 1633–1635 (2000)
7. Kumar, S., Mauro, J.C., Raghavan, S., Chowdhury, D.Q.: Intrachannel nonlinear penalties in dispersion-managed transmission systems. IEEE J. Sel. Top. Quant. Electron. **8**, 626–631 (2002)
8. Mecozzi, A., Clausen, C.B., Shtaif, M., Park, S.G., Gnauck, A.H.: Cancellation of timing and amplitude jitter in symmetric links using highly dispersed pulses. IEEE Photon. Technol. Lett. **13**, 445–447 (2001)
9. Striegler, A.G., Schmauss, B.: Compensation of intrachannel effects in symmetric dispersion-managed transmission systems. J. Lightwave Technol. **22**, 1877–1882 (2004)
10. Wei, H., Plant, D.V.: Intra-channel nonlinearity compensation with scaled translational symmetry. Opt. Express **12**, 4282–4296 (2004)
11. Liu, X., Wei, X., Gnauck, A.H., Xu, C., Wickham, L.K.: Suppression of intrachannel four-wave-mixing–induced ghost pulses in high-speed transmissions by phase inversion between adjacent marker blocks. Opt. Lett. **27**, 1177–1179 (2002)
12. Winzer, P.J., Gnauck, A.H., Raybon, G., Chandrasekhar, S., Su, S.Y., Leuthold, J.: 40-Gb/s return-to-zero alternate-mark-inversion (RZ-AMI) transmission over 2000 km. IEEE Photon. Technol. Lett. **15**, 766–768 (2003)
13. Kanaev, A.V., Luther, G.G., Kovanis, V., Bickham, S.R., Conradi, J.: Ghost-pulse generation suppression in phase-modulated 40-Gb/s RZ transmission. J. Lightwave Technol. **21**, 1486–1489 (2003)
14. Gill, D.M., Liu, X., Wei, X., Banerjee, S., Su, W.: $\pi/2$ alternate-phase ON–OFF keyed 40-Gb/s transmission on standard single-mode fiber. IEEE Photon. Technol. Lett. **15**, 1776–1778 (2003)
15. Appathurai, S., Mikhailov, V., Killey, R.I., Bayvel, P.: Investigation of the optimum alternate-phase RZ modulation format and its effectiveness in the suppression of intrachannel nonlinear distortion in 40-Gbit/s transmission over standard single-mode fiber. IEEE J. Sel. Top. Quant. Electron. **10**, 239–249 (2004)
16. Alic, N., Fainman, Y.: Data-dependent phase coding for suppression of ghost pulses in optical fibers. IEEE Photon. Technol. Lett. **16**, 1212–1214 (2004)
17. Zou, M., Chen, M., Xie, S.: Suppression of ghost pulses in 40 Gbps optical transmission systems with fixed-pattern phase modulation. Opt. Express **13**, 2251–2255 (2005)
18. Batshon, H.G., Djordjevic, I.B., Vasic, B.V.: An improved technique for suppression of intrachannel four-wave mixing in 40-Gb/s optical transmission systems. IEEE Photon. Technol. Lett. **19**, 67–69 (2007)
19. Forzati, M., Berntson, A., Martensson, J., Djupsjobacka, A.: Asynchronous phase modulation for the suppression of IFWM. J. Lightwave Technol. **25**, 2969–2975 (2007)

20. Shapiro, E.G., Fedoruk, M.P., Turitsyn, S.K., Shafarenko, A.: Reduction of nonlinear intrachannel effects by channel asymmetry in transmission lines with strong bit overlapping. IEEE Photon. Technol. Lett. **15**, 1473–1475 (2003)
21. Vasic, B., Rao, V.S., Djordjevic, I.B., Kostuk, R.K., Gabitov, I.: Ghost pulse reduction in 40 Gb/s systems using line coding. IEEE Photon. Technol. Lett. **16**, 1784–1786 (2004)
22. Djordjevic, I.B., Chilappagari, S.K., Vasic, B.: Suppression of intrachannel nonlinear effects using pseudoternary constrained codes. J. Lightwave Technol. **24**, 769–774 (2006)
23. Shafarenko, A., Turitsyn, K.S., Turitsyn, S.K.: Information-theory analysis of skewed coding for suppression of pattern-dependent errors in digital communications. IEEE Trans. Commun. **55**, 237–241 (2007)
24. Slater, B., Boscolo, S., Shafarenko, A., Turitsyn, S.K.: Mitigation of patterning effects at 40 Gbits/s by skewed channel pre-encoding. J. Opt. Netw. **6**, 984–990 (2007)
25. Shafarenko, A., Skidin, A., Turitsyn, S.K.: Weakly-constrained codes for suppression of patterning effects in digital communications. IEEE Trans. Commun. **58**, 2845–2854 (2010)
26. Liu, X., Chandrasekhar, S., Winzer, P.J., Tkach, R.W., Chraplyvy, A.R.: Fiber-nonlinearity-tolerant superchannel transmission via nonlinear noise squeezing and generalized phase-conjugated twin waves. J. Lightwave Technol. **32**, 766–775 (2014)
27. Yim, H.B., Lee, S.R.: Compensation characteristics of the distorted WDM signals depending on the variation degree of RDPS random variables in the optical links with dispersion management and optical phase conjugation. Int. J. Control Autom. **7**, 1–8 (2014)
28. Eliasson, H., Johannisson, P., Karlsson, M., Andrekson, P.A.: Mitigation of nonlinearities using conjugate data repetition. Opt. Express **23**, 2392–2402 (2015)
29. Kuo, B.P.P., Myslivets, E., Wiberg, A.O.J., Zlatanovic, S., Bres, C.S., Moro, S., Gholami, F., Peric, A., Alic, N., Radic, S.: Transmission of 640-Gb/s RZ-OOK channel over 100-km SSMF by wavelength-transparent conjugation. J. Lightwave Technol. **29**, 516–523 (2011)
30. Ip, E.M., Kahn, J.M.: Fiber impairment compensation using coherent detection and digital signal processing. J. Lightwave Technol. **28**, 502–519 (2010)
31. Millar, D.S., Makovejs, S., Behrens, C., Hellerbrand, S., Killey, R.I., Bayvel, P., Savory, S.J.: Mitigation of fiber nonlinearity using a digital coherent receiver. IEEE J. Sel. Top. Quant. Electron. **16**, 1217–1226 (2010)
32. Kumar, S., Shao, J., Liang, X.: Impulse response of nonlinear schrodinger equation and its implications for pre-dispersed fiber-optic communication systems. Opt. Express **22**, 32282–32292 (2014)
33. Turitsyn, S., Sorokina, M., Derevyanko, S.: Dispersion-dominated nonlinear fiber-optic channel. Opt. Lett. **37**, 2931–2933 (2012)
34. Zweck, J., Menyuk, C.R.: Analysis of four-wave mixing between pulses in high-data-rate quasi-linear subchannel-multiplexed systems. Opt. Lett. **27**, 1235–1237 (2002)
35. Shiner, A.D., Reimer, M., Borowiec, A., Gharan, S.O., Gaudette, J., Mehta, P., Charlton, D., Roberts, K., O'Sullivan, M.: Demonstration of an 8-dimensional modulation format with reduced inter-channel nonlinearities in a polarization multiplexed coherent system. Opt. Express **22**, 20366–20374 (2014)
36. Möller, L., Su, Y., Raybon, G., Liu, X.: Polarization mode dispersion supported transmission in 40 Gb/s long Haul systems. IEEE Photon. Technol. Lett. **15**, 335–337 (2003)
37. Kumar, S.: Intrachannel four-wave mixing in dispersion managed RZ systems. IEEE Photon. Technol. Lett. **13**, 800–802 (2001)
38. Essiambre, R.J., Raybon, G., Mikkelsen, B.: Optical fiber telecommunication, vol. 4B. Academic Press, San Diego, CA (2002)
39. Ablowitz, M.J., Hirooka, T.: Resonant nonlinear intrachannel interactions in strongly dispersion-managed transmission systems. Opt. Lett. **25**, 1750–1752 (2000)

Modeling of Networked Control Systems Based on Multiple Communication Channels with Event-Triggering

Li-Ying Zhao, Qian Li and Jun Wang

Abstract The work here examines the stability problem of a linear system and controller design of the networked control system with transmission delays. An innovative model is constructed based on the discrete event-trigger communication mechanism and multiple communication channels between sensor and controller. Lyapunov stability theory and liberty matrix method are applied to derive sufficient conditions for the exponential stability and the design controller. Two simulation examples are given to show that the proposed theorems are superior to other event-trigger methods in some published literature.

Keywords Networked control systems (NCSs) · Discrete event-triggered scheme · Multiple communication channels · Exponentially stability · Linear matrix inequality

1 Introduction

Recently, Networked Control Systems (NCS) have become popular among scholars due to a range of advantages such as easy installation and high efficiency [1, 2]. The signal transmits through a common network medium rather than point-to-point wirings. Due to the insertion of a network, induced delay can decrease the stability of the systems. In an NCS, the signals of the plant are sampled periodically and then

L.-Y. Zhao · Q. Li (✉)
School of Mathematics and Physics, University of Science and Technology
of Beijing, Beijing 100083, China
e-mail: 15110794231@163.com

L.-Y. Zhao
e-mail: liyingzhao@ustb.edu.cn

J. Wang
Institute of Automation and Electrical Engineering, Beihang University,
Beijing 100191, China
e-mail: dwill-wang@buaa.edu.cn

released to the controller through the network. If the sampling period is too small, the number of data packets will increase quickly, which can lead to an overloaded communication bandwidth. However, when there is little fluctuation of the measurement signals, fewer sampled signal are must be transmitted for control design task [3].

Therefore, the event-triggered mechanism has been shown to be more effective than the time-triggered method in terms of decreasing data transfer. As to event-triggered mechanism, the data packets are updated when required and needless signals can be avoided. Studies show that when the event-triggered condition is satisfied, the sampling is triggered in [4, 5]. The event-triggered condition is given as an inequality based on the state vector, when one side of the inequality exceeds a fixed threshold, the signals are sampled. The event-triggering mechanism has three advantages: (1) the mechanism can work to sample when the signals are necessary; (2) reduces the burden of the network bandwidth; (3) reduces the cost of the controller design in [5].

Most researchers in the NCSs field use a single communication channel for the transmission of information where time-delay and information missing are inevitable in [6, 7]. If the communication channel is improved, the conservation of the system can be reduced. The problem of stability analysis and controller designs are studied with multiple communication channels, which are used in NCS not concerned with event-trigger in [8, 9]. The work presented here focuses on designing the event-triggered controller for networked systems considering the multiple communication channels and time-delay. For convenient calculation, we have modeled the system with two channels and study its stability by a suitable Lyapunov functional and liberty matrix method. We than solve the controller matrix by LMI technique.

Notation in this paper is quite standard. '*' denotes the entry of matrices implied by symmetry.

2 Problem Description

2.1 The Plant

Consider the following continuous-time linear system

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0 \\ y(t) = Cx(t) \end{cases} \tag{1}$$

where $x(t) \in R^n$ is the system state vector, and $y(t) \in R^r$ is the measurement output vector; system matrices $A \in R^{n \times n}$, $B \in R^{n \times n}$, $C \in R^{r \times n}$ are real constant matrices; and x_0 is the initial condition.

In this paper, the purpose is to design a state feedback controller $u(t) = Kx(t)$ to make the system stability, where K is a matrix to be determined later.

2.2 Event Generator

In this section, event generator consists of a register and a comparator [3]. The register stores the last released data $x(i_k h)$ $(i_k = 1, 2...)$, while the comparator is used to check whether the next sampled signal $x((i_k + j)h)$ $(j = 1, 2...)$ satisfies the following judgement algorithm:

$$\left\|\Omega^{\frac{1}{2}}[x((i_k + j)h) - x(i_k h)]\right\|_2 \leq \sigma \left\|\Omega^{\frac{1}{2}} x((i_k + j)h)\right\|_2 \tag{2}$$

where Ω is a positive definite matrix, $\sigma > 0$, and $\sigma \in [0, 1)$.

Remark 1 Under the constraint (1), the sampled state $x((i_k + j)h)$ will not be transmitted if it is satisfying the inequality (2). Assuming that h is the sampled period, release times are $i_0 h, i_1 h...$, where $i_0 = 0$ is the initial time, and $i_k h$ is the sensor sampling instant. Supposing that the delay in the network communication is τ_k and $\tau_k \in (0, \bar{\tau})$. t_k $(k = 1, 2...)$ is the time instants at which the data packet x $(i_k h)$ arrives at the controller. It is concluded that $t_1 < t_2 < \cdots t_k...$ because packet dropouts and packet disordered do not occur in the transport process. The signal transmitted to the controller can be written as $x(t) = x(i_k h)$, $t \in [t_k, t_{k+1})$.

Let:

$$\rho_k = \min\{j | t_k + jh \geq t_{k+1}, j = 1, 2...\}. \tag{3}$$

The interval $[t_k, t_{k+1})$ can thus be written as $[t_k, t_{k+1}) = \cup_{j=1}^{\rho_k} I_j$, and $I_j = [t_k + (j-1)h, t_k + jh)$, $j = 1, 2, ..., \rho_k - 1$, $I_{\rho k} = [t_k + (\rho_k - 1)h, t_{k+1})$.

Two functions $\tau(t)$ and $e_k(t)$ in $[t_k, t_{k+1})$ can be described as

$$\tau(t) = \begin{cases} t - i_k h & t \in I_1 \\ t - i_k h - h & t \in I_2 \\ \vdots & \vdots \\ t - i_k h - (\rho_k - 1)h & t \in I_{\rho_k} \end{cases},$$

$$e_k(t) \in \begin{cases} x(i_k h) - x(i_k h) & t \in I_1 \\ x(i_k h) - x(i_k h + h) & t \in I_2 \\ \vdots & \vdots \\ x(i_k h) - x(i_k h + (\rho_k - 1)h) & t \in I_{\rho_k} \end{cases}.$$

Then $x(i_k h) = e_k(t) + x(t - \tau(t))$, $\tau_m \leq \tau_k \leq \tau(t) \leq h + \bar{\tau} = \tau_M$.

Remark 2 From the definition of $e_k(t)$ and the triggering Algorithm (2), it can be seen that, for $t \in [t_k, t_{k+1})$,

$$e_k^T(t)\Omega e_k(t) \leq \sigma x^T(t - \tau(t))\Omega x(t - \tau(t)). \tag{4}$$

2.3 Multiple Channels

Since the system transmits the useful information through multiple channels, we consider two channels to transmit the data packet and Eq. (1) can be rewritten as

$$\dot{x}(t) = Ax(t) + BKM_\sigma(x(t - \tau(t)) + e_k(t)), \tag{5}$$

where $M_\sigma = \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix}$, $\sigma_i = \begin{cases} 1 & \text{though the ith channel} \\ 0 & \text{though the other channel} \end{cases}$ $(i = 1, 2)$, $P(\sigma_1 = 1) = p$, $P(\sigma_2 = 1) = 1 - p$, $K \in R^{1 \times 2}$.

Remark 3 As shown in Fig. 1, the closed-loop NCS is composed of a plant, a sensor, an Event-trigger mechanism, a controller and an actuator. Two channels are established which change the original characteristics of single channel data transmission. A number of sampled data packets will be discarded under the discrete event-triggering mechanism, which will make the network loads reduced greatly. If we can improve the channel, a good quality of the network can be possibly ensured.

2.4 Main Lemmas

Definition 1 [10] The system (5) is said to be exponentially stable if for $\forall \varepsilon > 0$, there exist scalars $\alpha > 0$, $\beta > 0$, such that

$$E\left\{\|x(t)\|^2\right\} \le \alpha e^{\beta t} \sup_{-2\bar{\tau} \le s \le 0} E\left\{\|\phi(s)\|^2\right\}$$

where $\phi(.)$ is the initial function in Eq. (5), $x(t) = \phi(t)$, $t \in [-\bar{\tau}, 0]$.

Lemma 1 [5] *For positive matrices $R > 0, X > 0$, ρ is any chosen constant, we have*

$$-XR^{-1}X \le \rho^2 R - 2\rho X \tag{6}$$
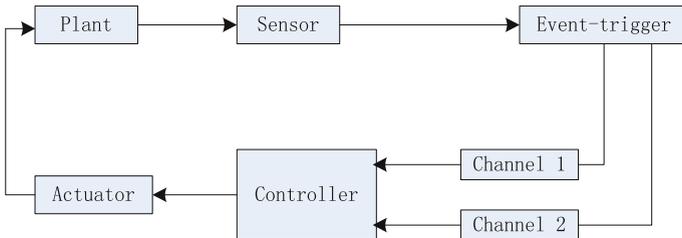
Fig. 1 The diagram for event-triggered NCS structure

Lemma 2 [11] *For any vectors* $x, y \in R^n$, *and positive definite matrix* $Q > 0$, *then the inequality holds*:

$$-2x^T y \le x^T Q x + y^T Q^{-1} y \tag{7}$$

3 Stability Analysis and Controller Design

Theorem 1 *For given* p, σ *and matrix* $K \in R^{1 \times 2}$, *the nominal system is exponentially stable if there exist positive matrices* $P_1 > 0$, $P_2 > 0$, $\Omega > 0$, $R > 0$.

$$\begin{bmatrix} W & \Phi_{12} & \Phi_{13} & \sqrt{\tau_M}N & \sqrt{\tau_M}M \\ * & -p^{-1}R^{-1} & 0 & 0 & 0 \\ * & * & -(1-p)^{-1}R^{-1} & 0 & 0 \\ * & * & * & -R & 0 \\ * & * & * & * & -R \end{bmatrix} < 0 \tag{8}$$

where $W = \Phi_{11} + \Gamma + \Gamma^T$, $\Gamma = [N \ M\text{-}N \ \text{--}M \ 0]$,

$$\Phi_{11} = \begin{bmatrix} pP_1A + (1-p)P_2A + (pP_1A + (1-p)P_2A)^T & pP_1BKE_1 + (1-p)P_2BKE_2 & 0 & pP_1BKE_1 + (1-p)P_2BKE_2 \\ * & \sigma\Omega & 0 & 0 \\ ** & -Q & 0 \\ ** & * & -\Omega \end{bmatrix},$$

$\Phi_{1i} = \begin{bmatrix} \sqrt{\tau_M}A & \sqrt{\tau_M}BKE_i & 0 & \sqrt{\tau_M}BKE_i \end{bmatrix}^T$, $i = 1, 2$, $N = \begin{bmatrix} N_1^T & N_2^T & N_3^T & N_4^T \end{bmatrix}^T$,

$M = \begin{bmatrix} M_1^T & M_2^T & M_3^T & M_4^T \end{bmatrix}^T$ $E_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ $E_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$.

Proof Consider the following functional:

$V(t) = V_1(t) + V_2(t) + V_3(t)$, $V_1(t) = x^T(t)P_\sigma x(t)$, $V_2(t) = \int_{t-\tau_M}^t x^T(s)Qx(s)ds$, $V_3(t) = \int_{t-\tau_M}^t \int_s^t \dot{x}^T(v)R\dot{x}(v)dvds$. Using free weighting matrix method [9], the following equation holds:

$$2\xi^T(t)M\left[x(t - \tau(t)) - x(t - \tau_M) - \int_{t-\tau_M}^{t-\tau(t)} \dot{x}(s)ds\right] = 0,$$

$$2\xi^T(t)N\left[x(t) - x(t - \tau(t)) - \int_{t-\tau(t)}^{t} \dot{x}(s)ds\right] = 0,$$

where N, M are appropriate dimensions matrices, and by Lemma 2,

$$- 2\xi^T(t)M \int_{t-\tau_M}^{t-\tau(t)} \dot{x}(s)ds \le (\tau_M - \tau(t))\xi^T(t)MR^{-1}M^T\xi(t) + \int_{t-\tau_M}^{t-\tau(t)} \dot{x}^T(s)R\dot{x}(s)ds,$$

$$- 2\xi^T(t)N \int_{t-\tau(t)}^{t} \dot{x}(s)ds \le \tau(t)\xi^T(t)NR^{-1}N^T\xi(t) + \int_{t-\tau(t)}^{t} \dot{x}^T(s)R\dot{x}(s)ds.$$

From Eq. (8), we can determine that $E(\dot{V}) < 0$, which means that there exists a constant $\lambda > 0$, such that

$$E(\dot{V}) \le -\lambda E\{\xi^T(s)\xi(s)\}$$

Using a method similar to that of [12], we have

$$E(x^T(t)x(t)) \le \alpha e^{-\varepsilon t} \sup_{-\bar{\tau} \le s \le 0} E\{\|\phi(s)\|^2\} \tag{9}$$

Theorem 2 *For given scalars σ, ρ, if there exist positive matrices $X > 0$, $\tilde{R} > 0$, $\Omega > 0$, N_i, M_i ($i = 1, 2\ldots 8$), $\tilde{N}_i, \tilde{M}_i$ ($i = 1, 3$) with appropriate dimension*

$$\begin{bmatrix} \Xi_{11} & \Xi_{12} & -\tilde{M}_1 + \tilde{N}_3^T & \Xi_{14} & \sqrt{\tau_M}XA^T & \sqrt{\tau_M}XA^T & \sqrt{\tau_M}\tilde{N}_1 & \sqrt{\tau_M}\tilde{M}_1 \\ * & \Xi_{22} & -(M_6 + M_7 - N_7)^T & (M_4 - N_4)^T & \sqrt{\tau_M}E_1^T K^T B^T & \sqrt{\tau_M}E_2^T K^T B^T & \sqrt{\tau_M}N_6^T & \sqrt{\tau_M}M_6^T \\ ** & -\tilde{M}_3 - \tilde{M}_3^T - \tilde{Q} & -M_8 & 0 & 0 & \sqrt{\tau_M}\tilde{N}_3 & \sqrt{\tau_M}\tilde{M}_3 \\ ** & * & -\Omega & \sqrt{\tau_M}E_1^T K^T B^T & \sqrt{\tau_M}E_2^T K^T B^T & \sqrt{\tau_M}N_8^T & \sqrt{\tau_M}M_8^T \\ ** & ** & p^{-1}(\rho^2\tilde{R} - 2\rho X) & 0 & 0 & 0 \\ ** & ** & * & (1-p)^{-1}(\rho^2\tilde{R} - 2\rho X) & 0 & 0 \\ ** & ** & ** & -\tilde{R} & 0 \\ ** & ** & ** & * & -\tilde{R} \end{bmatrix} < 0 \tag{10}$$

where

$$\Xi_{11} = (1 - p)\varepsilon AX + (1 - p)\varepsilon XA^T + pAX + pXA^T + \tilde{Q} + \tilde{N}_1 + \tilde{N}_1^T,$$
$$\Xi_{12} = (1 - p)\varepsilon BKE_2 + pBKE_1 + M_5 - N_5 + N_6,$$
$$\Xi_{14} = (1 - p)\varepsilon BKE_2 + pBKE_1 + N_8$$
$$\Xi_{22} = \sigma\Omega + M_2 - N_2 + (M_2 - N_2)^T.$$

To solve the linear matrix inequality in Eq. (10), we can get the feedback matrix K.

Proof In Eq. (8), making $P_2 = \varepsilon P_1$, defining matrix $X = P_1^{-1}$, and multiplying (8) left and right by $diag\{X, I, X, I, I, I, X, X\}$. New matrix variables can be defined as:

$\tilde{Q} = XQX, \tilde{R} = XRX, M_5 = XM_1, N_5 = XN_1, M_6 = XM_2, N_6 = XN_2,$

$M_7 = XM_3, N_7 = XN_3, M_8 = XM_4, N_8 = XN_4, \tilde{M}_i = XM_iX \ (i = 1, 3), \tilde{N}_i = XN_iX \ (i = 1, 3).$

Using Lemma 1, Eq. (10) will be obtained.

4 Two Applications to Different Dimension Systems

Example 1

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

When $\sigma = 0.3$, though Matlab program, we can get K = [−0.4696 −0.6150].

Remark 4 From Fig. 2, two curves $x_1(t)$ and $x_2(t)$ eventually can merge together, so the system will reach the stability if the controller insert to the open loop system. Besides, when $p = 0.3$, the value of $\tau_M = 6.7996$ is larger than the value of $\tau_M = 5.9976$ when $p = 0.5$.

Example 2

$$A = \begin{bmatrix} -1 & -5 & 0 \\ 0 & -0.5 & 1 \\ -3 & -1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 0.8 & 0.5 \\ 1.4 & -1.5 & 1 \\ 0.9 & 0.8 & -1.7 \end{bmatrix}.$$
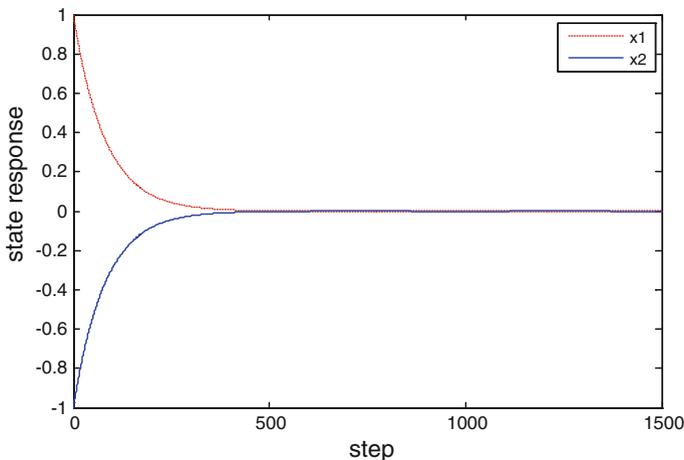
Fig. 2 State response of the system
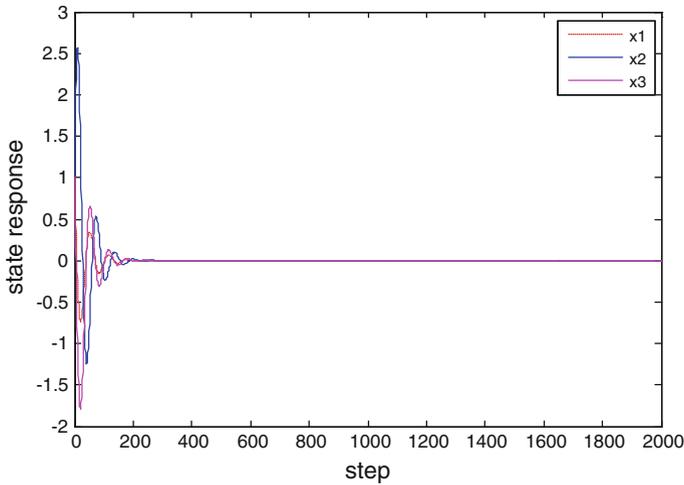
Fig. 3 State response of the system

Using Theorem 2 with $\sigma = 0.3$, $p = 0.3$, the upper bound in this system is computed as 0.4225 and corresponding controller gain and the event-trigger matrix are

$$K = \begin{bmatrix} -26.4253 & -13.0952 \end{bmatrix}, \quad \text{and}$$

$$\Omega = 1.0\text{e}^{+03} \begin{bmatrix} 1.4471 & -0.1002 & 0.0124 \\ -0.1002 & 1.0006 & -0.0011 \\ 0.0124 & -0.0011 & 0.9836 \end{bmatrix}.$$

Remark 5 From Fig. 3, three curves eventually can merge together, which demonstrates the usefulness of the controller design for NCS with multiple communication channels. Besides, theorems in this paper are also applied for higher dimension.

5   Conclusion

In this study, an event-trigger sampling model has been constructed which can determine the transfer of the sampled signals. The new communication channel constructed in this scheme can avoid both unnecessary sampling signal transmission and bandwidth limitations. Criteria for stability and control design have been obtained based on Lyapunov stability method. Finally, the simulation examples show that the theorems in the work can deal with the higher dimension control problem and make the system stability, which are effectiveness and usefulness of the proposed scheme.

References

1. Xiao, S.P., Guo, L.X., Wang, L.Y.: The stability analysis for a class of time-delay networked control systems. In: The 26th Chinese Control and Decision Conference, pp. 4977–4980. IEEE, Piscataway (2014)
2. Liu, K., Fridman, E.: Delay-dependent methods and the first delay interval. Syst. Control Lett. **64**(1), 57–63 (2014)
3. Zhang, X.M., Han, Q.L.: Event-triggered dynamic output feedback control for networked control systems. IET Control Theor. Appl. **8**(4), 226–234 (2014)
4. Kia, S.S., Cortés, J., Martínez, S.: Distributed event-triggered communication for dynamic average consensus in networked systems. Automatic **59**(3), 112–119 (2015)
5. Yue, D., Tian, E.G., Han, Q.L.: A delay system method to design of event-triggered control of networked control systems. In: 2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), vol. 413, pp. 1668–1673. IEEE press, Piscataway (2011)
6. Lunze, J., Lehmann, D.: A state-feedback approach to event-based control. Automatic **46**(1), 211–215 (2010)
7. Lemmon, M., Chantem, T., Hu, X., Zyskowski, M.: On self-triggered full-information H-infinity controllers. In: Proceedings of the 10th International Conference on Hybrid Systems: Computation and Control, vol. 4416, pp. 371–384. Springer, Berlin (2007)
8. Wang, Y.L., Yang, G.H.: Multiple communication channels-based packet dropout compensation for networked control systems. IET Control Theor. Appl. **2**(8), 717–727 (2008)
9. Liu, K., Fridman, E., Hetel, L.: Networked control systems in the presence of scheduling protocols and communication delays. Soc. Indust. Appl. Math. **53**(4), 1768–1788 (2015)
10. Yue, D., Tian, E.G., Wang, Z.D., James L.: Stabilization of systems with probabilistic interval input delays and its applications to networked control systems. IEEE Trans. Syst. **39**(4), 939–945 (2009)
11. Wang, Y., Xie, L.C., de Souza, C.E.: Robust control of a class of uncertain nonlinear systems. Syst. Control Lett. **19**(2), 139–149 (1992)
12. Mao, X., Koroleva, N., Rodkina, A.: Robust stability of uncertain stochastic differential delay equations. Syst. Control Lett. **35**(5), 325–336 (1998)

Modeling of Spatial-Temporal Associations on a Mobile Trajectory

Xiang Ren, Rong Xie, Juan Du and Xiang-Yi He

Abstract Despite the advanced state of development of mining association rules for time series data, simultaneous spatial and temporal data mining—particularly for mobile trajectories—remains a challenge, and requires huge amounts of data to be scanned during the mining process. This paper proposes a method that improves upon the incremental Godin algorithm, using a concept lattice to establish association rules based on spatial-temporal data with only a single scan of the dataset. Data was sourced from the mobile trajectories of individuals in a Tokyo subway station, then cleaned up, normalized, and compressed into one large locomotion dataset. Analysis of this data was executed using the newly proposed Indexed Lattice method, with results compared to those achieved by the traditional Godin and Apriori algorithms. It was shown that the Indexed Lattice method performed more efficiently and with greater stability, especially when processing data possessing more attributes.

Keywords Spatial-temporal data · Association rules mining · Concept lattice · Hasse diagram · Indexed Lattice method

X. Ren · R. Xie (✉) · J. Du · X.-Y. He
International School of Software, Wuhan University, Wuhan
Hubei, China
e-mail: xierong@whu.edu.cn

X. Ren
e-mail: xiangren2000@gmail.com

J. Du
e-mail: dj_nancy@163.com

X.-Y. He
e-mail: happyriver.love@163.com

1   Introduction

The recent, rapid proliferation of intelligent devices with built-in position tracking has created abundant information on the movements of people, their vehicles, and their smart electronics. Analyzing the spatial-temporal data generated can elucidate a variety of previously unknown behavioral patterns exhibited by users of sensor-equipped devices. Although locomotion data has become readily available in the digital era, evaluating the simultaneous spatial and temporal components of this data continues to present a challenge [1, 2]. In response, data miners have turned to association rules based on market basket transaction data to generate spatial-temporal association patterns [3]. The Apriori algorithm has been influential in its application toward itemsets mining and association rule learning from transactional databases [4]. It uses k-1 frequent itemsets to search for k frequent itemsets based on pre-defined minimum support. The frequent itemsets determined can then be utilized to establish association rules with pre-defined minimum confidence. However, most algorithms of this kind must scan datasets multiple times, which is both tedious and expensive due to the quantity of data involved.

Concept lattices have recently risen to prominence as a more elegant and efficient solution. A concept lattice is a conceptual knowledge representation model first proposed by Rudolf Wille. A formal concept is defined to be a pair consisting of a set of objects (the "extent") and a set of attributes (the "intent") [5], and concept lattice analysis involves extracting relationships and hierarchies from the common attributes shared by objects. Formal concept lattice analysis theory expresses both the intent and extent of the concept using formal mathematical language. It is capable of forming spatial associations without a heavy dependence on numerical data, and does not assume statistical independence between neighbors—an unrealistic expectation when associations between neighboring regions exist. Lattices can be represented by Hasse diagrams—in which each node represents a concept—comprised of two parts: intension and extension. Links between nodes indicate sub-concept and super-concept relationships, and each dataset is represented by one concept lattice, which covers all instances in the dataset. Hasse diagrams present a convenient means of visualizing the spatial associations generated.

The Godin algorithm is a classical method of producing lattices and Hasse diagrams incrementally [6]. Using concept lattices, this study aims to improve the incremental Godin algorithm for spatial-temporal data mining using the movement of individuals in a Tokyo subway station as a data source. The investigators propose combining formal concept lattices with traditional data mining spatial association rules, and visualizing the mining results using Hasse diagrams. Our intent is to generate features from the considerable quantity of mobile trajectory data, and certify all association rules in a short time relative to Godin and Apriori algorithms.

In summary, our goals in this study are:

1. Generate mobile trajectories from a large dataset of locomotion—cleaned up, normalized, and compressed for further processing.

2. Propose an algorithm for spatial-temporal association rule mining based on concept lattices. Improve the indexing portion of the Godin algorithm, reducing storage requirements and improving retrieval efficiency.
3. Create a framework that uses spatial-temporal locomotion data mining and concept lattices to reveal association rules.

2 Related Works

Data miners have worked to combat the challenges analyzing positional information since the mining association rules problem presented in 1993 by Agrawal et al. [7]. In reality, when tackled using traditional methods, these problems always consume extraordinary time and computational resources because of the need to scan relevant datasets multiple times. Thus, a major component of association rules innovation has focused on reducing time and storage needs. Traditional mining methods operate in two phases:

1. Mining frequent item sets.
2. Mining association rules from these itemsets.

Vo and Le [8] discovered that these phases could be expedited by first building a lattice of frequent itemsets, and then mining association rules from the lattice—with greater time savings as more itemsets become involved. Mishra and Khare [9] used a standard Apriori algorithm to mine frequent itemsets from fingerprint data, and then classified the fingerprints by gender. Guo and Cui [10] designed a model to detect environment constraint factors affecting vehicle tracking, in order to generate spatial-temporal traffic information. Wen et al. [11] proposed an effective project-based algorithm, an OSAF-tree (one ScAn full tree) for moving sequential pattern mining, which obtains the maximum number of sequential patterns in a single scan of available datasets. Candia et al. [12] used time and space information recorded in mobile electronics to analyze human activities.

3 Spatial-Temporal Association Model based on Concept Lattice

3.1 Concept Lattice

A concept lattice is a conceptual knowledge representation model first proposed by Rudolf Wille in 1984—building on applied lattice and order theory developed by Garrett Birkhoff, among others, in the 1930s—concept lattices is a means of analyzing formal concepts from a collection of objects and their properties. Kuznetsov and Obiedkov [13] compared incremental and batch algorithms for generating concept lattices, and analyzed their merits and faults. Levy and Baklouti, Krajca

et al. developed a distributed version of the algorithms employed for building concept lattices that proved faster and more effective than traditional methods [14–16]. Table 1 is an example of a dataset in which a, b, c, d and e each represent one of five different attributes, and there are 9 instances of these attributes. The Hasse diagram generated from this dataset is shown in Fig. 1.

Definition 1 The formal context is a triple $k = (O, D, R)$, where O is a set of objects, D is a set of attributes, and R is the relationship between O and D. i.e. $R \subseteq O \times D$. R can be defined as two projections $f(x)$ and $g(x)$:

$$\forall m \in O : f(m) = \{n \in D | \forall m \in O, mRn\} \tag{1}$$

$$\forall n \in D : g(n) = \{m \in O | \forall n \in D, mRn\} \tag{2}$$

Table 1 An example dataset where rows represent instances and columns represent attributes

Instance	Attributes				
	a	b	c	d	e
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0
6	0	1	1	0	0
7	1	0	1	0	0
8	1	1	0	0	1
9	1	1	1	0	0

"1" denotes an instance with the attribute; "0" denotes an instance without the attribute
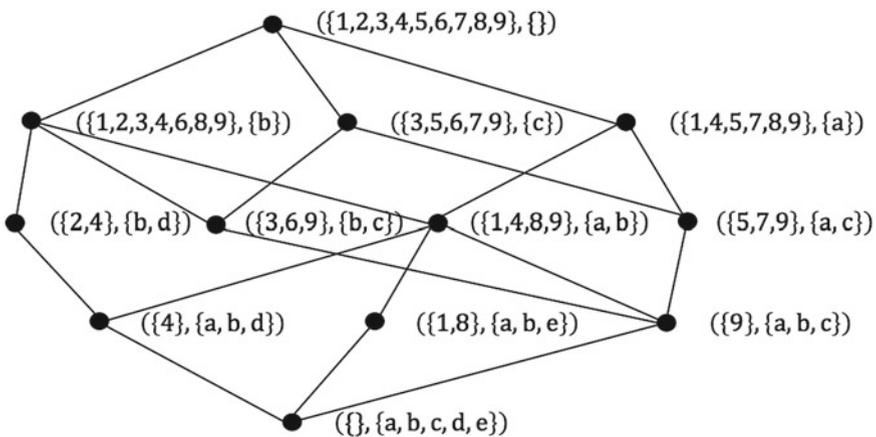
Fig. 1 Hasse diagram generated from dataset in Table 1

A couple $C_1(O_1, D_1)$ derived from the formal context is called a concept, where $O_1 \in O, D_1 \in D, D_1' = g(D_1) = O_1$. O_1 is the extension, and D_1 is the intension of the concept C_1.

Definition 2 Given two concepts $C_1(O_1, D_1)$ and $C_2(O_2, D_2)$, if $O_1 \subseteq O_2$ and $D_2 \subseteq D_1$ (denoted as $C_1 \leq C_2$), and there is not other concept C_3 in the lattice such that $C_1 \leq C_3 \leq C_2$, then C_2 is the parent of C_1 (immediate super concept), and C_1 is the child of C_2 (immediate sub concept).

3.2 Spatial-Temporal Association Rules Mining

The primary function of association rules mining is discovery of common patterns within a large dataset [3]. Spatial-temporal association rules indicate certain relationships among a set of positions or locations as they change over time. If A and B are two itemsets, a spatial-temporal association rule between them can be represented as $A \Rightarrow B$. Support and confidence are two important measures of the availability and certainty of rules—support indicates the probability of a pattern's occurrence in a set of objects, and confidence indicates the probability of B whenever A occurs.

1. $Support(A \Rightarrow B) = P(A \cup B)$
2. $Confident(A \Rightarrow B) = P(B|A)$

Definition 3 In a generated Hasse diagram, if there exists a couple (C_1, C_2), where $C_1 = (g(A), f(g(A)))$, $C_2 = (g(A \cup B), f(g(A \cup B)))$, then

1. $support(A \Rightarrow B) = |Extension(C_2)|/|O|$
2. $confidence(A \Rightarrow B) = |Extension(C_2)|/|Extension(C_1)|$

where $|Extension(C_2)|(or|Extension(C_1)|)$ is the extent cardinal of concept $C_2(or\ C_1)$, and $|O|$ is the number of objects in the spatial-temporal dataset. Given user defined minimum support thresholds and confidence thresholds, association rules satisfying both are called strong association rules.

3.3 Building Concept Lattice

Concept lattices are a classical framework for discovering space-associated information, which can not only improve the accuracy and efficiency of mining algorithms, but more recently can also detect all association rules [17]. Thus far, two types of algorithms have been used to construct concept lattices: incremental algorithms and batch algorithms. For this study we implemented an incremental algorithm that is an improvement upon the traditional Godin algorithm [6]. In the Godin algorithm, objects are inserted from top to bottom. When a new concept is inserted, it must be compared to every node in the existing lattice, and introduced according to the relationships established.

```
Input: Lattice L, new object x*, and its intent f{x*}.
Output: updated L*.
BEGIN
  IF inf(L) = (∅,∅) THEN Update inf(L) with ({x*},{f{x*}});
  IF f{x*} ⊄ Intent(inf(L)) THEN
    Add node H: (∅,Intent(inf(L)∪f{x*})
    inf(L) := H
    Add link from original inf(L) to H
  ENDIF
  B[i] := {C : ||Intent(C)|| = i};
  // Group all nodes C with the same cardinals in L
  Initialize index datasets B'[i] := ∅;
  FOR i := 0 TO maximum cardinality DO
    FOR each pair C in B[i]
      IF Intent(C) ⊆ f{x*} THEN
        // C is a updating concept
        Extent(C) := Extent(C)∪{x*};
        Add C to B'[||C||];
      ELSE
        Int := Intent(C)∩f{x*}
        IF ∄C₁ ∈ B'[||Int||], where Intent(C₁) = Int
        THEN
          Add new node Cₙ := (Extent(C)∪{x*},Int);
          // C is a generated sub concept
          Add Cₙ to B'[||Int||]
          Add link from Cₙ to C
          Update links
    END FOR
  END FOR
```

4 Improved Godin Algorithm

While the aforementioned traditional Godin algorithm inserts new concepts from top to bottom, making comparisons to every other concept in the lattice, this study introduces a new approach where indexing is used to improve lattice-building efficiency.

4.1 Concept Lattice based on Indexing

In our proposed method, we construct an index tree where each attribute is treated as a character, sorted by the order in which it appears in the alphabet. Consider an attribute set $D = \{d_1, d_2, d_3, \ldots, d_k\}$ mapping to the alphabet $a_1, a_2, a_3, \ldots, a_k$, where $a_1 < a_2 < a_3, \ldots, < a_k$. If we represent attributes with lines, we can use an index tree to re-organize concept nodes, as shown in Fig. 2. From Fig. 2 we can see that $T1$ is the
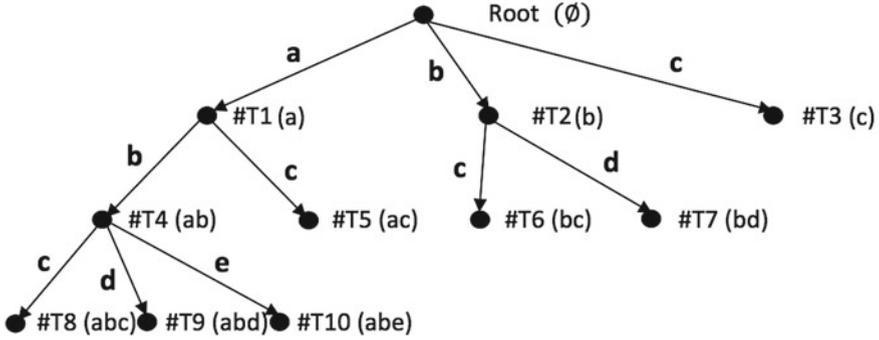
Fig. 2 Index tree for example concept lattice

parent node of $T5$, denoted by $T5 = T1.Children[c]$. Given that each concept is mapped to a node, we define a mapping function λ: for root node, $\lambda(root) = null$; if $T5 = T1.Children[c]$, then $\lambda(T5) = \lambda(T1) + c$. Notably, if there is a concept C, where $\gamma(Intent(C)) = \lambda(T)$, then T is a valid tree node. In order to construct an index tree, we first initiate the tree as a root, which when mapped to a concept has a content of null. We then individually insert each concept into this index tree according to the following function. The purpose of this function is to find D, the attribute to be inserted in index tree. If the tree node does not exist or is invalid, then this function builds a corresponding tree node and returns that tree node.

```
// where T.root is the root node of the index tree, D
is an attribute set.
BEGIN
   IndexNode = T.root;
   FOR each attribute d ∈ D in ascending order DO
      IF (IndexNode.children[d]! = null) THEN
         IndexNode = IndexNode.children[d];
      ELSE
         IndexNew.Lattice = null;
         IndexNode.children[d] = IndexNew;
         IndexNode = IndexNode.children[d];
      END IF
   END FOR
   Return IndexNode
END
```

4.2 Indexed Lattice

Using the Search and Insert functions, we can construct a map from concept lattice to index tree—when searching for a concept node, it is much faster to search in the tree. By combining the concept lattice with the index tree, we can significantly

reduce time and storage needs. Each *TreeNode*, which is a combination of a concept node and an index tree node, has five attributes: parents, children, tree children, intent and extent. In this context, parents and children are parent and child nodes in the concept lattice, while tree children are child nodes in the index tree. Our improved algorithm is as follows:

```
Input: Lattice L, new object x*, and its intent f{x*}.
Output: updated L*.
BEGIN
   IF  inf(L) = (∅,∅) THEN
   Update  inf(L) with ({x*},{f{x*}});
   IF  f{x*} ⊄ Intent(inf(L)) THEN
      Add node H: (∅,Intent(inf(L)∪f{x*})
      inf(L) := H
      Add link from original inf(L) to H
   END IF

   // Depth-First Traversal
   //Establish stack container and initialize it
   Stack:=null;
   Stack.add(inf(L));
   Stack.add(TreeRoot);
   WHILE (!stack.is empty())
      indexTreeNode:=stack.pop
      //push child nodes of indexTreeNode into stack,
      and initialize them
      FOR each child node d in index tree
         Stack.add(d)
      END FOR
      IF  indexTreeNode.intent.size == 0 &&
         indexTreeNode.! = TreeRoot
      CONTINUE
      END IF
      IF  Intent(indexTreeNode) ⊆ f{x*}
         // indexTreeNode is a updating concept
         Extent(indexTreeNode) := Extent(indexTreeNode)∪{x*}
         Add indexTreeNode to B'
      ELSE
         int := Intent(indexTreeNode)∩f{x*}
         AddNode:=SearchOrInsert(T.root,int)
         IF  Extent(AddNode) == 0 && int.! = 0 THEN
            // AddNode is a generated sub concept
            Extent(AddNode) := Extent(indexTreeNode)∪{x*}
            Add AddNode to B'
            Add link from indexTreeNode to AddNode
            update links
         END IF
      END IF
   END WHILE
```

With this algorithm, we construct the index tree and concept lattice at the same time. Whenever a new instance is inserted, using the search or insert functions, we can add or validate this node in the index tree. We can then use the index tree to locate nodes in the concept lattice that intersect with the new instance and update the lattice accordingly.

5 Experiment and Analysis

5.1 Data Description

The movements of individuals within a Tokyo subway station were recorded. Each person, passenger or staff, was scanned several times while inside the station, with each scan stored in frames. Each frame recorded the time and location of the person.

5.2 Experiment Design

Time and location features were extracted from raw data such that each instance had the same dimensions in the data cleaning step. As position sensors treated the individuals to which they were attached as origin points, coordinates (x, y) were normalized according to the following method:

$$x' = (x - x_{min})/(x_{max} - x_{min}) \tag{3}$$

$$y' = (y - y_{min})/(y_{max} - y_{min}) \tag{4}$$

where x_{max} and x_{min} are the maximum and minimum coordinates in the x axis, respectively; while y_{max} and y_{min} are the maximum and minimum coordinates in the y axis, respectively. After normalization all data sets were projected onto a map with the coordinates from (0,0) to $(x_{max} - x_{min}, y_{max} - y_{min})$.

To reduce complexity, normalized data was discretized in the compression step. To do this, the normalized map was divided into $n \times n$ grids. Time series data located in a particular grid area was treated as if it was located at the center point of that grid. To establish the value of parameter n, 500 instances were randomly selected, from which compression results showed that an optimal value for n was 250. We then applied the Godin and Apriori algorithms and compared the results to those achieved by the Indexed Lattice method. This entire process is represented in Fig. 3.
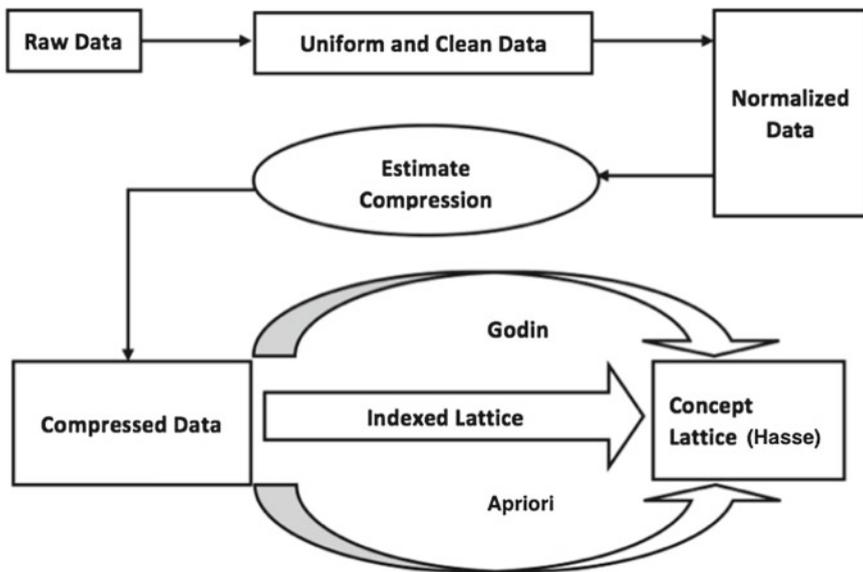
Fig. 3 Experiment process flow

5.3   Results

As the Hasse diagrams produced by all algorithms are identical, we can only compare the time required to execute the Godin, Apriori and Indexed Lattice algorithms, shown in Fig. 4. We can see that when support = 0.01, the Apriori algorithm generates frequent itemsets most quickly. Additionally, when the number of points exceeds 600 the Apriori algorithm's time consumption decreases
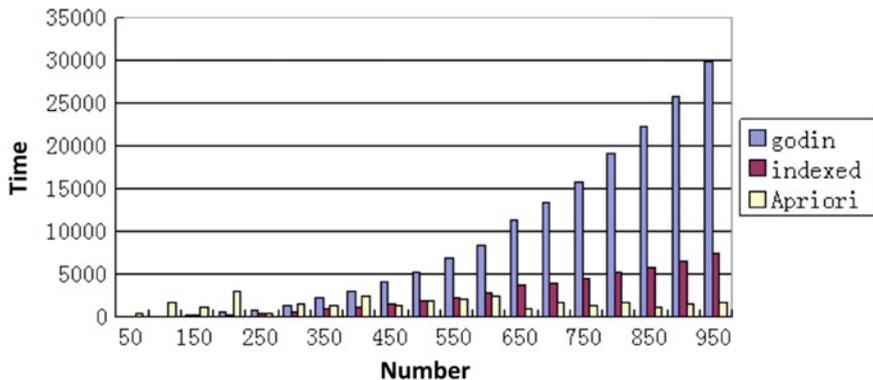
Fig. 4 Time required for Godin, Apriori and indexed lattice algorithms to generate frequent item sets

dramatically due to excessive pruning, likely resulting in information loss. Meanwhile at support values greater than 0.01, the time used by the Apriori algorithm increases significantly, even beyond that used by the Godin algorithm.

The new proposed Indexed Lattice algorithm requires considerably less time than the incremental Godin algorithm. Additionally, the Indexed Lattice algorithm should prove even more efficient when a greater number of attributes are present.

6 Conclusion

Though spatial-temporal data mining has recently pervaded many fields, this study represents the first attempt to incorporate the use of concept lattices. This novel Indexed Lattice approach, when used to analyze data collected from individuals moving within a Tokyo subway station, proved more efficient and stable than traditional incremental and Apriori algorithms. Even greater benefits are anticipated when handling data with more attributes.

References

1. Shrivastava, J., Shrivastava, N.: A review of data reduction/extraction in data mining from the large set of database. Int. J. Electr. Electron. Comput. Eng. **3**(2), 149 (2014)
2. Hira, S., Deshpande, P.S.: Data analysis using multidimensional modeling, statistical analysis and data mining on agriculture parameters. Procedia Comput. Sci. **54**, 431–439 (2015)
3. Han, J., Kamber, M., Pei, J.: Data mining: concepts and techniques: concepts and techniques. Elsevier, Waltham (2011)
4. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference Very Large Data Bases (VLDB), vol. 1215, pp. 487–499. Morgan Kaufmann Publ Inc., San Mateo (1994)
5. Wille, R.: Concept lattices and conceptual knowledge systems. Comput. Math Appl. **23**(6), 493–515 (1992)
6. Godin, R., Missaoui, R., Alaoui, H.: Incremental concept formation algorithms based on Galois (concept) lattices. Comput. Intell. **11**(2), 246–267 (1995)
7. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. ACM SIGMOD Rec. **22**(2), 207–216 (1993)
8. Vo, B., Le, B.: Mining traditional association rules using frequent itemsets lattice. In:2009 International Conference on Computers and Industrial Engineering (CIE 2009), pp. 1401–1406. IEEE, Piscataway (2009)
9. Mishra, A., Khare, N.: A review on gender classification using association rule mining and classification based on fingerprints. In: 2015 Fifth International Conference on Communication Systems and Network Technologies (CSNT), pp. 930–934. IEEE, Piscataway (2015)
10. Guo, D.H., Cui, W.H.: Vehicle trajectory data mining based on real-time traffic information extraction. J. Wuhan Univ. Technol. (Trans. Sci. Eng.) (2010)
11. Wen, N.X., Qing, Y.D., Wei, T.S., Jiao, W.T.: OSAF-tree—an interactive and incremental algorithm for moving sequential pattern mining. J. Comput. Res. Develop. **10**, 022 (2004)

12. Candia, J., Gonzalez, M.C., Wang, P., Schoenharl, T., Madey, G., Barabsi, A.L.: Uncovering individual and collective human dynamics from mobile phone records. J. Phys. A: Math. Theor. **41**(22), 224015 (2008)
13. Kuznetsov, S.O., Obiedkov, S.A.: Comparing performance of algorithms for generating concept lattices. J. Exp. Theor. Artif. Intell. **14**(2–3), 189–216 (2002)
14. Levy, G., Baklouti, F.: A distributed version of the ganter algorithm for general galois lattices. Concept Lattices Appl. (CLA) **2005**, 207–221 (2005)
15. Krajca, P., Outrata, J., Vychodil, V.: Parallel recursive algorithm for FCA. Concept Lattices Appl. (CLA) **2008**, 71–82 (2008)
16. Krajca, P., Vychodil, V.: Distributed algorithm for computing formal concepts using map-reduce framework. Advances in Intelligent Data Analysis VIII, pp. 333–344. Springer, Heidelberg (2009)
17. Niu, J., Zhang, Y., Feng, W., Ren, L.: Spatial association rules mining for land use based on fuzzy concept lattice. In: 2011 19th International Conference on Geoinformatics, pp. 1–6. IEEE, Piscataway (2011)

Moving Vehicle Detection in Dynamic Traffic Contexts

Chun-Ling Tu and Sheng-Zhi Du

Abstract We address the problem of moving vehicle detection for dynamic traffic environment with moving cameras. The proposed system obtains visual data from cameras mounted on a moving host vehicle. The optical flow method is employed to estimate the movement of the camera and the relative movement of the background. Moving vehicles are detected by difference of velocity to the background movement model.

1   Introduction

Different methods to detect vehicles in the traffic context were proposed with various sensing techniques applied, such as radar sensors [1], lidar sensors [2, 3], and remote sensing [4, 5].

Among the vehicle detection approaches based on remote sensing devices, the computer vision techniques play very important roles. Because images and videos contain more useful information and wider view than radar data, relevant research obtains more and more attention recently. Based on these research, traffic situation of the vehicle can be estimated for different purposes, such as vehicle collision avoidance system [6], vehicle detection and tracking [7], etc.

Although many vehicle detection methods were reported, it is still an open area for effective vehicle detection in real world situation. In this paper, a vehicle

C.-L. Tu (✉)
Department of Computer Systems Engineering,
Tshwane University of Technology, Pretoria 0001, South Africa
e-mail: tclchunling@gmail.com

S.-Z. Du
Department of Mechanical Engineering, Mechatronics and Industrial Design,
Tshwane University of Technology, Pretoria 0001, South Africa
e-mail: dushengzhi@gmail.com

detection technique based on optical flow is proposed. The model of background movement is proposed and used to detect moving vehicles.

2   Related Work

Over past decades, with the fast development of computer technology then followed by the decreasing of electronic component price, vehicle detection based on computer vision techniques has got a considerable growing over the automatic driving, traffic monitoring and driving assistance, etc. In these applications, the moving vehicle detection is the core task, which mostly determines the total system performances, such as accuracy and reliability.

2.1   Moving Vehicle Detection

The computer vision based vehicle detection systems usually have one or more cameras. Stationary cameras are commonly used to detect moving vehicles on entrance, free way, and public places. In these cases, the cameras are mounted on fixed position, such as poles and the roof of a building, to monitor the road lanes, road crossing and parking. Because of the fact that cameras are stationary, the background (buildings and lanes etc.) is not moving which makes it easier to detect the moving vehicles. Various algorithms were developed for stationary cameras?

Another type of moving vehicle detection systems uses the information obtained from moving cameras. The cameras are usually mounted to a host vehicle and capture the video from the front or back or both scenes. Contrast to the stationary cameras, more problems are concerned, such as disturbances from the vibration of camera, complex dynamical background, host vehicle movement and so on. Therefore, the background and disturbing objects removal becomes the critical problem that seriously affects the system reliability and accuracy.

2.2   Optical Flow

Optical flow concerns the relative motion in consecutive image frames. A basic assumption is that the brightness of a pixel from different image frames is constant, i.e.

$$I(x+u, y+v, t+1) - I(x, y, t) = 0 \qquad (1)$$

where $I(x, y, t)$ is the brightness of a pixel (x, y) at time t, $I(x+u, y+v, t+1)$ is the brightness at time $t+1$, u and v are the displacements of the pixel from time t to

$t + 1$. Equation (1) expresses the intensity constancy, which is then expanded to the Taylor series [8].

$$\frac{\Delta I}{\Delta x}u + \frac{\Delta I}{\Delta y}v + \frac{\Delta I}{\Delta t} = 0. \tag{2}$$

With additional constraints [9], the solution of Eq. (2) estimates the optical flow.

3   The Proposed Method

In this research, the case of moving camera is considered where most pixels have considerable velocities which makes the stationary camera methods infeasible. The motivation of the proposed method is that the video obtained by a moving camera contains the relative motion between the background and the camera. By modelling the motion of background, it is possible to detect the foreground objects with different motion model.

3.1   Imaging System Settings

Physically, the relative velocity indicates the movement of an object referencing to others. In the real world traffic situation, all moving vehicles have relative velocity referencing to the "background" (like stationary vehicles, lanes, trees, and buildings, and so on). The objects that can be considered as "background" have very small or almost no relative velocity to each other. However, from the perspective projection which are the base nature of our vision systems (eyes, and cameras), these background objects locating on different positions and distances to the camera get different velocities in the video.

Another interesting factor deserves being noted when the focal position of a camera is setup to infinity. When the camera is moving approaching or departing the focal position, all background objects (stationary referencing to the earth) move departing or approaching the focal position in the video.

Based on these two factors, a camera with the focal position (FoP) at infinity is mounted on a vehicle and is rearward oriented. When the host vehicle moves, the camera moves on the axial direction of the view field, therefore, the background objects moving approaching to or departing from the focal position.

3.2 Background Movement Model

The ideal view field of the camera is depicted in Fig. 1a, b when the camera moves approaching to and departing from the FoP respectively. The pixels closer to the FoP move slower than the ones closer to the edge of the view field. Of course, these differences happen under the ideal situation, i.e. all background objects are far enough to the camera.

In fact, some background objects in real world traffic scenes such as the trees and buildings, are very close to the camera. So they move quite fast in the video. Fortunately the velocities of these objects keep orienting to the FoP when the camera moves on the axial direction of the view field.

As the FoP of cameras are calibrated in the center of the view field in manu-factory, the FoP is reliable for a full scene image. Theoretically, a pixel (x, y) of the background objects satisfies the following velocity constrain.

$$\frac{v_x}{v_y} = \frac{x - x_{FoP}}{y - y_{FoP}},$$ (3)

where (x_{FoP}, y_{FoP}) is the coordinates of the FoP.

Equation (3) can be used to validate whether a pixel belongs to the background or not. In the velocity field estimated by optical flow method [8], pixels consisting an object might have different velocity values and directions, therefore, it is risky to determine the object motion by a single pixel. In this research, the sum velocity of the pixels of an object is considered.
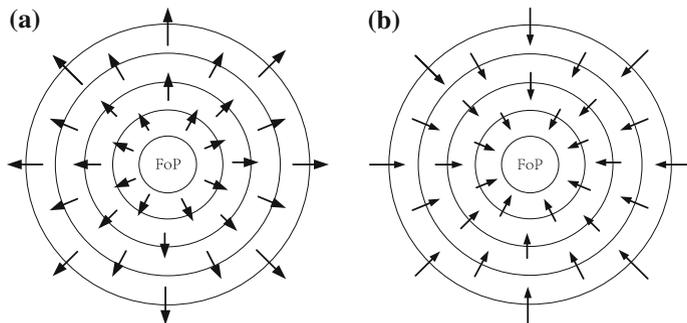
Fig. 1 The velocity field for camera axial motion. **a** Approaching. **b** Departing

4 Experiments and Results

Because of the random errors in single pixel velocity estimation, in this research, the image was split into small cells. To decrease the random errors, the average velocity of the pixels in a cell is used to present the cell motion. The average velocity indicates the movement and orientation of the group of pixels. For estimating the background movement, a voting scheme is proposed. Each cell votes to the cells lying on the direction of its velocity. After the voting process, the cell getting maximum votes is considered as the FoP, which is used to detect the camera movement and then moving vehicles.

4.1 Detecting the FoP in the View Field of a Moving Camera

As discussed in Sect. 3.2, in theory the FoP locates on the center of the images. However, because of the vibration of the host vehicle, the camera slight swings up and down. This vibration results in the FoP movement. So in a given full scene image, the FoP needs to be detected before the model in Eq. (3) is applied to detect moving vehicles. Figure 2a depicts the detected FoP. To reduce the effect of vehicle vibration, average smoothing operation is applied to the original optical flow. From Fig. 2b one finds the FoP of smoothed optical flow is much more stable, which can be used to detect moving vehicles.
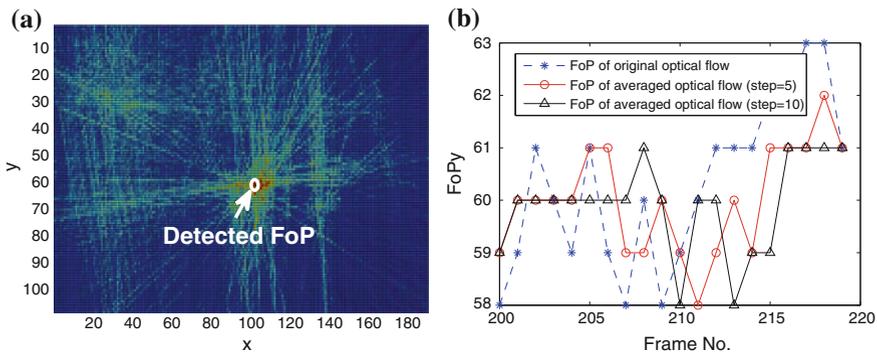
Fig. 2 FoP detection for moving camera. **a** FoP detected. **b** The change of FoP for camera vibration

4.2   Moving Vehicle Detection

Based on the FoP detected by the proposed method, the estimated object velocity obtained from optical flow is compared with the proposed background model in Eq. (3). All objects having velocity approaching to or departing from the FoP are considered as the background, but the ones having considerable velocity value and not oriented to the FoP are considered as moving objects. Then the size and position are further considered to determine whether the moving objects are vehicles or not.

The detecting result is shown in Fig. 3, where Fig. 3a is obtained by the proposed method and Fig. 3b is the traditional optical flow. In Fig. 3a the green arrows indicate the moving orientation of background objects and the blue arrows are the moving vehicles that are not oriented to the FoP. However, due to the fact that the traditional optical flow methods just consider the velocity magnitude, everything having considerable velocity are considered as moving objects. As shown in Fig. 3b, the false positive rate of the traditional method is high, where the poles, lights, trees and buildings are all considered as moving objects. That is because these objects are actually moving relative to the camera, but are stationary to the background. The proposed method effectively reduces the false positive rate by merging these motion to the background model.

To be noted that optical flow has flaws. The insensitivity to dark objects results in the missing of a dark vehicle in Fig. 3 (marked by white box), which is not successfully detected due to the low velocity value obtained from optical flow. The computation cost is another concerns. The proposed method inherits these flaws.
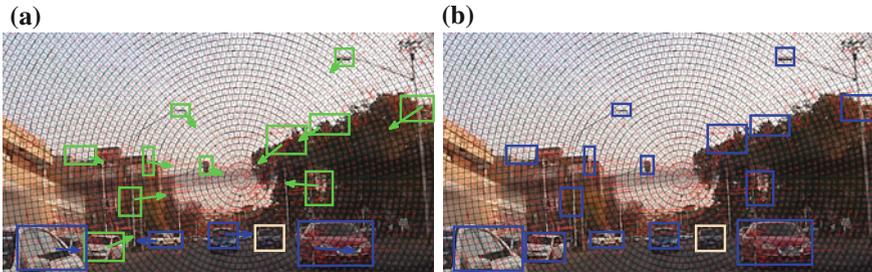
Fig. 3   Moving vehicle detection results. **a** The proposed method. **b** Traditional optical flow

5 Conclusion

In this paper, we proposed a moving vehicle detection system based on optical flow technique. The ideal model of background motion was proposed, and then validated in real world traffic scenes. Static vehicles are excluded since their consistent velocity with the background model. The proposed method can be used for traffic situation awareness, driving alarm, and anti-tracking surveillance systems, etc. The flaws of the proposed method inspire future work.

References

1. Fang, J., Meng, H., Zhang, H., Wang, X.: A low-cost vehicle detection and classification system based on unmodulated continuous-wave radar. In: IEEE Intelligent Transportation Systems Conference (ITSC 2007), pp. 715–720. IEEE, Piscataway (2007)
2. Mendes, A., Bento, L.C., Nunes, U.: Multi-target detection and tracking with laser range finder. In: IEEE Intelligent Vehicles Symposium (IV 2004), pp. 796–801. IEEE, Parma (2004)
3. Garcia, F., Cerri, P.: Vehicle Detection Based on Laser Radar. Springer, New York (2009)
4. Momin, B.F., Mujawar, T.M.: Vehicle detection and attribute based search of vehicles in video surveillance system. Circuit, Power and Computing Technologies (ICCPCT), pp. 1–4. Piscataway, IEEE (2015)
5. Satzoda, R.K., Ohn-Bar, E., Lee, J., Song, H., Trivedi, M.M.: On-road vehicle detection with monocular camera for embedded realization: robust algorithms and evaluations. In: SoC Design Conference (ISOCC), pp. 150–151. IEEE, Piscataway (2014)
6. Mukhtar, A., Xia, L., Tang, T.: Vehicle detection techniques for collision avoidance systems: a review. IEEE Intell. Transp. Syst. Mag. **16**(5), 2318–2338 (2015)
7. Satzoda, R.K., Trivedi, M.M.: Efficient lane and vehicle detection with integrated synergies (ELVIS). In: Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 708–713. IEEE, Piscataway (2014)
8. Horn, B.K., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**, 185–203 (1981)
9. Nagel, H.H.: On the estimation of optical flow relations between different approaches and some new results. Artif. Intell. **33**, 299–324 (1987)

Natural Image-Orientated Hybrid Filter Using Pulse Coupled Neural Network

Yun-Dong Li and Jia-Hao Pan

Abstract Noise introduced during capture and transmission is inevitable for natural images generated by Complementary Metal-Oxide-Semiconductor (CMOS) sensors, including quantization error during digitalization, transmission disturbance and other sources of noise. To process natural images from a CMOS sensor, a hybrid filter combining Pulse Coupled Neural Network (PCNN), median filter and Wiener filter is proposed in this paper. First, salt-and-pepper noise is located via PCNN, and processed by a median filter. Then, Gaussian noise is removed by a self-adaptive Wiener filter. Simulation results indicated that compared to other methods (hybrid filter containing median and Wiener filter, hybrid filter containing median and wavelet filter), the hybrid filter with PCNN demonstrates better performance in the preservation of image detail and edge in the premise of similar Signal-Noise Ratios (SNRs).

Keywords PCNN · Median filter · Wiener filter · Image denoising

1 Introduction

Image denoising is a persistent focus in the field of digital image processing, since the sources of noise in natural images are complex and a single filter is not able to remove all noise. The common noise in natural images can be classified as additive noise and multiplicative noise. Traditional median filters, mean filters, and Gaussian filters can improve or remove noise derived from certain sources, but often blur the image during denoising. The object of natural image denoising is to remove noise and preserve image details as much as possible.

Y.-D. Li (✉) · J.-H. Pan
North China University of Technology, Beijing, China
e-mail: liyundong@ncut.edu.cn

J.-H. Pan
e-mail: panjiahao@126.com

Currently, the most popular denoising tool is the use of mean filters and median filters [1]. The self-adaptive centre weighted median filter proposed by Jin et al. [2] can smooth self-adaptively per noise scale. Gao and Gao [3] used PCNN to identify a salt-and-pepper point, then applied a median filter accordingly to the impulsive positions. Zou [4] proposed an improved PCNN method to remove the salt-and-pepper point, which harms the image detail to some degree. Cuo and Wang [5] removed noise by using PCNN and the wavelet transform method to maintain the image details; however, this method is difficult to implement on Field Programmable Gate Array (FPGA). The above mentioned methods all have certain advantages, but not for natural images. Natural images usually contain both Gaussian and salt-and-pepper noises, thus making it difficult to achieve satisfactory results with a single filter. In order to remove noise and retain image details simultaneously, hybrid filters are proposed.

This paper proposes a novel hybrid median and Wiener filter with PCNN for natural image denoising implemented on FPGA. The denoising results verify the improved denoising performance of the proposed method compared to other methods including a median and Wiener filter, and a wavelet transform filter and median filter.

2 Simplified PCNN Model

Noises are inevitable during image capture and transmission, which normally contain both salt-and-pepper noise and Gaussian noise. This paper introduces PCNN to locate salt-and-pepper points according to the firing patterns of different neurons. Then, a median filter is applied to the determined salt-and-pepper points only, which will maintain image details and edges as much as possible [6–14].

A PCNN model is a back-feeding network constructed by a number of connected neurons, in which each neuron contains three parts: input portion, linking portion and the pulse generator portion, as shown in Fig. 1.
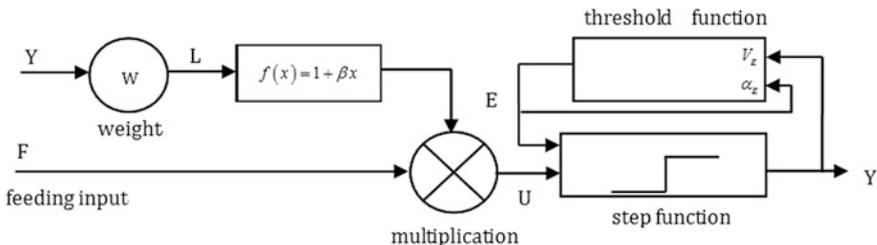
Fig. 1 Simplified PCNN neuron

Iterations of neurons are computed according to from the following Eqs. (1)–(5):

$$F_{ij}[n] = I_{ij} \tag{1}$$

$$L_{ij}[n] = \sum_k w_{ijk} Y[n-1] \tag{2}$$

$$U_{ij}[n] = F_{ij}[n]\left(1 + \beta L_{ij}[n]\right) \tag{3}$$

$$Y_{ij}[n] = \begin{cases} 1 & U_{ij}[n] \geq E[n-1] \\ 0 & U_{ij}[n] < E[n-1] \end{cases} \tag{4}$$

$$E_{ij}[n] = \exp(-\alpha_E) E_{ij}[n-1] + V_E Y_{ij}[n] \tag{5}$$

where I is the grayscale of pixel at point (i, j). The input portion consists of F and L channels. The F channel receives stimulus from neurons and L is the secondary input of lateral connections to neighboring neurons, where W is the weight applied to the grayscale of other pixels within the filter window. Internal activity U is generated by the modulation of F and L. β is the linking strength constant between synapses. Y is the output of the neuron, and E is the dynamic threshold. In the linking domain, the weighted summary of the pulse output of lateral neurons is the input of the neuron, while in the input domain, external stimulus is the direct input of the neuron. When the internal activity U is greater than E, the neuron produces a pulse, also called a fire. The dynamic threshold is related to the neuron output state, indicating that the dynamic threshold will instantly increase instantly to a pre-defined value V when the neuron outputs pulse; an increased threshold prevents the immediate second pulse output from the same neuron. Alternatively, when the threshold is less than the internal activity U, the dynamic threshold attenuates exponentially until the neuron produces another pulse. At the same time, the neuron's pulse output (acting as the input of other neurons), has impact on their subsequent outputs. In this article, to adjust the value of parameters according to the experimental phenomena, the value of W is set to [0.707 1 0.707;1 0 1;0.707 1 0.707], β is 0.1 and is 0.1.

3 Hybrid Denoising Algorithm

The algorithm is verified directly with natural images captured by CMOS. There are two primary types of noise in digital images: salt-and-pepper noise and Gaussian noise. In order to achieve a satisfactory denoising result, the filter must be designed based on the premise of preserving image details and boundaries as much as possible.

In this paper, a hybrid filter designed for natural images is composed of three parts: noise detection based on PCNN, a median filter, and a self-adaptive Wiener

filter. The denoising procedure consists of three steps: (1) locate salt-and-pepper points within image; (2) apply median filter to remove the determined salt-and-pepper points; (3) apply self-adaptive Wiener filter to remove Gaussian noise.

3.1 Removal of Salt-and-Pepper Noise

Traditional filter algorithms for salt-and-pepper noise treats all pixels as noise points, which blurs the image and deteriorates the image quality when removing the noise. Thus, locating the noise point before denoising and applying a filter to the noise point only would avoid image deterioration.

In the algorithm proposed in this paper, a PCNN model is first used to locate salt-and-pepper noise. Since there is a large difference between the grayscale of a pulse noise point and its neighboring points, the fire state of most noise points are different from that of lateral points. Therefore, a simplified PCNN model is used to pre-process an image. If one neuron fails to fire while most of its neighboring neurons fire, then the corresponding pixel of this neuron is designated as a noise point. Similarly, if one neuron fires while most of its neighbors fail to fire, a median filter is applied to restore the grayscale of this pixel after it is determined to be a noise point.

3.2 Removal of Gaussian Noise

A self-adaptive Wiener filter is applied to remove Gaussian noise after the salt-and-pepper noise is filtered out [15, 16]. A Wiener filter is optimal in terms of mean-square-error, and exhibits better preservation of image details and edges than a linear filter. Self-adaptive image denoising is implemented via calculating the local square error and mean value of each pixel. The formula to determine the image local mean value and square error are described as follows:

$$u = \frac{1}{MN} \sum_{(x,y) \in s} a(x,y), \quad \delta^2 = \frac{1}{MN} \sum_{(x,y) \in s} \alpha^2(x,y) - u^2$$

where S is the chosen neighborhood of each pixel in an image, M * N is the size of the neighborhood, and $a(x, y)$ is the grayscale of the neighborhood pixels. The pixel grayscale after application of a Wiener filter is described as follows:

$$b(x,y) = u + \frac{\sigma^2 - \delta^2}{\sigma^2} [a(x,y) - u]$$

where $b(x, y)$ is the pixel grayscale after denoising, u is the local mean value, σ^2 is the local square error, and δ^2 is the square error of noise. Different sizes of filter window, i.e., 3 * 3, 5 * 5, and 7 * 7, are used to verify algorithm performance.

4 Experimental Results and Comparison

Three hybrid filters with different window sizes are used to filter out noise in many images with resolution 1280 * 1024. Results of two images (Image 1, Image 2 in Fig. 2), are presented as an example. The experiments are conducted via MATLAB. Because natural images are used directly, which contain complex noise and for which there are no corresponding ideal noiseless images available for comparison, it is difficult to judge the denoising effect of the proposed method by SNR alone. The Signal-Noise-Ratio (SNR) and global average gradient (GAG) are both used to compare the denoising effect during verification.

In the SNR formula, $g(i, j)$ is the pixel grayscale before processing, and $f(i, j)$ is the pixel grayscale after processing. In this case, the original image containing noise is used as a comparative standard. If denoising is effective, the value of $[g(i, j)–f(i, j)]^2$ will increase, and SNR will decrease accordingly. As shown in Table 1 (filter1: median filter and wavelet transform filter, filter2: median and wiener filter, filter3: median filter with PCNN and Wiener filter, filter4: median filter), for each seethe

Image1   filter2 (3*3)   filter2 (5*5)   filter2 (7*7)   filter3 (3*3)   filter3 (5*5)   filter3 (7*7)

Image2   filter2 (3*3)   filter2 (5*5)   filter2 (7*7)   filter3 (3*3)   filter3 (5*5)   filter3 (7*7)
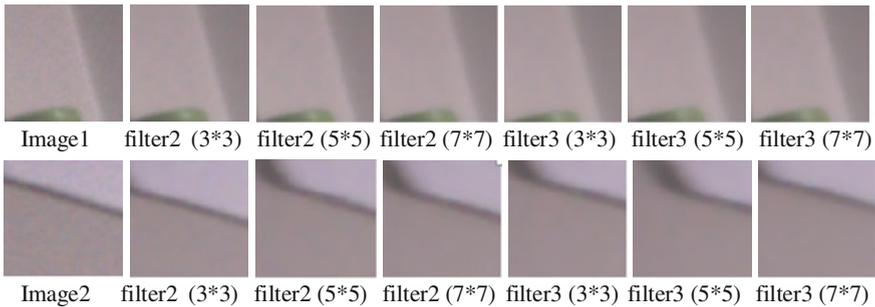
Fig. 2 Original images and denoised images

Table 1 SNR and GAG for images

Size	3 * 3		5 * 5		7 * 7	
Filters	SNR/GAG		SNR/GAG		SNR/GAG	
	Im1	Im2	Im1	Im2	Im1	Im2
Filter1	15.43/0.16	14.65/0.19	15.43/0.16	14.65/0.19	15.43/0.16	14.65/0.19
Filter2	16.80/0.18	15.95/0.21	15.80/0.18	14.75/0.20	15.15/0.17	14.09/0.19
Filter3	17.10/0.36	15.72/0.39	16.13/0.35	14.77/0.38	15.46/0.34	14.18/0.38
Filter4	16.68/0.19	15.52/0.23	15.92/0.18	14.58/0.20	14.89/0.17	13.85/0.19

parate image regardless of which filter combination is selected, when the size of Wiener filter window increases, the denoising effect improves, and SNR decreases accordingly. For different filter combinations applied to the same image, when the size of the filter window increases, hybrid filters with PCNN should outperform median and Wiener filters used separately. However, as shown in Table 1, the SNR values for hybrid filters with PCNN are larger than SNRs for median and Wiener filters. This is reasonable because image details and edges are better maintained via hybrid filters with PCNN, and thus the value of $[g(i, j)–f(i, j)]^2$ tends to decrease, which in turn causes SNR to increase. The unavailability of a noiseless ideal image is clear in that SNR alone is not enough to determine the denoising effect. Therefore, GAG was employed to evaluate the blurriness of denoised image.

GAG is used to evaluate the quality of a denoised image due to its sensitivity to image definition; the greater the average gradient, the higher the definition. As shown in Table 1, when the size of the filter window is fixed, hybrid filters with and without PCNN are applied to remove the noise from a natural image, demonstrating only a 0.12–0.3 difference between SNRs for the processed images, indicating similar denoising results. Under this premise, GAG is introduced as a criterion for the further evaluation of denoised image quality. Three filter combinations are used to filter noise out of two natural images; the GAGs of denoised images are listed in Table 1. For one specific image, regardless of filter combinations, with the enlargement of the filter window, the output image blurs and its GAG decreases. When the filter window size is fixed, the hybrid filters with PCNN outperform median and Wiener filters in terms of maintaining image details and edges; the former GAGs are almost twice of the latter. It is also shown in the Table 1 that with similar denoising results, the proposed hybrid filter with PCNN outperforms median and wavelet transform filters in terms of retaining image definition. As shown in Table 1, the hybrid filter with PCNN also outperforms filter4.

The original images and denoised images are shown in Fig. 2. Since it is difficult to observe image details in decreased dimensions, there is a magnified section to facilitate observation and comparison.

SNR and GAG are combined to evaluate the overall denoising result of natural image. Based on the premise of similar SNR values, the proposed algorithm is optimal for maintaining image details and edges. The novel hybrid filter with PCNN proposed in this paper achieves satisfactory results for natural image denoising and can be implemented on FPGA.

5 Conclusion

For natural image denoising, it is difficult to objectively evaluate the denoising result due to the lack of a noiseless ideal image. This paper combines SNR and GAG to provide a method by which to compare the natural image denoising results.

Compared with the other two studied combinations of filters, the proposed hybrid filter with PCNN maintains more image details and edges, verified by similar experimental SNR values.

References

1. Zhong, C.L.: Comparison and research of digital image denoising method. Inf. Technol. **15**, 41 (2010). (in Chinese)
2. Jin, L.H., Xiong, C.Q., Li, D.H.: Adaptive center-weighted median filter. Huazhong Univ. Sci. & Tech. (Nat. Sci. Edn) **36**, 9–12 (2008). (in Chinese)
3. Gao, M.J., Gao, K.L.: Image niose-erased and edge extraction algorithm upon PCNN and strengthening edge. Digital Technol. Appl. **11**, 135–137 (2011). (in Chinese)
4. Zou, W.J.: AN image de-noising algorithm based on PCNN. Comput. Simul. **25**, 235–237 (2008). (in Chinese)
5. Cuo, Y.C., Wang, S.B.: Method of medical ultrasonic image de-noising based on PCNN in the wavelet domain. J. Anhui Univ. **34**, 54–59 (2010). (in Chinese)
6. Liu, L., Tan, W.R.: An effective method of image gauss noise filtering based on PCNN. J. Southwest Univ. Nat. (Nat. Sci. Edn.). **38**(4), 642–647 (2012) (in Chinese)
7. El-taweel, G.S., Helmy, A.K.: Image fusion scheme based on modified dual pulse coupled neural network. Image Process. **7**, 407–414 (2013)
8. Taravat, A., Latini, D.: Fully automatic dark-spot detection from SAR imagery with the combination of nonadaptive Weibull multiplicative model and pulse-coupled neural networks. Geosci. Remote Sens. **52**, 2427–2435 (2014)
9. Mohammed, M.M., Abdelhalim, M.B., Badr, A.: An optimized PCNN for image classification. In: 2014 IEEE International Conference on Computer Engineering Conference (ICENCO), pp. 16–20. IEEE, Piscataway (2014)
10. Qayyum, A., Malik, A.S., Naufal Bin Muhammad Saad, M., Iqbal, M.: Segmentation of satellite imagery based on pulse-coupled neural network. In: 2015 IEEE International Conference on Space Science and Communication (IconSpace), pp. 393–397. IEEE, Piscataway (2015)
11. Nathiya, R., Sivaradje, G.: An efficient threshold based pulse coupled neural network model for stem cell image segmentation. In: 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), pp. 1–6. IEEE, Piscataway (2015)
12. Chen, Y., Ma, Y., Kim, D.H., Park, S.K.: Region-based object recognition by color segmentation using a simplified PCNN. IEEE Trans. Neural Networks Learn. Syst. **26**, 1682–1697 (2015)
13. Mohammed, M.M., Badr, A., Abdelhalim, M.B.: Image classification and retrieval using optimized pulse-coupled neural network. Expert Syst. Appl. **42**, 4927–4936 (2015)
14. Ikuta, C., Zhang, S.J., Uwate, K., Yang, G.A., Nishio, Y.: A novel fusion algorithm for visible and infrared image using non-subsampled contourlet transform and pulse-coupled neural network. In: 2014 IEEE International Conference on Computer Vision Theory and Applications (VISAPP), pp. 160–164. IEEE, Piscataway (2014)
15. Yin, F.P., Su, J.: Research on improved kernel wiener filter for image denoising. Laser Infrared **40**(5), 549–553 (2010) (in Chinese)
16. Guo, S.X., Tang, Y.J.: Extension and application of wiener filter in image processing. Comput. Eng. Appl. **44**(14), 178–180 (2008) (in Chinese)

Novel DDS-Based Method to Realize OFDM Baseband Signals

San-jun Liu, Meng Ma, Zi-jun Zhao, Chang Liu and Bing-li Jiao

Abstract In this paper, we propose a novel Orthogonal Frequency Division Multiplexing (OFDM) baseband signal generator to substitute the traditional Inverse Fast Fourier Transform (IFFT) based structure in an OFDM transmitter. In our architecture, we employ the Direct Digital Synthesizer (DDS) method to generate digital sub-carriers and transform frequency domain data to time domain. By making full use of the OFDM signal's two distinctive properties, we simplified the proposed architecture to the extent that multipliers can be discarded, and the DDS modules can share a small number of read-only memory (ROM) banks. Moreover, the proposed architecture has no latency while outputting the first time domain data, which leads to a quicker response compared to conventional structure. Finally, we conduct experiments on Field Programmable Gate Array (FPGA) which demonstrates that our DDS-based architecture saves more than half of the hardware resources and doubles the achievable frequency compared with traditional structure.

Keywords OFDM · Transmitter · DDS · IFFT

S. Liu · M. Ma · Z. Zhao · C. Liu · B. Jiao (✉)
School of Electronics Engineering and Computer Science, Peking University, Beijing, China
e-mail: jiaobl@pku.edu.cn

S. Liu
e-mail: lsjbox@pku.edu.cn

M. Ma
e-mail: mam@pku.edu.cn

Z. Zhao
e-mail: zhaozijun@pku.edu.cn

C. Liu
e-mail: liuchang0812@pku.edu.cn

S. Liu
Science and Technology College, Hubei University for Nationalities, Enshi, China

1 Introduction

OFDM has been accepted as a promising technique for high speed digital communications [1], and thus numerous communication standards, such as IEEE 802.11 protocol family for wireless local area network [2], and so on, have been created. Ever since OFDM was first proposed by Weinstein and Ebert [3] in 1970s, for the traditional OFDM transmitters, the time domain baseband digital signal is generated from frequency domain by inverse fast Fourier transform (IFFT), up-sampling and low pass shaping filter (LPSF) modules. The current IFFT processors are mainly based on Cooley and Tukey's structure. However, after closely inspecting these structures, we observe that they have three obvious limitations when implementing OFDM transmitter. First, the transformed data length should be a power of 2, which also constrains the number of OFDM sub-carriers to being a power of 2. Second, IFFT requires a large number of arithmetic operations, and thus its hardware complexity is very high. Third, it always takes a long time for the IFFT processors to achieve the transform, which impairs system response speed.

In view of these limitations, we propose a direct digital synthesis (DDS) [4] based OFDM transmitter architecture to substitute the traditional IFFT-based part. The proposed DDS-based architecture constructs all the sub-carriers by DDS modules, and then adds all the modulated DDS signals together to obtain the time domain digital data. The DDS method can make full use of the OFDM signal's two distinctive properties; first all the sub-carriers have the same frequency interval, and second all the frequency domain data belongs to a communal set with small number of elements. By using the first property, we make all the DDS modules share a small number of ROM banks, which consume the ROM amount even less than IFFT module. Per the second property, we use simple switches to substitute the multipliers which dramatically reduce system hardware complexity. Furthermore, since the proposed architecture is highly parallelized, it is capable of outputting the transformed time domain data one-by- one per clock cycle soon after the frequency domain data has been input to the system. Some related works, such as [5, 6], have mentioned using DDS to implement OFDM system. However, they neither simplified the OFDM transmitter's architecture to the extent that multipliers can be discarded, nor identified these advantages obtained by the DDS-based architecture. The proposed architecture is especially suitable for implementation on FPGA.

The rest of this paper is organized as follows. Section 2 gives a brief overview of OFDM transmitter, which derives the primary DDS-based architecture. Since the primary architecture is very complicated, some approaches are further investigated to simplify it in Sect. 3. In Sect. 4, experiments are carried out on FPGA to verify the performance of the proposed final architecture. Finally, Sect. 5 concludes this paper.

2 Overview of OFDM Transmitter Structure and the Primary DDS-Based Architecture

In traditional OFDM transmitter, the serial information data in frequency domain is first grouped into parallel, and then transformed by IFFT, up-sampling and LPSF modules into time domain. Assuming the frequency domain data is denoted as $X(k), k = 0, 1, \ldots, N − 1$, and the time domain data is $S(n)$, $n = 0, 1, \ldots, NM − 1$, where N is the number of sub-carriers, and M represents the interpolation factor of up-sampling. Finally, $S(n)$ is delivered into digital-to-analog converter (DAC) to generate the analog signal $s(t)$.

According to the prototype definition of $s(t)$, if the first sub-carrier's analog angular frequency is Ω_1, then the baseband complex $s(t)$ can be expressed as

$$s(t) = A \cdot \sum_{k=0}^{N-1} [X(k) \cdot \exp(jk\Omega_1 t)], t \in [0, T] \tag{1}$$

where T is the OFDM symbol's effective period, A represents the amplitude factor of each sub-carrier, and j is the imaginary unit. To investigate the relationship among $X(k)$, $S(n)$ and $s(t)$, let $s(t)$ be sampled by the frequency of the before-mentioned DAC. Then we can obtain the expression of $S(n)$ by substituting $t = nT/NM$ into Eq. (1), i.e.,

$$S(n) = A \cdot \sum_{k=0}^{N-1} \left[X(k) \cdot \exp\left(j\frac{2\pi kn}{NM} \right) \right], n = 0, 1, \ldots, NM − 1 \tag{2}$$

where k is the sub-carrier's index, and equals to $0, 1, \ldots, N − 1$.

As indicated in Eqs. (1) and (2), both the analog and digital OFDM signals are generated by summing a number of modulated sub-carriers together. Although the analog sub-carriers in Eq. (1) are difficult to generate by analog devices since their phases are hard to control and the costs are expensive, it is convenient to construct these digital sub-carriers by DDS modules. DDS is a frequency synthesis technique that employs digital data to generate sinusoidal signals of variable frequencies [4]. The advantages for DDS rely on its convenient control to its frequency and phase. Therefore, according to Eq. (2), the primary architecture of the DDS-based OFDM transmitter is shown in Fig. 1.

As shown in Fig. 1, the DDS-based OFDM transmitter comprises of N branches, each of which contains a complex exponential DDS module and a complex multiplier, denoted by "$\otimes$". The kth branch's DDS module is used to output the complex exponential signal $A \cdot \exp(jk\omega_1 n), n = 0, 1, \ldots, NM − 1$, where ω_1 denotes the first sub-carrier's digital angular frequency, and it is equal to $2\pi/NM$. Accordingly, the kth branch's output can be expressed as

Fig. 1 The primary
DDS-based architecture of
OFDM transmitter
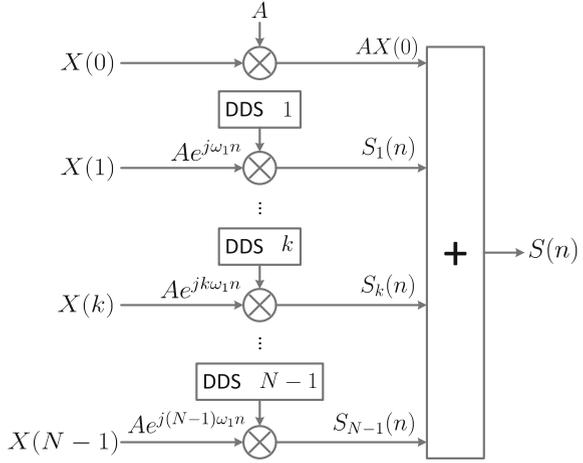
$$S_k(n) = A \cdot X(k) \cdot \exp(jk\omega_1 n)$$
$$= A \cdot [X_I(k) + jX_Q(k)] \cdot [\cos(k\omega_1 n) + j\sin(k\omega_1 n)], \tag{3}$$

where $X_I(k)$ and $X_Q(k)$ represent the in-phase and quadrature parts of $X(k)$ respectively. The complex exponential DDS module is implemented by combining two conventional DDS modules together, with one module generating the real part and the other the imaginary part. More details of DDS principle can be found in [7, 8], and the references therein. The notation "+" in Fig. 1 stands for a complex adder that can sum up all the branches' outputs in one clock cycle.

3 DDS-Based OFDM Baseband Signal Generator Structure and Its Optimization

In this section, we further optimize the proposed primary DDS-based architecture according to OFDM signal's two distinctive properties.

Property 1 *The digital angular frequency of each DDS module is integer multiple of ω_1.*

After closely inspecting the branches in Fig. 1, one can find that all the angular frequencies of DDS modules are integer multiples of ω_1. By exploiting this property, we are capable of designing a new structure, in which all the DDS modules share a communal ROM bank (we name this method as "DSAR"). The DSAR is twofold.

(1) The real part and the imaginary part of a complex exponential DDS module can share a communal ROM bank (named as "RISAR"). The communal ROM

bank stores the data sequence either in the form of $\sin(2\pi n/NM)$ or in the form of $\cos(2\pi n/NM)$, where $n = 0, 1, \ldots, NM - 1$. To generate the data sequence from one format to the other format, one just need to define the correct initial phase for the first datum, in other words, to fetch data from the right data index. For example, supposing the communal ROM bank stores the data $\sin(2\pi n/NM)$, we can generate the data sequence $\cos(2\pi n/NM)$ by fetching data from the index $NM/4$.

(2) All the DDS modules with different angular frequencies can share the communal baseband ROM bank (named as "DAFSAR"). Although different DDS modules have different digital angular frequencies, they are integer multiples of the baseband angular frequency. Therefore, if we use only one ROM bank that stores the baseband data sequence $\sin(2\pi n/NM), n = 0, 1, \ldots, NM - 1$, and decimate data one-by-one per clock cycle at different decimation ratio, then we can obtain all the desired $\sin(2\pi kn/NM)$ sequences. If one DDS module is going to fetch the data with index (denoted as y) over $NM - 1$, then the DDS module should perform a modulo NM operation, i.e. the next data index will be $y - NM$.

By combining RISAR method and DAFSAR method, each DDS can output the complex data sequence that confirms to the expression in Eq. (2). Therefore, the proposed architecture can precisely generate OFDM baseband signal defined by Eq. (1). Moreover, since the proposed architecture does not really realize a single sinusoidal signal, and it can be summarized as a simple realization method of Eq. (2), so the proposed architecture does not encounter the spurious [9] problems introduced by the traditional DDS modules.

Property 2 *Both $X_I(k)$ and $X_Q(k)$ belong to the same set with finite elements.*

Expanding Eq. (3) into real data computations, i.e.

$$
\begin{aligned}
S_k(n) = {} & A \cdot [X_I(k)\cos(k\omega_1 n) - X_Q(k)\sin(k\omega_1 n)] \\
& + j \cdot A \cdot [X_I(k)\sin(k\omega_1 n) + X_Q(k)\cos(k\omega_1 n)],
\end{aligned}
\tag{4}
$$

we remark that each complex computation includes four real calculations, two of which calculate the real part of $S_k(n)$, and the other two calculate the imaginary part. In this optimization method, we realize each real calculation in a sub-branch, therefore, each branch contains four sub-branches. Then we note that each sub-branch contains a multiplier that multiply $X_I(k)$ (or $X_Q(k)$) by $A \cdot \sin(k\omega_1 n)$ (or $A \cdot \cos(k\omega_1 n)$) per clock cycle. Usually, $X_I(k)$ and $X_Q(k)$ belong to the same set with finite elements. For convenience, we denote the set as "modulation set".

For example, in the Quadrature Phase Shift Keying (QPSK) modulation, $X_I(k)$ and $X_Q(k) \in \{1, -1\}$; in the 16 Quadrature Amplitude Modulation (16QAM), the modulation set is $\{3, 1, -1, -3\}$. Since the two input data of each multiplier belongs to their respective set, we can pre-calculate these products and store them in ROM banks. Therefore, multipliers can be replaced by switches which are used to fetch data from the corresponding ROM bank. By making use of the
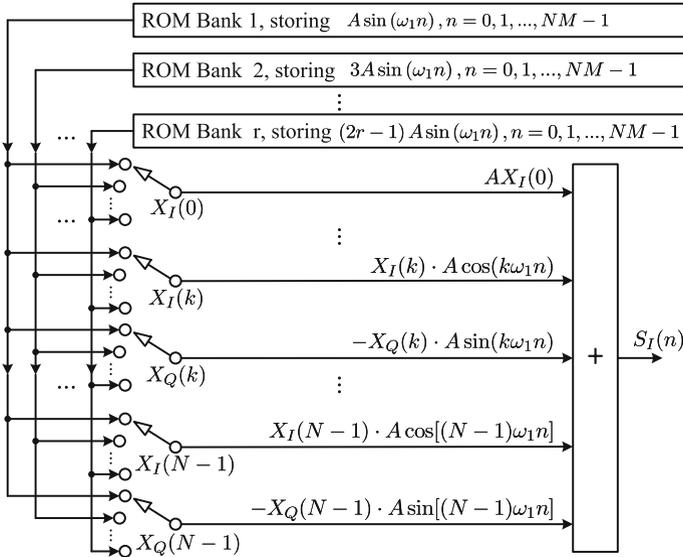
Fig. 2 The proposed final DDS-based architecture of OFDM transmitter

aforementioned two properties and their optimization methods, the proposed final DDS-based architecture can be shown in Fig. 2, where $S_I(n)$ is the in-phase part of $S(n)$, and the quadrature part $S_Q(n)$ has been omitted due to its similarity.

In Fig. 2, each sub-branch uses a switch to replace its multiplier. Data is fetched from one of the r memory banks per clock cycle, where r is equal to the number of different absolute values of the elements in the modulation set. For the lower order QAM modulation, r is usually very small, such as $r = 1$ for QPSK (the same as 4QAM) and $r = 2$ for 16QAM. Assume that the modulation scheme is x-QAM, where x is an integer power of 4, then

$$r = \frac{\sqrt{x}}{2}. \tag{5}$$

Each ROM bank stores NM data and these absolute values of the elements in the modulation set are equal to $2m - 1$, where $m = 1, 2, \ldots, r$, thus, the data sequence stored in the mth memory bank is expressed as

$$(2m - 1) \cdot A \cdot \sin(\omega_1 n), n = 0, 1, \ldots, NM - 1. \tag{6}$$

All these switches are driven by a common clock. When the system starts working, the clock generator begins to generate clock ticks. Then, at every positive edge of the clock tick, each switch fetches a datum from a ROM bank with the right initial phase at its decimation ratio. Meanwhile, the two adders sum up all their input data in a single clock cycle. These initial phases are 0°, 90°, 180° or 270°,

Table 1 The initial phases related to the data expressions

Data expression	$+\sin(\cdot)$	$+\cos(\cdot)$	$-\sin(\cdot)$	$-\cos(\cdot)$
Initial Phase	$0°$	$90°$	$180°$	$270°$
First data index	0	$NM/4$	$NM/2$	$3NM/4$

which are related to fetching the first datum with index 0, $NM/4$, $NM/2$ or $3NM/4$ respectively. Let

$$d_k(n) = \pm \left| X_{IQ}(k) \right| \cdot A \cdot f(\omega_1 n), n = 0, 1, \ldots, NM - 1 \qquad (7)$$

be the expression of these data fetched by the sub-branch's switch located in the kth branch, where $X_{IQ}(k)$ represents $X_I(k)$ or $X_Q(k)$, and $f(\cdot)$ stands for $\sin(\cdot)$ or $\cos(\cdot)$. The ROM bank number is determined by $\left| X_{IQ}(k) \right|$ according to Eq. (6). If $\left| X_{IQ}(k) \right| = 2m - 1$, then the ROM bank number is m. The initial phase depends on Eq. (7)'s sign being positive or negative and $f(\cdot)$ being $\sin(\cdot)$ or $\cos(\cdot)$. In general, there are four cases for these initial phases as shown in Table 1.

For example, if one sub-branch intends to output the sequence $-\left| X_{IQ}(k) \right| \cdot A \cdot \sin(\omega_1 n)$, $n = 0, 1, \ldots, NM - 1$, then its sign is negative, and it has the format "$-\sin(\cdot)$". As a result, its initial phase is $180°$ according to Table 1, and the first datum should be fetched from the $NM/2$th element.

4 Hardware Experiments on FPGA

In order to investigate the performance of the proposed architecture, we conducted hardware experiments on Altera FPGA platform DE2. The OFDM system framework is based on the IEEE 802.11a standard, and for comparison, two representative IFFT algorithms are employed. The first one is Quartus II Megacore [10] and the other one is Mahdavi's algorithm [11].

According to the IEEE 802.11a standard, the OFDM baseband signals contain 64 sub-carriers, thus, the IFFT transform length is 64. For our hardware tests, we use 16QAM as the modulation scheme. From Eq. (5), we can derive that $r = 2$, which means that the proposed architecture only employs two ROM banks. Assuming $M = 4$, then the data sequences stored in the two ROM banks are $A \cdot \sin(2\pi n/256)$ and $3A \cdot \sin(2\pi n/256)$ respectively, where $n = 0, 1, \ldots, 255$. Overall, the hardware complexity of the traditional OFDM transmitter contains a 64-point IFFT module, an up-sampling module and a LPSF (we use 41 taps), while in our DDS-based architecture the complexity relies on two ROM banks, 253 switches and two adders. The experimental hardware consumptions and the performance comparisons are given in Table 2.

In this experiment, we mainly compare the hardware consumption (HC), speed, and latency. HC is measured by total logic elements (TLEs) and embedded multipliers (Mults). Speed is measured using available maximum frequency

Table 2 Performance comparisons of the experiment

OFDM transmitter method	OFDM with Megacore IFFT	OFDM with Mahdavi's IFFT	Proposed DDS based architecture
Hardware complexity	64-point IFFT, interpolator, 41-tap filter	64-point IFFT, interpolator, 41-tap filter	2 ROM banks, 253 switches, 2 adders
TLEs	1362	1163	631
Mults	54	46	0
Mem/bit	19837	16934	9193
Latency/CLKs	228	136	0
Total HC ratio (%)	4.1	3.5	1.9
Max Freq/MHz	115	125	220

(Max Freq), while latency refers to the clock cycles consumed by the data transformation to output $S(0)$. All these experimental results can be observed from the development software.

To quantify the hardware saving by the proposed architecture against the traditional architecture, we complementally implemented several OFDM transmitters with different sub-carrier numbers, and plotted the hardware resource consumption as a function of sub-carrier number in Fig. 3.

Both Table 2 and Fig. 3 demonstrate that our proposed OFDM transmitter occupies less hardware resources compared with the traditional IFFT-based architecture, whereas it enables higher operation speed and lower latency. The proposed
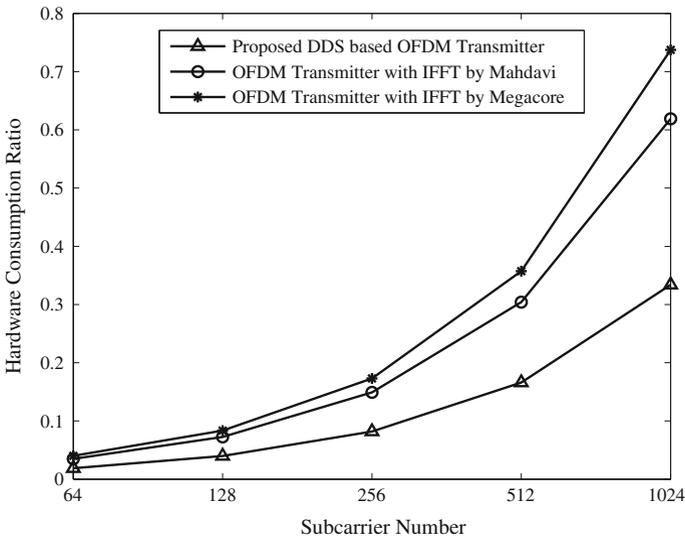
Fig. 3 Hardware resource consumption as a function of sub-carrier number

structure can save more than half of the hardware resource relative to that consumed by Megacore IFFT based architecture, and save almost half of the hardware resource compared to that consumed by Mahdavi's IFFT-based architecture. Particularly, our proposed OFDM transmitter has zero latency during the transformation to output the first time domain data, which can dramatically improve system response.

5   Conclusions

In this paper, we mainly focused on improving the traditional IFFT-based OFDM transmitter, in terms of hardware consumption, latency, achievable maximum frequency, and the arbitrary transform length. A novel DDS-based OFDM transmitter has been presented. By making use of DDS signal's two distinctive properties, we have proposed two further optimizations. The first one has made all the DDS modules share a small number of ROM banks so as to save ROM consumption; while the second one has replaced all the multipliers with switches, which enables high operating speed. The advantage lies in that the proposed OFDM transmitter can avoid using IFFT, LPSF and multipliers, and has zero latency during the transform of obtaining the first time domain data from the frequency domain. Compared with the traditional OFDM transmitter, the proposed DDS-based architecture has saved hardware resources by more than 50 %, and the available maximum frequency has outperformed the IFFT-based structure by more than 40 %.

References

1. Wang, Y., Yang, C., Ai, B.: Iterative companding transform and filtering for reducing PAPR of OFDM signal. IEEE Trans. Consum. Electron. Soc. **61**(2), 144–150 (2015)
2. Abrate, S., Straullu, S., Nespola, A., Savio P.: Self-coherent reflective FDMA-PON for next generation access and LAN architectures. In: Proceedings of IEEE International Conference on Communications, (ICC), pp. 988–993. IEEE, Piscataway (2015)
3. Weinstein, S.B., Ebert, P.M.: Data transmission by frequency-division multiplexing using the discrete fourier transform. IEEE Trans. Comm. Tech. **19**(5), 628–634 (1971)
4. Romashov, V.V., Khramov, K.K.: The comparative analysis of the noise curves of the signal shapers based on direct digital synthesizer. In: Proceedings of International Siberian Conference on Control and Communications, (SIBCON), pp. 1–5. IEEE, Piscataway (2015)
5. Sarria, D., Pallares, O., del Rio-Fernandez, J., Manuel-Lazaro, A.: Low cost OFDM based transmitter for underwater acoustic communications. In: Proceedings of IEEE OCEANS-Bergen, 2013 MTS. pp. 1–4. IEEE, Piscataway (2013)

6. Lai, W.C., Huang, J.F., Fu, C.H.: Frequency synthesizer with digital calibration in a GSM directs conversion application and MB-OFDM UWB communication. In: Proceedings of IEEE International Conference on Consumer Electronics-Taiwan, (ICCE-TW), pp. 398–399. IEEE, Piscataway (2015)
7. Cui, W.B., Zhang, X., Yu, X., Ren, Y.X.: The design of high performance X-band frequency synthesizer based on DDS and PLL. In: Proceedings of Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, (CSQRWC), pp. 97–100. IEEE, Piscataway (2013)
8. Du, X.M., Zhang, J.: DDS phase-locked swept source and study design. In: Proceedings of International Conference on Computer Science and Network Technology, (ICCSNT), pp. 146–149. IEEE, Piscataway (2011)
9. Song, Y.Y., Zhu, B.C.: Vector synthesis algorithm for spur reduction in direct digital synthesis. In: Proceedings of IEEE 11th International Conference on Signal Processing, (ICSP), pp. 115–118. IEEE, Piscataway (2012)
10. Altera corporation: Quartus II MegaCore function introduction, http://www.altera.com.cn/literature. June (2015)
11. Mahdavi, N., Teymourzadeh, R., Bin Othman, M.: VLSI implementation of high speed and high resolution FFT algorithm based on radix 2 for DSP application. In: Proceedings of the 5th Student Conference on Research and Development, pp. 1–4. IEEE, Piscataway (2007)

Observer-Based Controller Design for a Discrete Networked Control System with Random Delays

Xue Cheng and Li-Ying Zhao

Abstract This study investigates the mean square asymptotic stability problem for a networked discrete-time control system with random modeled measurement and actuation delays. The design process of the delay-dependent observer and controller is conducted to determine system stability. The gain matrices of the observer and controller are both solved according to the Lyapunov method and obtained by solving linear matrix inequalities. Finally, a simulated example is applied to illustrate the validity of the theoretical outcomes.

Keywords Observer-based control · Networked control system · Network-induced random delays · Mean-square stability · Lyapunov functional

1 Introduction

The past few years have seen increased study of networked control systems (NCSs) in the control system field because of their special type, in which the primary components transmit information through communication networks. The primary advantages of NCSs are their improved flexibility and high reliability of integrated applications, lower cost, and easier installation and maintenance [1–3]. Despite the great advantages and potential, the stability problems of networked system still persist due to their complex dynamics. For restricted bandwidths and irregular changes of date traffic, time delay and date dropout issues often emerge and degrade the performance and efficiency of the systems. Among these factors, network-induced time delays have great influence on degradation of NCS performance [4, 5]. The stability of NCSs had been previously investigated [6–8], in

X. Cheng (✉) · L.-Y. Zhao
College of Mathematics and Physics, University of Science
& Technology Beijing, Beijing, China
e-mail: cheng_xue0126@163.com

L.-Y. Zhao
e-mail: liyingzhao@ustb.edu.cn

which network-induced delays were reported as a time-varying form with a boundary.

Recently, many researchers have paid increasing attention to the modeling of random communication delays [9] using probability methods, and have made some achievements. Delays were modeled as finite-state homogeneity Markov chains in continuous and discrete systems, respectively in [10, 11]. Random time delays were also treated as Bernoulli distributed sequences [12]. Most existing results are based on a fundamental assumption, which is that network-induced delays are bounded in the studied NCS. However, in some extreme situations, some delay values might occur a great number of times while some occur less frequently, indicating that the delay values are regularly distributed.

In the current work, the stabilization problem of closed-loop NCSs is taken into account for a discrete-time plant with random modeled measurement and actuation delays. The delay-dependent observer and controller were concurrently designed to achieve mean-square asymptotic stability of the system. According to the Lyapunov stability theory, sufficient conditions in light of the existence of the aforementioned controller are demonstrated by applying and solving linear matrix inequalities (LMIs). An example and its simulation result is used to illustrate the validity and feasibility of the presented theory.

2   System Description and Problem Statement

The main components including controlled plant, sensor, controller and actuator constitute shown in Fig. 1.

The current work focuses on the discrete time-invariant linear system, supposing that the controlled plant is modeled and structured in the following form:

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) \\ y(k) = Cx(k) \end{cases} \tag{1}$$

where $x(k) \in R^n$ denotes the system sampled state vector of a plant at instant k, and $u(k) \in R^m$ and $y(k) \in R^l$ indicate the control input and measurement output
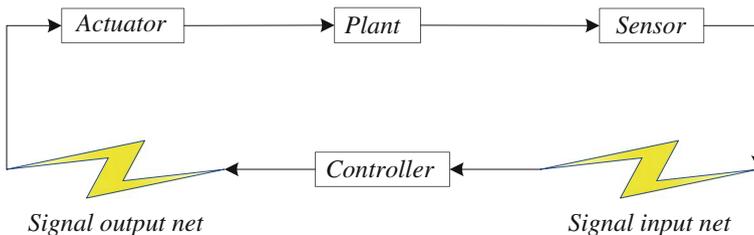
Fig. 1 Framework of NCS with random delays

vector, respectively. A, B and C are all given as constant matrices with appropriate dimensions.

Assumption 1. The transmission of the measurement output demonstrates a random variation communication delay τ_k^m which takes values from a finite set $\Delta_1 = \left\{ \tau_1^m, \tau_2^m, \ldots, \tau_p^m \right\}$, and each of the values has a certain probability given by: Prob $\left\{ \tau_k^m = \tau_i^m \right\} = \alpha_i$, here $\alpha_i \geq 0$ and $\sum_{i=1}^{p} \alpha_i = 1$, for i = 1, ... p. During actuation of each component of the system time delay τ_k^a, either might occur in some unpredictable condition which takes its values from another finite set $\Delta_2 = \left\{ \tau_1^a, \ldots, \tau_q^a \right\}$, Prob $\left\{ \tau_k^a = \tau_j^a \right\} = \beta_j$, here $\beta_j \geq 0$ and $\sum_{j=1}^{q} \beta_j = 1$, for j = 1, ..., q.

This paper introduces an indicator function : $\vartheta_{\left\{ \tau_k = s_j \right\}} = \begin{cases} 1, \tau_k = s_m \\ 0, \tau_k \neq s_m \end{cases}$ thus obtaining $E\left\{ \vartheta_{\left\{ \tau_k = s_m \right\}} \right\} = \text{Prob}\{ \tau_k = s_m \}$, for m = 1, ..., n . *Remark 1.* The idea of defining an indicator function is taken from previous research [13]. Compared to other studies, most existing conclusions assume that network-induced time delays are bounded. This differs from Assumption 1, which is more realistic according to data analysis and experience.

Hence, the measurement can be described as $y_c(k) = \sum_{i=1}^{p} \lambda_{\left\{ \tau_k^m = \tau_i^m \right\}} y(k - \tau_i^m)$, where $\lambda_{\left\{ \tau_k^m = \tau_i^m \right\}}$ is the indicator function as mentioned above.

Due to the complex structure and uncertainty, all state information of the plant is generally difficult to measure in an actual environment. Based on the forementioned analysis, the improved observer-based controller should be designed as follows:

$$\text{Observer} : \begin{cases} \tilde{x}(k+1) = A\tilde{x}(k) + Bu(k) + L(y_c(k) - \tilde{y}_c(k)) \\ \tilde{y}_c(k) = \sum_{i=1}^{p} \lambda_{\left\{ \tau_k^m = \tau_i^m \right\}} C\tilde{x}(k - \tau_i^m) \end{cases},$$

$$\text{Controller} : \begin{cases} u_c(k) = K\tilde{x}(k) \\ u(k) = \sum_{j=1}^{q} \pi_{\left\{ \tau_k^a = \tau_j^a \right\}} u_c(k - \tau_j^a) \end{cases} \quad (2)$$

where $\tilde{x}(k) \in \Re^n$ is the estimation value of state x(k), $\tilde{y}_c(k) \in R^l$ is the observer output, and $L \in R^{n \times l}$ and $K \in R^{m \times n}$ are the observer and the controller gains which are yet to be designed.

The error between the state vector and estimation value is given as follows: $\sigma(k) = x(k) - \tilde{x}(k)$. Denoting the augmented vector $\varsigma(k) = [x^T(k) \quad \sigma^T(k)]^T$, the system can be equivalent to the following equation:

$$\varsigma(k+1) = \overline{A}\varsigma(k) + \sum_{i=1}^{p} \lambda_{\left\{ \tau_k^m = \tau_i^m \right\}} \overline{B}\varsigma(k - \tau_i^m) + \sum_{j=1}^{q} \pi_{\left\{ \tau_k^a = \tau_j^a \right\}} \overline{C}\varsigma(k - \tau_j^a) \quad (3)$$
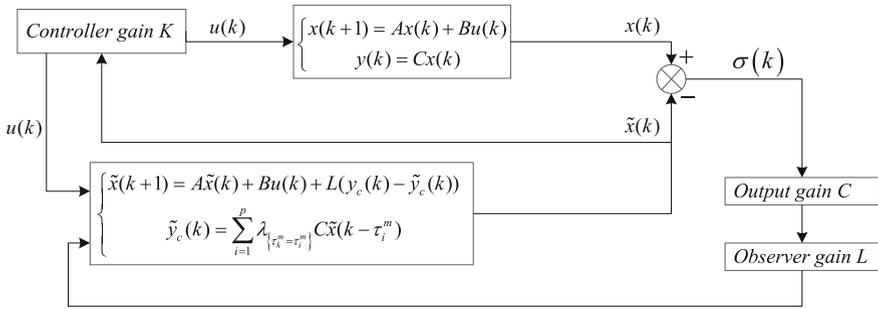
Fig. 2 The observer-based controller structure

where $\overline{A} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}, \overline{B} = \begin{bmatrix} 0 \\ -L \end{bmatrix} [0 \quad C] = \widetilde{L}\ \widetilde{C}, \overline{C} = \begin{bmatrix} B \\ 0 \end{bmatrix} [K \quad -K] = \widetilde{B}\ \widetilde{K}.$

Definition 1 For any $\varepsilon > 0$, there exists a real scalar $\delta(\varepsilon) > 0$. If $E\left\{\|x(0)\|^2\right\} < \delta(\varepsilon)$ holds true, allowing $E\left\{\|x(k)\|^2\right\} < \varepsilon$, k > 0, the closed-loop system (3) is denoted as the mean-square stable [13]. Moreover, if $\lim_{k\to\infty} E\left\{\|x(k)\|^2\right\} = 0$ holds true for arbitrary initial conditions, system (3) is said to achieve global mean-square asymptotic stability.

Based on the above description, a simple observer-based controller structure is presented in Fig. 2. The primary design concept is to estimate the delay state using the proposed observer. Though random delays occur in NCS, using the state estimation information, a state feedback input can be generated to guarantee system stability. The specific design process for the observer and controller is proposed in greater detail in Sect. 3.

3 Primary Results

In this section, sufficient conditions are presented for system (3) by applying the Lyapunov-Krasovskii functional theory.

Theorem 1 In consideration of the modeled networked control system (1) with the observer-based controller (2), the closed-loop system (3) is global mean-square asymptotically stable for any measurement delay τ_k^m and actuation delay τ_k^a that satisfies the conditions mentioned above if there exists matrices P, Q_i, R_j, M_i, N_j, Z_i, W_j, and they are all positive definite real matrices with appropriate dimensions (i = 1, ..., p, j = 1, ..., q) satisfying the following:

$$\begin{bmatrix} \Gamma_1 + \Gamma_2^T + \Gamma_2 + \Pi_1\Gamma_3\Pi_1^T + \Pi_2\Gamma_4\Pi_2^T & \Gamma_5 \\ * & -\Gamma_6 \end{bmatrix} < 0 \tag{4}$$

$$\Gamma_1 = diag\left\{ -P + \sum_{i=1}^{p} Q_i + \sum_{j=1}^{q} R_j, -Q_1, \ldots, -Q_p, -R_1, \ldots, -R_q \right\}, \Gamma_2 = \left[\sum_{i=1}^{p} M_i + \sum_{j=1}^{q} N_j - M_1 \cdots - M_p - N_1 \cdots - N_q \right],$$

$$\Gamma_3 = diag\{P, P, \ldots, P\}, \Gamma_4 = diag\{H, \ldots, H\}, \Gamma_5 = \left[\sqrt{\tau_1^m} M_1 \cdots \sqrt{\tau_p^m} M_p \sqrt{\tau_1^a} N_1 \cdots \sqrt{\tau_q^a} N_q \right],$$

$$\Gamma_6 = diag\{Z_1, \ldots, Z_p, W_1, \ldots, W_q\}, \Xi_i = \begin{bmatrix} \Xi_{i1} & \cdots & \Xi_{iq} \end{bmatrix}, \tilde{\Xi}_i = \begin{bmatrix} \tilde{\Xi}_{i1} & \cdots & \tilde{\Xi}_{iq} \end{bmatrix},$$

$$\Theta_i = \begin{bmatrix} \Theta_{i1} & \cdots & \Theta_{iq} \end{bmatrix}, H = \sum_{i=1}^{p} \tau_i^m Z_i + \sum_{j=1}^{q} \tau_j^a W_j, \Xi_{ij} = \sqrt{\alpha_i \beta_j} \bar{A}^T,$$

$$\tilde{\Xi}_{ij} = \sqrt{\alpha_i \beta_j} (\bar{A} - I)^T, \Theta_{ij} = \sqrt{\alpha_i \beta_j} \bar{B}^T,$$

$$\Upsilon_i = \begin{bmatrix} \sqrt{\alpha_i \beta_1} \bar{C}^T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\alpha_i \beta_q} \bar{C}^T \end{bmatrix}, \Pi_1 = \begin{bmatrix} \Xi_1 & \Xi_2 & \cdots & \Xi_p \\ \Theta_1 & 0 & \cdots & 0 \\ 0 & \Theta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Theta_p \\ \Upsilon_1 & \Upsilon_2 & \cdots & \Upsilon_p \end{bmatrix} \Pi_2 = \begin{bmatrix} \tilde{\Xi}_1 & \tilde{\Xi}_2 & \cdots & \tilde{\Xi}_p \\ \Theta_1 & 0 & \cdots & 0 \\ 0 & \Theta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Theta_p \\ \Upsilon_1 & \Upsilon_2 & \cdots & \Upsilon_p \end{bmatrix}$$

Proof At the beginning of the certification, make a definition $\phi(k) \overset{\Delta}{=} \varsigma(k+1) - \varsigma(k)$.

Then apply the Lyapunov stability theory, and select a *V*-function with the following structure:

$$V(\varsigma(k), k) = \varsigma^T(k)P\varsigma(k) + \sum_{i=1}^{p} \sum_{u=k-\tau_i^m}^{k-1} \varsigma^T(u)Q_i\varsigma(u) + \sum_{j=1}^{q} \sum_{v=k-\tau_j^a}^{k-1} \varsigma^T(v)R_j\varsigma(v)$$

$$+ \sum_{i=1}^{p} \sum_{u=-\tau_j^m}^{-1} \sum_{t=k+u}^{k-1} \phi^T(s)Z_i\phi(s) + \sum_{j=1}^{q} \sum_{v=-\tau_j^a}^{-1} \sum_{s=k+v}^{k-1} \phi^T(s)W_j\phi(s) \tag{5}$$

where P, Q_i, R_j, Z_i and W_j must be determined.

Differentiating $V(\varsigma(k), k)$ according to system (3) and obtaining the mathematical expectation yields $E\{\Delta V(\varsigma(k), k)\} = E\{V(\varsigma(k+1), k+1)\} - V(\varsigma(k), k)$.

By using the free-weighting matrix method according to the Leibniz-Newton formula, for any matrices with appropriate dimensions (i = 1, …, p, j = 1, …, q),

$$M_i = \begin{bmatrix} M_{0i}^T & M_{1i}^T & \cdots & M_{pi}^T & 0 & \cdots & 0 \end{bmatrix}^T, \quad N_j = \begin{bmatrix} N_{0j}^T & 0 & \cdots & 0 & N_{1j}^T & \cdots & N_{qj}^T \end{bmatrix}^T,$$

$$\xi^T(k)M_i \left[\varsigma(k) - \varsigma(k - \tau_i^m) - \sum_{r=k-\tau_i^m}^{k-1} \phi(r) \right] = 0,$$

$$\xi^T(k)N_j \left[\varsigma(k) - \varsigma(k - \tau_j^a) - \sum_{h=k-\tau_j^a}^{k-1} \phi(h) \right] = 0.$$

Set $\quad \xi(k) = \begin{bmatrix} \varsigma^T(k) & \varsigma^T(k-\tau_1^m) & \cdots & \varsigma^T(k-\tau_p^m) & \varsigma^T(k-\tau_1^a) & \cdots & \varsigma^T \end{bmatrix}$
$(k-\tau_q^a)]^T$, thus obtaining:

$$E\{\Delta V(\varsigma(k),k)\} \leq \xi^T(k)\Lambda\xi(k) - E\left\{ \sum_{i=1}^{p} \sum_{r=k-\tau_i^m}^{k-1} \left(\xi^T(k)M_i + \phi^T(r)Z_i \right) Z_i^{-1} \left(M_i^T\xi(k) + Z_i\phi(r) \right) \right\}$$

$$-E\left\{ \sum_{j=1}^{q} \sum_{h=k-\tau_j^a}^{k-1} \left(\xi^T(k)N_j + \phi^T(h)W_j \right) W_j^{-1} \left(N_j^T\xi(k) + W_j\phi(h) \right) \right\}$$

$$(6)$$

Note that according to the Schur complement theorem, (4) is equivalent to $\Lambda = \Gamma_1 + \Gamma_2^T + \Gamma_2 + \Pi_1\Gamma_3\Pi_1^T + \Pi_2\Gamma_4\Pi_2^T + \Gamma_5\Gamma_6^{-1}\Gamma_5^T < 0$; hence, we can obtain (6) <0, so that $E\{V(\varsigma(k+1),k+1)\} < V(\varsigma(k),k)$ hold true. According to existence theorem analysis, a real number must exist $0 < \gamma < 1$ satisfying $E\{V(\varsigma(k+1),k+1)\} \leq \gamma V(\varsigma(k),k)$. Recursive through this relationship, determine a summation and limit on both sides of the formula. Thus, $\lim_{N\to\infty} E\left\{ \sum_{k=0}^{N} [\varsigma^T(k)\varsigma(k)] \right\} \leq \frac{1}{(1-\gamma)\lambda_{\min}(P)} V(\varsigma(0),0)$ holds true, which indicates $E\left\{ \|\varsigma(k)\|^2 \right\} \to 0$ as $k \to \infty$. According to Definition 1, system (3) is mean-square asymptotically stable, and the proof for Theorem 1 is complete.

Next, based on Theorem 1, a solution is proposed to the observer-based controller design problem regarding system stability. Consider the matrix inequality (4), by pre- and post-multiplying $diag\left\{ \overline{P}, \overline{P} \otimes \Psi_1, \overline{P}^{-1} \otimes \Psi_1, \Psi_1, \cdots, \Psi_1 \right\}$ on both sides through a congruence matrix linear transformation, where $\Psi_1 = diag\{I, I, \cdots, I\}$.

By using $-\overline{P}Z_i^{-1}\overline{P} \leq \overline{Z}_i - 2\overline{P}$, $-\overline{P}W_j^{-1}\overline{P} \leq \overline{W}_j - 2\overline{P}$ to obtain linearization, define

$$P^{-1} = \overline{P}, Q_i = \overline{P}^{-1}\overline{Q}_i\overline{P}^{-1}, R_j = \overline{P}^{-1}\overline{R}_j\overline{P}^{-1}, Z_i = \overline{Z}_i^{-1}, W_j = \overline{W}_j^{-1}, \tilde{K} = \overline{K}\overline{P}^{-1}, \tilde{L}\tilde{C} = \overline{L}\overline{P}^{-1}$$

$$M_j = diag\left\{ \overline{P}^{-1}, \overline{P}^{-1} \otimes \Psi_1 \right\} \overline{M}_i diag\left\{ \overline{P}^{-1}, \overline{P}^{-1} \otimes \Psi_1 \right\}, N_j = diag\left\{ \overline{P}^{-1}, \overline{P}^{-1} \otimes \Psi_1 \right\} N_j diag\left\{ \overline{P}^{-1}, \overline{P}^{-1} \otimes \Psi_1 \right\}.$$

Then, based on Theorem 1, the observer and controller gains are determined.

4   Numerical Example

In this section, to test the feasibility of the proposed conclusion, consider the following example with the applied combination of the observer-based controller designed in this paper:

Fig. 3 State responses of system (3) under the observer-based controller
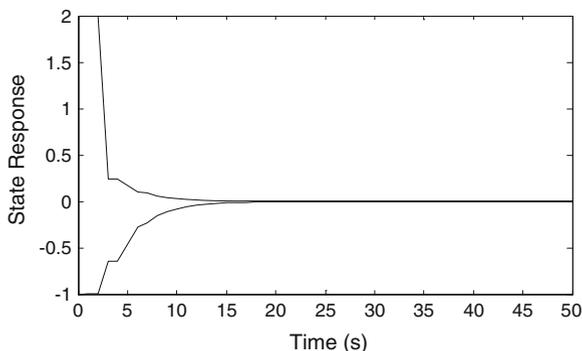
$$A = \begin{bmatrix} 0.0001 & 0.0003 \\ 0.0005 & 0.0001 \end{bmatrix}, \quad B = \begin{bmatrix} -0.002 \\ 0.0053 \end{bmatrix}, \quad C = \begin{bmatrix} 0.03 & 0.0036 \end{bmatrix}.$$

Assume that measurement delays exist in $\tau_k^m \in \{1, 2\}$ and that actuation delays exist in $\tau_k^a \in \{1, 2\}$ with probabilities respectively given by Prob $\{\tau_k^m = 1\} = 0.64$, Prob $\{\tau_k^m = 2\} = 0.36$, Prob $\{\tau_k^a = 1\} = 0.9$, Prob $\{\tau_k^a = 2\} = 0.1$

By calculating the matrix inequality after linear transformation through $\tilde{K} = \bar{K}\bar{P}^{-1}$, $\tilde{L}\tilde{C} = \bar{L}\bar{P}^{-1}$, where $\tilde{K} = [K - K]$, the observer and controller gain matrices are obtained, and represented as: $L = \begin{bmatrix} 9.3033 \\ 1.1267 \end{bmatrix}, K = \begin{bmatrix} -12.6127 & 35.5275 \end{bmatrix}$.

The state responses with the initial condition are assumed by $[2 \ -1]^T$. Though random delays occur in this system, using the observer-based controller, the values of the system sub-states all converge to zero as shown in Fig. 3. This simulation result demonstrates that the proposed observer-based controller is effective in solving the stabilization problem of NCSs with random modeled delays, while demonstrating the rationality of the design methodology presented in this paper.

5 Conclusion

The problem of stabilization for discrete-time NCSs with random delays which are modeled in a modality that adopts values according to certain probabilities has been discussed and solved in this study paper. The improved observer-based controller design procedure is proposed for system stabilization in order to achieve a mean-square asymptotically stable NCS. Using the Lyapunov method and LMI techniques, stability analysis has been conducted and sufficient stabilization condition has been addressed. Finally, the effectiveness and feasibility of the proposed conclusion have been illustrated by a numerical example.

References

1. Tipsuwan, Y., Chow, M.Y.: Control methodologies in networked control systems. Control Eng. Pract. **11**, 1099–1111 (2003)
2. Hespanha, J.P., Naghshtabrizi, P., Xu, Y.: A survey of recent results in networked control systems. Proc. IEEE **95**, 138–162 (2007)
3. Zhivoglyadov, P.V., Middleton, R.H.: Networked control design for linear systems. Automatic **39**, 743–750 (2003)
4. Hua, C., Yu, S., Guan, X.: Finite-time control for a class of networked control systems with time-varing delays and sampling jitter. Inter. J. Auto. Comput. **12**, 448–454 (2015)
5. Lu, Z., Wang, L., Zhang, F., Xu, F.: The design of robust controller for networked control system with time delay. Math. Problem. in Engineering. **2015**, 1–8 (2015)
6. Walshi, G.C., Ye, H., Bushnell, L.: An improved delay-dependent stability criterion of networked control system. J. Franklin Ins. **351**, 1540–1552 (2014)
7. Tang, B., Wang, J., Zhang, Y.: A delay-distribution approach to stabilization of networked control systems. IEEE Trans. Control Netw. Syst. **1** (2015)
8. Jungers, M., Castelan, E.B., Moraes, V.M., Moreno, U.F.: A dynamic output feedback controller for NCS based on delay estimates. Automatic **49**, 788–792 (2013)
9. Seshadhri, S., Ayyagari, R.: Dynamic controller for network control systems with random communication delay. Int. J. Syst. Control Commun. **3**, 178–193 (2011)
10. Zhang, L.Q., Shi, Y., Chen, T.W., Huang, B.: A new method for stabilization of ncss with random delays. IEEE Trans. Auto. Control **50**, 1178–1181 (2015)
11. Qiu, L., Shi, Y., Yao, F., Xu, G., Xu, B.: Network-based robust control for linear systems with two-channel random packet dropouts and time delays. IEEE Trans. Auto. Control **45**, 1450–1462 (2015)
12. Luan, X.L., Shi, P., Liu, F.: Stabilization of networked control systems with random delays. IEEE Trans. Ind. Electron. **58**, 4323–4330 (2011)
13. Gao, H.J., Meng, X.Y., Chen, T.W.: Stabilization of networked control systems with a new delay characterization. IEEE Trans. Auto. Control **53**, 2142–2148 (2008)

Performance Analysis of Optical Balanced Coherent Detection Based on an Acousto-Optic Deflection System

Hong-Yan Jiang, Ning He and Xin Liao

Abstract Based on the principle of acousto-optic (AO) diffraction, the mechanism of acousto-optic deflection for frequency measurement and the noise characteristics of balanced detection in a coherent system are analyzed theoretically. For signal to noise ratio (SNR) improvement and multiple simultaneous signals parallel processing in acousto-optic signals processing environment, a test platform for frequency measurement is established to discuss the method of SNR improvement while the ability of signals detection and frequency resolution are tested. In the operational bandwidth, the system using acousto-optic effect and balanced coherent detection can improve the output SNR to some extent. Experiments have shown that a balanced coherent configuration has 5 dB improvements in output SNR over a single-detector coherent configuration, and frequency resolution of the system is less than 0.1 MHz.

Keywords Acousto-optic deflection · Optical coherence · Balanced detection · Multiple signals processing · Frequency resolution

1 Introduction

Weak-signals detection and processing in increasingly complex and dense electromagnetic signal environment, such as radar signals sorting, threat identification and directed jamming, make it very important to develop radio frequency (RF) receivers with large bandwidth, high sensitivity and parallel processing ability,

H.-Y. Jiang · N. He (✉) · X. Liao
School of Information and Communication, Guilin University of Electronic Technology, Guilin, China
e-mail: eicnhe@guet.edu.cn

H.-Y. Jiang
e-mail: jhy8499@163.com

X. Liao
e-mail: liaoshin@guet.edu.cn

which can achieve quick, accurate, safe and seamless detection. With the development of acousto-optic technology and laser applications, receivers based on acousto-optic effect have great advantages over conventional receivers for frequency measurement and signal processing.

Wireless signal is susceptible to environmental random-noise interference in transmission, thus in receivers the accumulated noise lead to further deterioration of SNR. In addition, in optical signal transmission there is phase-fluctuation noise resulted from phase mismatch in optical coherent detection, which affects the output performance of optical receivers. Acousto-optic systems combining optical coherent detection technique with acousto-optic technique have several great characteristics, such as high-speed parallel processing ability, large dynamic range, high resolution, particular large bandwidth, high probability of intercept (POI) and so on. All these advantages make the receiver well suited to application in increasingly dense and complex electromagnetic signals environment. In this paper, based on the acousto-optic deflection and optical balanced coherent detection, a system for frequency measurement is established, and the method of SNR improvement is discussed. Meanwhile frequency resolution and noise characteristics are measured and analyzed.

2 The Principle of Acousto-Optic Deflection

The operating principle of acousto-optic deflection is based on acousto-optic interaction. Through a piezoelectric transducer, electrical signals are converted into acoustic waves which carry information of original electrical signals and propagate in acousto-optic interaction medium. This action causes refractive index variations by grating effect. Consequently diffraction beams, whose amplitude and phase are linearly proportional to those of the applied signals respectively, are not only deflected at prescribed angles as a linear function of the frequencies of the applied electrical signals but also, because of the Doppler effect, take a new circular frequency ω_N given by

$$\omega_N = \omega + N\Omega, \quad (N = 0, \pm 1, \pm 2 \ldots) \tag{1}$$

where N is the order of the diffracted beam, ω and Ω are the circular frequencies of the laser beam and acoustic waves respectively.

AO deflectors as shown in Fig. 1 operate at the Bragg regime, where the diffracted beam will appear predominantly in a single(+1 or −1) order and its intensity will be maximum when laser beam is incident at the Bragg angle θ_B, which is given by $\theta_B \approx \frac{\lambda}{2n\lambda_S} = \frac{\lambda}{2nv}f$. The deflection angle θ of the first order is described by

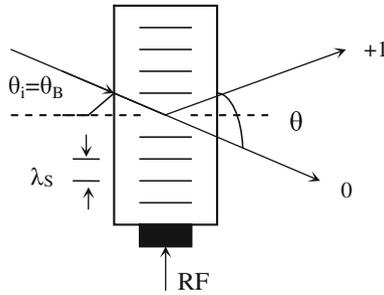$$\theta = \frac{\lambda}{nv}f \tag{2}$$

Fig. 1 Bragg diffraction geometry

where λ is the wavelength of the laser beam in free space, n is the index of refraction and λ_s is the acoustic wavelength, v is the acoustic velocity and f is the frequency of the signal launched into the transducer.

If multiple simultaneous electrical signals with different frequencies are applied to the piezoelectric transducer, multiple corresponding acoustic waves will be generated. Interacting with the laser beam, every acoustic wave will generate a principal diffracted beam separated from the incident beam by an angle [1, 2]:

$$\theta_j = \frac{\lambda}{nv} f_j, \quad (j = 1, 2, 3, \ldots, N) \tag{3}$$

3 Balanced Detection and Noise-Characteristics Analysis in Coherent Systems

A balanced configuration has superior characteristics over a single-detector configuration, i.e. local-oscillator excess-noise and shot-noise reduction, common-mode noise suppression [3]. A block diagram of the balanced coherent detection is presented in Fig. 2.

After received optical signal field Es and optical local oscillator wave E_L respectively are input to the coupler, the coupler will output two beams ($E_{S1} + E_{L1}$)
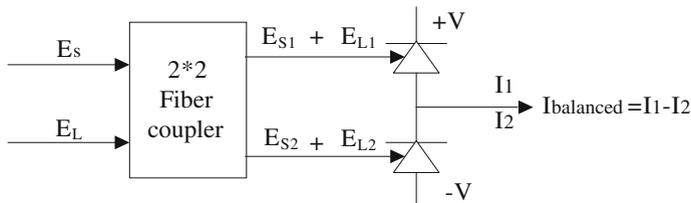

Fig. 2 Schematic diagram of balanced detection

and $(E_{S2} + E_{L2})$ into two photodetectors respectively producing two currents I_1 and I_2. The difference between the two currents is the output current of a balanced detection system which is given by

$$
\begin{aligned}
I_{balanced}(t) =\ & \frac{e}{h\upsilon}\left\{[(1-\varepsilon)\eta_1 - \varepsilon\eta_2]A_S^2 + [\varepsilon\eta_1 - (1-\varepsilon)\eta_2]A_L^2\right\} \\
& + \frac{e}{h\upsilon}\left\{2(\eta_1+\eta_2)\sqrt{\varepsilon(1-\varepsilon)}A_S A_L\cos[(\omega_L-\omega_S)t + (\varphi_L-\varphi_S)]\right\} \\
& + [n_1(t) - n_2(t)]
\end{aligned}
\tag{4}
$$

with A = amplitude, φ = phase, ω = circular frequency, ε = beam splitting ratio, η = quantum efficiency of the detector, υ = optical frequency, e = electron charge, h = Planck's constant and n(t) = noise current. Where subscripts S and L represent for the signal and local fields respectively, subscripts 1 and 2 represent for the upper and lower branches respectively.

In front of a receiver photodetectors have much impact on output-noise performance, after photoelectric conversion output current contains local oscillator (LO) excess noise, thermal noise and shot noise. In balanced detection the total shot (excess) noise is the difference of shot (excess) noise between the two detectors. The SNR of the balanced detection is given by

$$
\begin{aligned}
SNR_{balanced} &= \frac{P_{balanced}}{P_{b-shot} + P_{b-excess} + P_{b-thermal}} \\
&= \frac{2\varepsilon(1-\varepsilon)P_S(\eta_1+\eta_2)^2}{h\upsilon(\sqrt{\eta_1\varepsilon} - \sqrt{\eta_2(1-\varepsilon)})^2 BW + 2e\gamma P_L[\eta_1\varepsilon - \eta_2(1-\varepsilon)]^2 BW + k_1 N_{b-thermal}}
\end{aligned}
\tag{5}
$$

With

$$
k_1 = \frac{(h\upsilon)^2}{4P_L R_1 e^2}
\tag{6}
$$

With the same input (received signal and local oscillator fields are the same as that of the balanced detection), the SNR of a single-detector detection system (taking the first detector for example) can be expressed by

$$
\begin{aligned}
SNR_{single} &= \frac{P_{single}}{P_{s-shot} + P_{s-excess} + P_{s-thermal}} \\
&= \frac{2(1-\varepsilon)P_S\eta_1}{h\upsilon BW + 2e\gamma P_L\eta_1\varepsilon BW + k_2 N_{s-thermal}}
\end{aligned}
\tag{7}
$$

where

$$k_2 = \frac{(h\upsilon)^2}{4P_L R_2 \eta_1 \varepsilon e^2} \tag{8}$$

R_1 and R_2 represents for the load impedance of balanced and single-detector system respectively; γ represents for the change of LO-noise power as the square of LO average power, which has a typical number range from 10^2 to 10^6 A^{-1}; BW is the bandwidth of the receiver; P represents for power [4–7].

Assuming the signal and LO beam is generated by the same laser and $\gamma = 10^5$ A^{-1}, Ps = 3.5 %P_{laser}, P_L = 36 %P_{laser}, ε = 0.5. According to Eqs. (5) to (8), simulation about theoretical changes of the SNR as the laser power in balanced and single-detector coherent receivers is shown in Fig. 3.

As can be seen from Fig. 3, the SNR of the ideal balanced-detection system (in matching state) can be constantly enhanced by increasing the laser output power. However, in practical applications, the excess noise and shot noise cannot be completely eliminated, so the SNR of a balanced detection system (in mismatching state) or a single-detector system tends to a saturation value with the increase of the laser output power. In comparison, the SNR of the balanced system is greater than that of the single-detector system because some of excess noise and shot noise is eliminated by balanced detection. Here only the case of $\varepsilon = 0.5$ is discussed, more information about the relation between the SNR and beam splitting ratio in balanced coherent detection systems can be found in the previous studies of Wang [4]. Therefore, to get superior performance of balanced detection in coherent receivers, optimal splitting ratio and a suitable laser should be used according to actual situation.
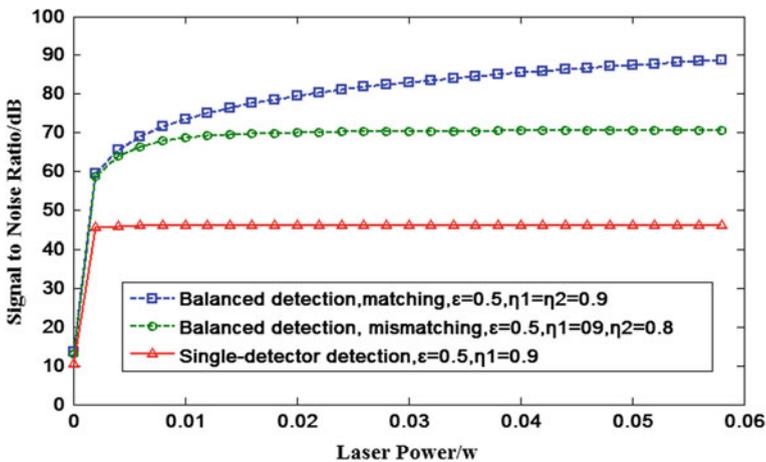

**Fig. 3** SNR simulation of balanced detection and single-detector detection

4 System Composition and Performance Analysis

The block diagram of the coherent system for frequency measurement is shown in Fig. 4, which uses acousto-optic deflection, balanced detection and optical fiber.

The test platform mainly contains six parts: a single longitudinal mode laser, a preprocessing circuit, two injectors, an acousto-optic deflector, a 2×2 optical fiber coupler and a set of balanced detectors. Laser beam is split up into signal-carrier and reference beam (equal to LO beam) by a dispersion prism. RF signals are input to acousto-optic deflector to establish phase grating for frequency modulation. Photodetection of the mixing between signal and reference beam is made in the photodetectors and then output current is filtered and amplified.

In the system, techniques used to improve the SNR are as follows [8–10]:

1. The laser with a narrow linewidth and high stability is helpful to enhance the stability of the system.
2. Compared to narrow-band filters of the direct optical detection, IF filters of coherent detection have higher performance, which can improve frequency selectivity so that anti-interference ability is greatly enhanced in transmission.
3. Filtering characteristic of acousto-optic deflectors is used. Because rand noise cannot cause steady grating in acousto-optic deflectors, its diffraction light is unavailable. Thus noise can be effectively restrained.
4. A balanced coherent detection system can reduce local oscillator excess and shot noise and has a more efficient use of the local oscillator power.
5. Polarization-maintaining (PM) fibers are used to improve matching in coherent detection. By injecting laser beams into PM fibers, the influence that the background noise and vibration has on the combination of signal and reference beams is reduced and polarization matching is fulfilled. Hence all of SNR, detected efficiency and stability can be improved.
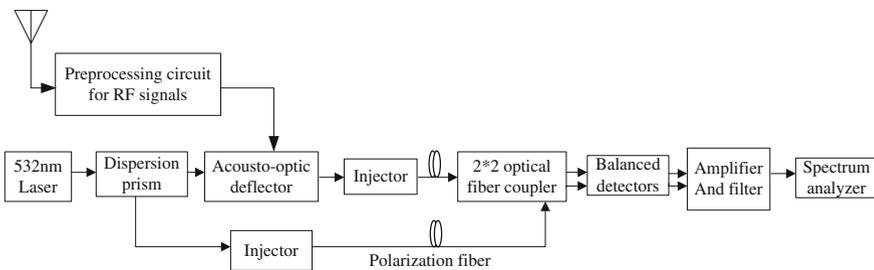
Fig. 4 System block diagram

5 Experimental Methods and Results

In the experimental test platform, a 50 mw single longitudinal mode laser, which is with a wavelength of 532 nm and a linewidth of 30 MHz, provides illumination for both signal and reference arms. The acousto-optic deflector used is with 800 MHz center frequency and 100 MHz bandwidth.

The frequency of the RF signal input to the acousto-optic deflector is 800 MHz. In the case that RF signal source is turn off, noise characteristics of the system with or without laser beams can be analyzed. In order to make comparison, both of the single-detector and balanced coherent optical systems are tested. The spectrum analyzer uses the function of measuring channel power to measure noise characteristics of the two systems, with integration bandwidth set as 130 MHz (3 dB). Noise spectrums of the single-detector and balanced coherent systems have been shown in Fig. 5.

When the laser is turned off, noise mainly consists of dark current and noise floor, and noise spectrums of the two systems can be seen from Fig. 5a, c. While the laser is turned on, noise is mainly produced from quantum noise and noise spectrums of the two systems are shown in Fig. 5b, d.

By comparing Fig. 5a with c, in the case of no laser beams, it is known that the noise of the single-detector detection system and that of the balanced detection system is almost equal, about −42 dBm. For the single-detector system, compared with that in the case of no laser beams, the noise floor is increased by about 13 dB when the laser is turned on, and the results are shown in Fig. 5a, b respectively. But seen from 5c, d, for the balanced detection system, noise floor is just increased by about 2 dB when the state of the laser is changed from off to on. Hence balanced detection has good performance in noise suppression. With same signals input, the output carrier to noise ratio (CNR) of balanced detection can be improved by about 5 dB compared with that of single-detector in the coherent system, which is shown in Fig. 6.

The spectrum of two RF signals received by this balanced coherent detection system is shown in Fig. 7. The frequency of curse 1 is 800 MHz and that of curse 2 is 799.9 MHz, thus two signals with frequency difference 0.1 MHz can be properly detected.

Based on experiments, the optical balanced coherent detection system has been proven to be stable and reliable. It can achieve frequency measurement for multiple simultaneous signals. In the operational bandwidth, frequency resolution is less than 0.1 MHz and output SNR can be improved by about 5 dB compared with the single-detector system.
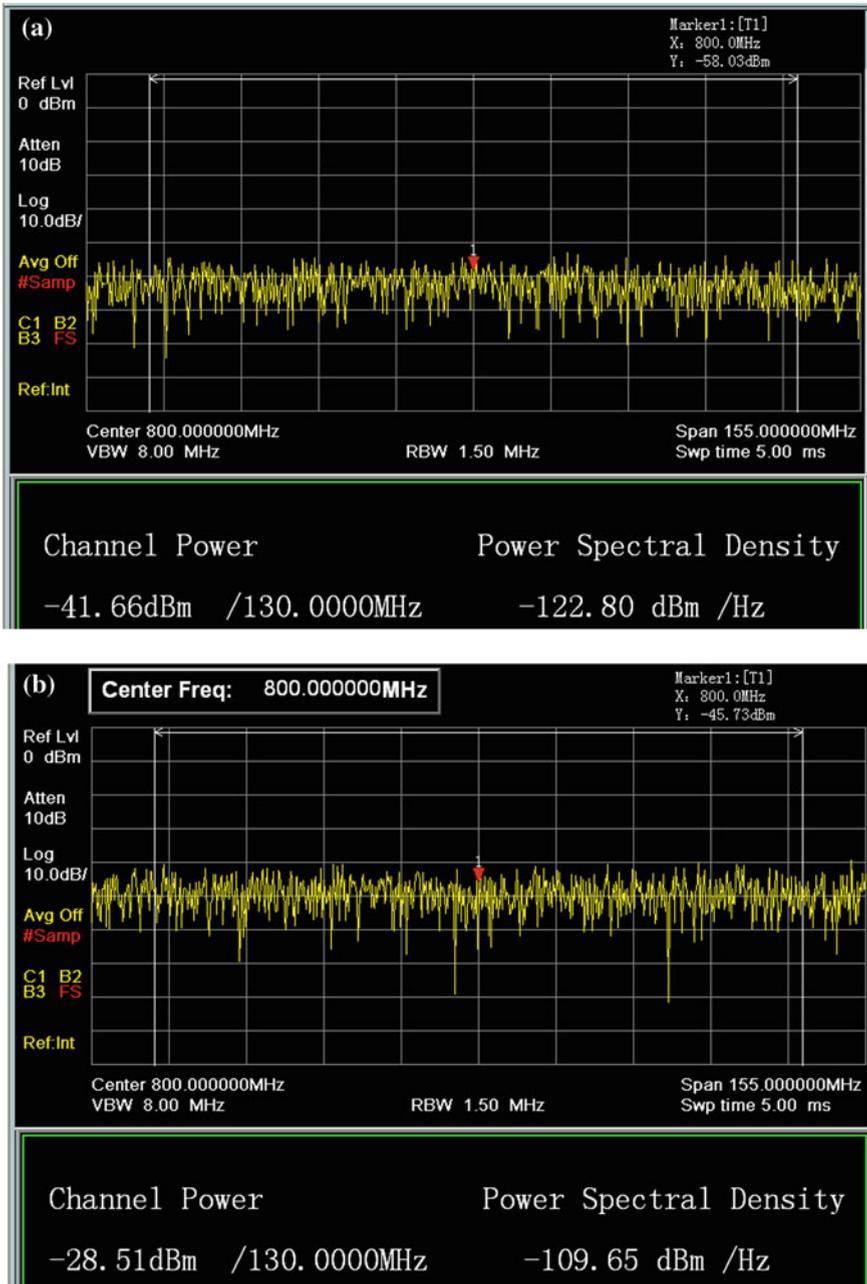
Fig. 5 Noise spectrums of the single-detector detection system and balanced detection system, **a** single-detector detection without laser beams, **b** single-detector detection with laser beams, **c** balanced detection without laser beams, **d** balanced detection with laser beams
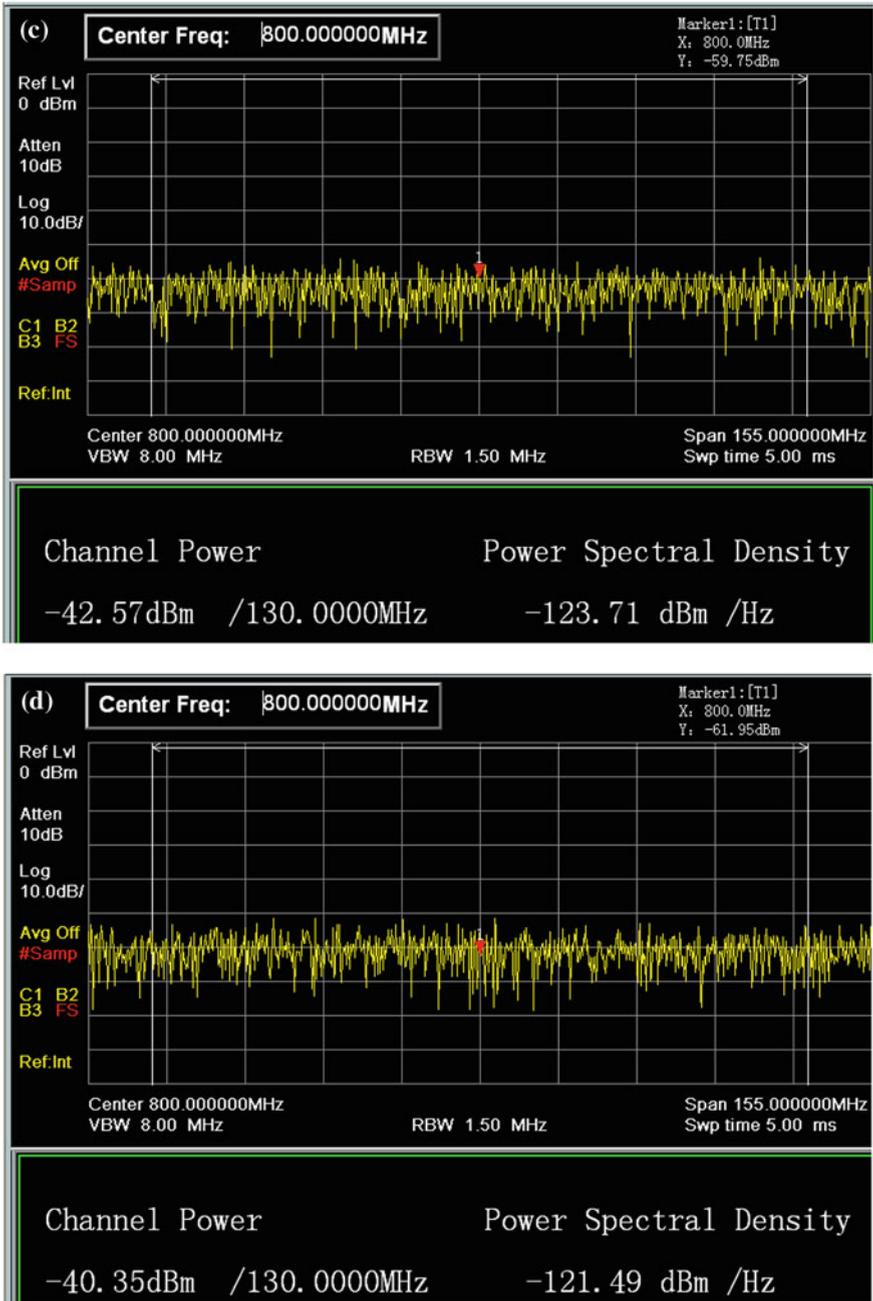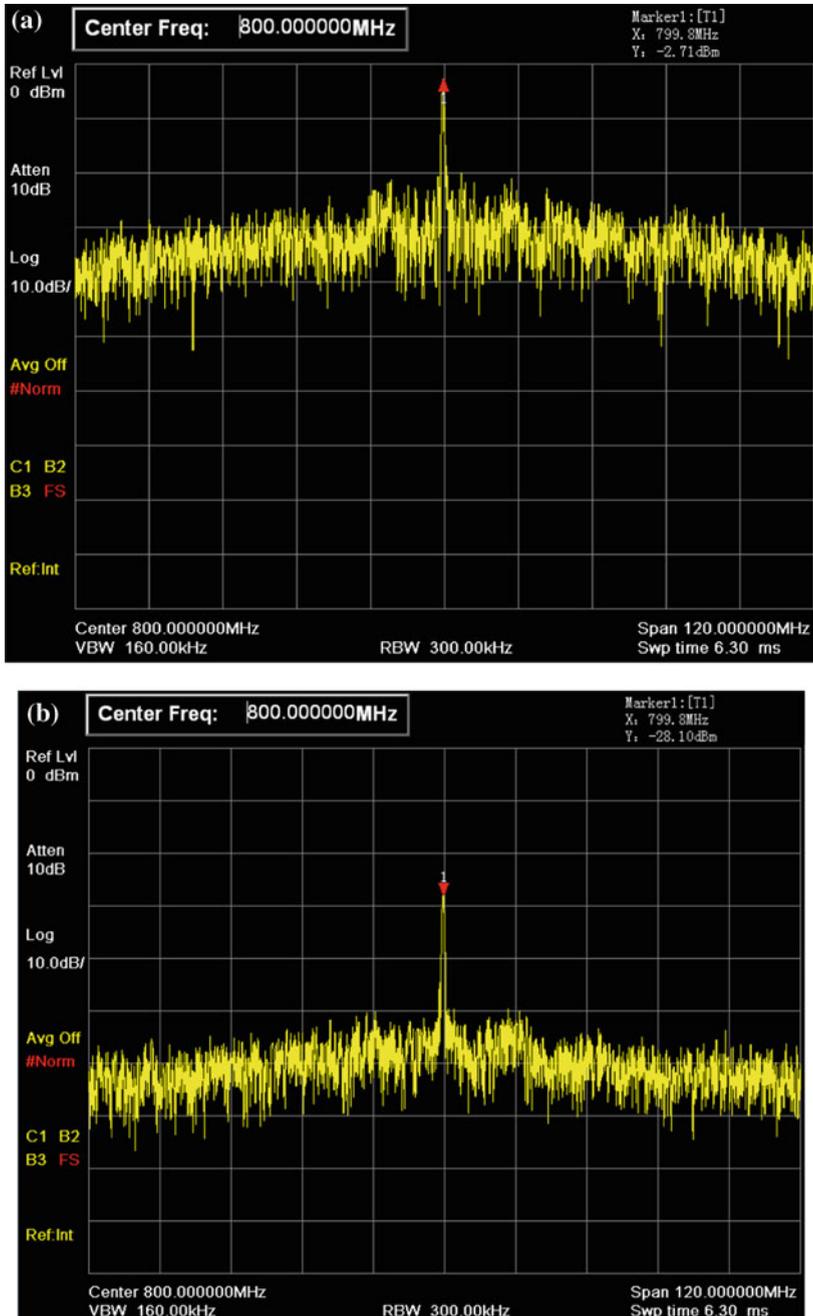
Fig. 5 (continued)

Fig. 6 The comparison between the CNRs of single-detector and balanced detection, **a** the CNR of single-detector detection is about 17 dB, **b** the CNR of balanced detection is about 22 dB
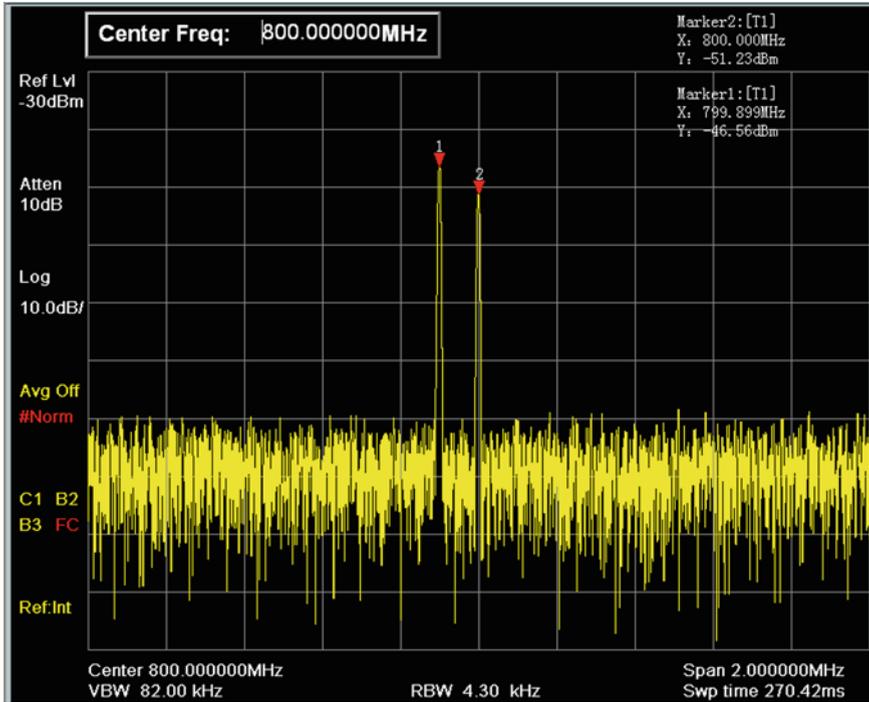
Fig. 7 The spectrum for measuring frequency resolution

6 Conclusion

In this paper, based on acousto-optic effect and balanced coherent detection, a technical method for noise suppression is presented. It has been proven that the approach can greatly improve the ability of the system to detect weak signals. The main reasons for SNR improvement can be concluded as follows: (1) Only signals with stable phase can establish phase grating in acousto-optic devices, so random noise in transmission is suppressed; (2) A balanced coherent detection system can reduce local oscillator excess and shot noise; (3) Injecting laser beams into PM fiber, which reduces the influence from background light and vibration, makes great contributions to SNR and stability of the system. Experiments have proven that the system can achieve carrier- frequency measurement quickly and accurately. Hence a new technical method for practical application is supplied to reduce noise in feeble signals environment.

References

1. Farley, G.J., Galko, P.: Acousto-optic spectrum analyzer techniques for monitoring CW. IEEE Trans. Aerosp. Electron. Syst. **27**(3), 430–440 (1991)
2. Hecht, D.L.: Multifrequency acoustooptic diffraction. IEEE Trans. Sonics Ultrason. **24**(1), 7–18 (1977)
3. Abbas, G.L., Chan, V.W., Yee, T.K.: Local-oscillator excess-noise suppression for homodyne and heterodyne detection. Opt. Lett. **8**(8), 419–421 (1983)
4. Wang, C.H., Gao, L., Pang, Y.J. et al.: Experimental investigation for relation between beam splitter coefficient and signal-to noise ratio of 2 μm balanced coherent system. Acta Optica Sinica **31**(11), 1104002-1–1104002-6 (2011)
5. Liu, H.Y., Zhang, Y.G., Ai, Y. et al.: Design and implementation of balance detector used in coherent optical communication system. Laser Optoelectron. Process. **51**, 070601-1–070601-7 (2014)
6. Xu, S.L., Hu, Y.h., Guo, L.R.: Design and performance analysis of aircraft wake vortex coherent laser detection system. Laser Optoelectron. Process. **51**, 081202-1–081202-6 (2014)
7. Painchaud, Y., Poulin, M., Morin, M., et al.: Performance of balanced detection in a coherent receiver. Opt. Express **17**(5), 3659–3672 (2009)
8. Barry, J.R., Lee, E.A.: Performance of coherent optical receivers. IEEE **78**(8), 1369–1394 (1990)
9. Gong, Y.X., He, N., Zhang, L.: Research into the approaches of coherent detection with optical fiber delay line based on phase-controlling. Optical Commun. Technol. **29**(7), 14–16 (2014)
10. He, N., Feng, T.Q., Liao, X.: Research on noise cancellation based on acouto-optic effect in coherent detection. Acta Optica Sinica **35**(7), 0706001-1–0706001-8 (2015)

Performance Evaluation of a Two-Dimensional Savitzky-Golay Filter for Image Smoothing Applications

Jeong-Hwan Kim, Gyeo-Wun Jeung, Jun-Woo Lee and Kyeong-Seop Kim

Abstract The smoothing ability of the one-dimensional Savitzky-Golay (SG) filter has typically been applied to spectroscopic measurement and biomedical signal processing for suppressing noise in electrocardiogram or electroencephalogram data. However, interpretation of an image using the SG filter was rarely considered. In this study, a formulation of a two-dimensional (2D) SG filter was proposed specifically for image smoothing applications. The performance of image smoothing using the 2D SG filter under the corruption of additive Gaussian random noise was evaluated in terms of signal to noise ratio and mean square error.

Keywords Savitzky-Golay filter · Noise · Smoothing · Gaussian filter · Mean filter

1 Introduction

The Savitzky-Golay (SG) filter is a finite impulse response (FIR) filter that can be applied to a set of data points in a signal for the purpose of reducing noise while simultaneously preserving high-frequency components. With the assumption of equally time-spaced samples, the polynomial coefficients of a one-dimensional SG filter were estimated optimally by minimizing the least-square error between the polynomial curve and the given data [1–3]. The smoothing property of the one-dimensional SG filter was originally applied to measure accurate spectra.

J.-H. Kim · G.-W. Jeung · J.-W. Lee · K.-S. Kim (✉)
School of Biomedical Engineering, Konkuk University, Chungju, Korea
e-mail: kyeong@kku.ac.kr

J.-H. Kim
e-mail: zindevil@kku.ac.kr

G.-W. Jeung
e-mail: wjdrudns92@naver.com

J.-W. Lee
e-mail: junu418@naver.com

The SG smoothing filter has also applied to suppress noise in electrocardiogram [4, 5] and to enhance geophysical signals or remote sensing data [6–8]. However, formulations of two-dimensional (2D) SG filters were not well established in terms of image processing applications. With this goal, a derivation of the 2D SG smoothing filter was presented and its image smoothing capability under the corruption of additive Gaussian random noise was compared in terms of signal to noise ratio (SNR) and mean square error (MSE) to the mean and Gaussian smoothing filter results.

2 Derivation of 2D SG Filter Kernel

Consider a sub-image, $g(i, j)$, $(i = 1, …, M_l, j = 1, …, N_l)$ of an $M \times N$ image that represents two-dimensional digital data that are time-evenly sampled. To fit a two-dimensional polynomial to $g(i, j)$ in the sense of the least square minimization, the nth -degree two-dimensional polynomial $p(x, y)$ can be expressed as:

$$p(x, y) = \sum_{r_1=0}^{n_1} \sum_{r_2=0}^{n_2} a_{r_1 r_2} \cdot x^{r_1} \cdot y^{r_2}, \quad n_1, n_2 = 0, 1, \ldots n \text{ and } 0 \le (n_1 + n_2) \le n \quad (1)$$

where $a_{r_1 r_2}$ is the coefficient of two-dimensional polynomial with degree of n and (x_i, y_i) is a spatial coordinate in a sub-image that allocates the center pixel at the $(0, 0)$ position. The kernels of the two-dimensional SG filter can be determined by choosing the $a_{r_1 r_2}$ coefficient that minimizes the total mean-square error, ε_n:

$$\varepsilon_n = \sum_{i}^{M_\ell} \sum_{j}^{N_\ell} \{g(i,j) - p(i,j)\}^2 \quad (2)$$

In this study, the value of M_l was selected to be equal to N_l, i.e., $M_l = N_l = L$ by considering a square image-patch. In a vector format, the data in the considered region can be expressed by G, as in the following:

$$\begin{aligned}
G = \Bigg[& g\left(\left[-\frac{L}{2}\right], \left[-\frac{L}{2}\right]\right) g\left(\left[-\frac{L}{2}\right], \left[-\frac{L}{2}\right]+1\right) \cdots g\left(\left[-\frac{L}{2}\right], 0\right) g\left(\left[-\frac{L}{2}\right], 1\right) \\
& \cdots g\left(\left[-\frac{L}{2}\right], \left[\frac{L}{2}\right]\right) g\left(\left[-\frac{L}{2}\right]+1, \left[-\frac{L}{2}\right]\right) g\left(\left[-\frac{L}{2}\right]+1, \left[-\frac{L}{2}\right]+1\right) \\
& \cdots g\left(\left[-\frac{L}{2}\right]+1, 0\right) \quad g\left(\left[-\frac{L}{2}\right]+1, 1\right) \cdots g\left(\left[-\frac{L}{2}\right]+1, \left[\frac{L}{2}\right]\right) \\
& \cdots g\left(\left[-\frac{L}{2}\right], [L-1]\right) \cdots g\left(\left[-\frac{L}{2}\right]+(L-1), \left[\frac{L}{2}\right]\right) \Bigg]^t
\end{aligned} \quad (3)$$

where t is a transpose operator and $[\cdot]$ is an operator that retains the largest integer value less than or equal to the original vale. Similarly, L^2 dimensional-polynomial basis vectors $S_0, S_1, S_2,\dots S_{k-1}$, in which k is the total number of coefficients, can be defined as the following, whose components are:

$$
S_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}, S_1 = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_i \\ \vdots \\ x_{(L^2-1)} \end{bmatrix}, S_2 = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_i \\ \vdots \\ y_{(L^2-1)} \end{bmatrix}, \dots,
$$

$$
S_\ell = \begin{bmatrix} x_0^{n_1} \cdot y_0^{n_2} \\ x_1^{n_1} \cdot y_1^{n_2} \\ \vdots \\ x_i^{n_1} \cdot y_i^{n_2} \\ \vdots \\ x_{(L^2-1)}^{n_1} \cdot y_{(L^2-1)}^{n_2} \end{bmatrix}, \dots, S_p = \begin{bmatrix} x_0^{n} \\ x_1^{n} \\ \vdots \\ x_i^{n} \\ \vdots \\ x_{(L^2-1)}^{n} \end{bmatrix}, \dots, S_{(k-1)} = \begin{bmatrix} y_0^{n} \\ y_1^{n} \\ \vdots \\ y_i^{n} \\ \vdots \\ y_{(L^2-1)}^{n} \end{bmatrix}
$$

$$(4)$$

where $n_1, n_2 = 0,1,\dots n$, $0 \le (n_1 + n_2) \le n$ and n is the degree of two-dimensional polynomial $p(x, y)$. The values of x_i and y_i are determined by (x_i, y_i) spatial coordinate of the corresponding pixel in G. The polynomial fitting coefficients can be estimated by formulating the following equation in a vector-matrix format:

$$
G = S \cdot A
$$

$$
= \begin{bmatrix} \vdots & \vdots & & \vdots \\ S_0 & S_1 & \cdots & S_{k-1} \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} a_{00} \, a_{10} \, a_{01} \, \cdots \, a_{ij} \, \cdots a_{n0} \, \cdots \, a_{0n} \end{bmatrix}^t \qquad (5)
$$

where i, j = 0, 1, … n and $0 \le (i + j) \le n$

If the sub-space is defined as C, which is spanned by the vectors $S_0, S_1, S_2,\dots S_{k-1}$, the optimal vector S^* lies in a *Hilbert* space, C. The closest distance from S^* to G can be resolved by the projection theorem using the fact that $G–S^*$ is orthogonal to S^*, as shown in Fig. 1 [9, 10].

Fig. 1 Illustration of
projection theorem in which
$\|\cdot\|$ is the norm operator
defined in a *Hilbert* space, C
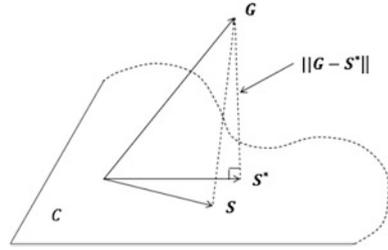
The two-dimensional fitted-polynomial coefficients $\widehat{A}$ can be estimated by
computing a pseudo-inverse, $(S^t \cdot S)^{-1}$ such that the following is true:

$$\widehat{A} = (S^t \cdot S)^{-1} \cdot S^t \cdot G = W \cdot G = \begin{bmatrix} \cdots W_1 \cdots \\ \cdots W_2 \cdots \\ \vdots \\ \cdots W_k \cdots \end{bmatrix} \cdot G \qquad (6)$$

where W is the weight matrix which comprises the row vectors, $W_1, W_2, \ldots, W_k \in R^{L^2}$ and k is the number of SG-polynomial coefficients. Therefore, the smoothed
image $\widehat{G}$ using the 2D SG filter can be obtained by the following:

$$\widehat{G} = S \cdot \widehat{A} = S(S^t \cdot S)^{-1} \cdot S^t \cdot G = P \cdot G \qquad (7)$$

where P is a projection matrix. Note that the elements in a G vector denotes the
intensities of the pixels in a sub-image, $g(i, j)$, $(i, j = 1, \ldots, L)$ and the SG filtered
intensity of the target pixel at $(0, 0)$ spatial coordinate is obtained by evaluating
Eq. (7). In order to get the new SG filtered data, the pixels in the sub-image need to
be shifted one sample from the left to right and from the top to the bottom direction
(if necessary) to define the new position of a center pixel. Therefore, SG smoothing
filtering can be achieved by applying the two-dimensional convolution operation by
resolving the two-dimensional convolution kernels by interpreting the nth-order
fitting polynomial in Eq. (1) such that the following is true:

$$p(0,0) = a_{00} \qquad (8)$$

Thus, the two-dimensional SG smoothing filter-kernel can be identified by a_{00},
which is related to the first row vector in a matrix W in Eq. (6).

3 Performance Evaluation of Image Smoothing with 2D SG Filter

3.1 Image Smoothing with 2D SG Filter (L = 3, n = 2)

Figure 2 shows the 3 × 3 convolution kernels in a mesh plot by taking account of masks defined by the 2D SG mean and Gaussian smoothing filter. These kernels were convolved with a noisy image *g* to suppress the noise. For the experimental simulations, the "Lenna" image was corrupted with the additive Gaussian-distribution random noise (mean = 0.0041, variance = 164.8225). Figure 3 illustrates the filtered images using the various smoothing filters.

Table 1 summarizes the performance indexes in terms of SNR and MSE. Here, the SNR was calculated in dB scale by computing the relative ratio of the variance of the filtered image to the variance of a local region that contained only noisy statistics.
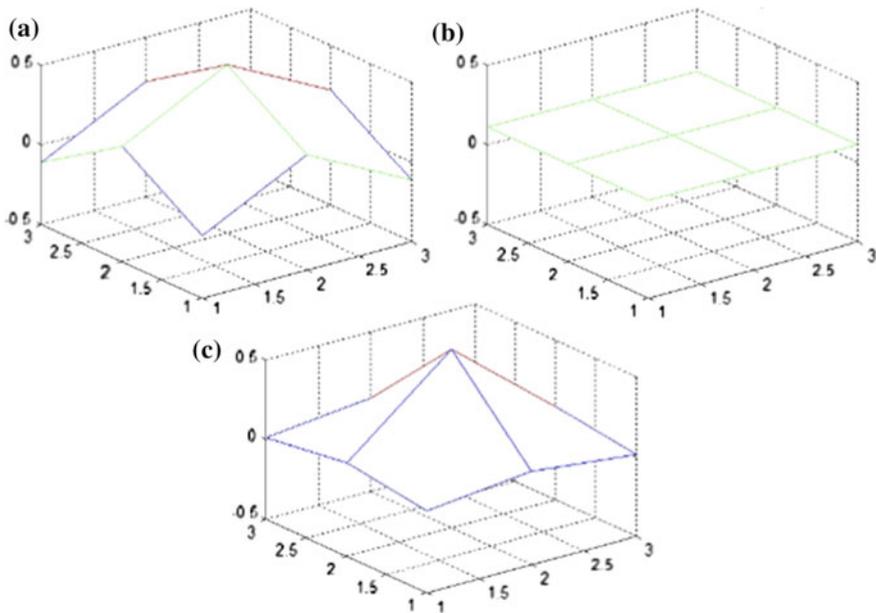
Fig. 2 Mesh plot of the 3 × 3 convolution kernels for image smoothing: **a** 2D SG; **b** Mean; **c** Gaussian filter
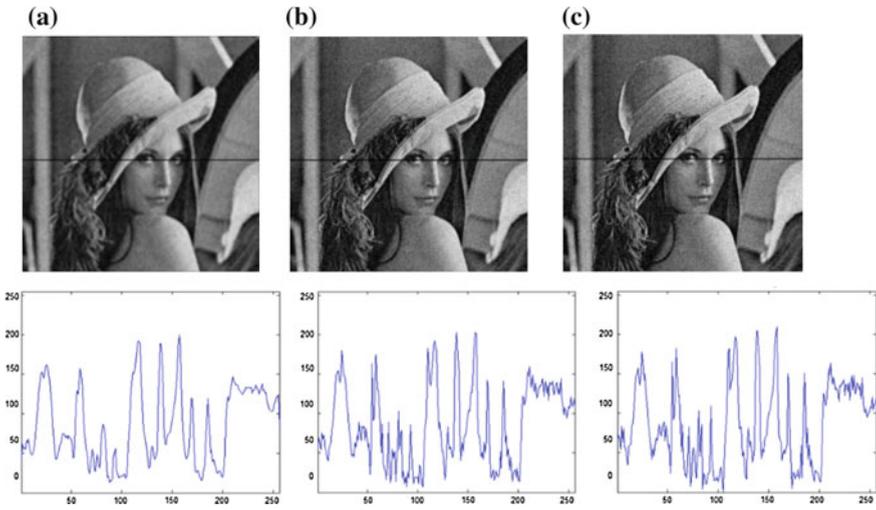
Fig. 3 The smoothing filtered images and their line profile-intensities by **a** Mean; **b** Gaussian; **c** 2D SG filter ($L = 3$, $n = 2$)

Table 1 The performance indexes of SNR and MSE

Filter type	SNR (dB)	MSE (dB)
Noisy "Lenna" image	13.7752	22.0492
Smoothed image by Mean filter	20.9307	21.9886
Smoothed image by Gaussian filter	16.5015	19.4405
Smoothed image by 2D SG filter ($L = 3$, $n = 2$)	16.4047	20.8695

3.2 Image Smoothing by 2D SG Filter (L = 5, n = 3, 4)

Figure 4 shows the filtered images after applying the smoothing filters.

Table 2 shows the performance indexes in terms of SNR and MSE.
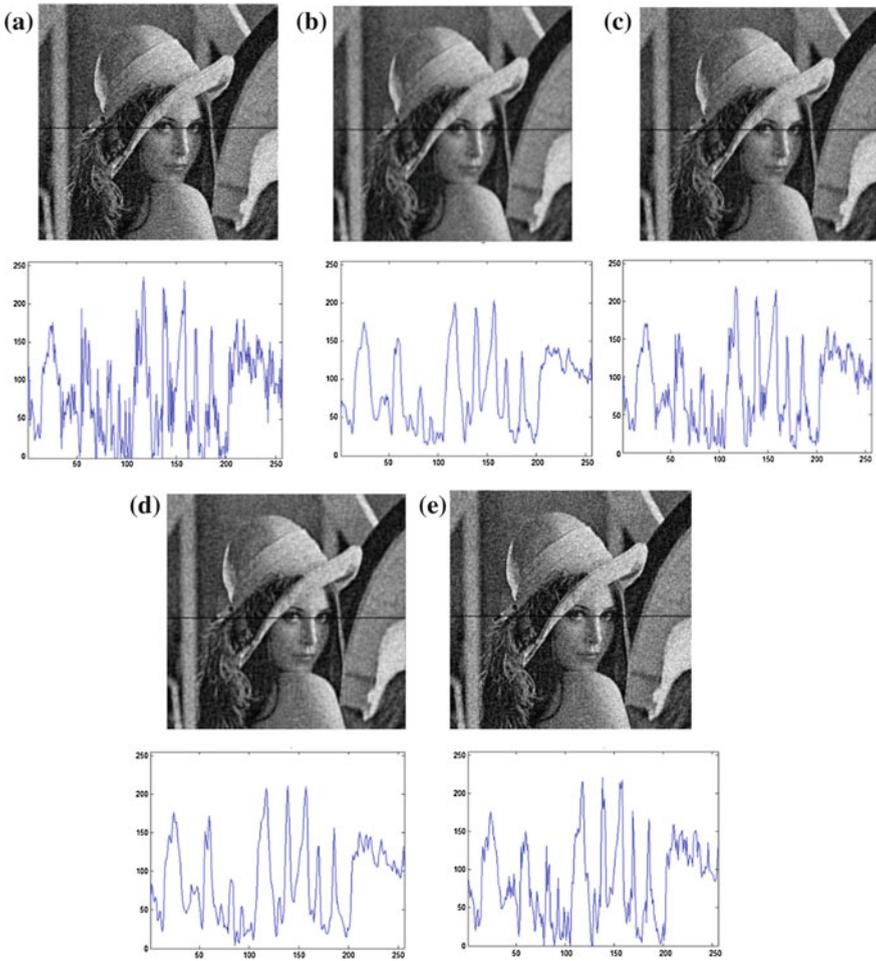
Fig. 4 The smoothing filtered images and their line profile-intensities: **a** the noisy image, g; **b** Mean; **c** Gaussian; **d** 2D SG filter ($L = 5$, $n = 3$); **e** 2D SG filter ($L = 5$, $n = 4$)

Table 2 The performance indexes of SNR and MSE

Filter type	SNR (dB)	MSE (dB)
Noisy "Lenna" image	7.8373	27.8890
Smoothed image by Mean filter	15.2093	23.2550
Smoothed image by Gaussian filter	11.0771	24.4367
Smoothed image by 2D SG filter ($L = 5$, $n = 3$)	13.8736	23.4288
Smoothed image by 2D SG filter ($L = 5$, $n = 4$)	11.6916	25.0338

4   Conclusion

In this study, a two-dimensional SG smoothing filter was formulated. Simulation results showed that the two-dimensional SG smoothing filter reduced noise without loss of resolution by identifying high-frequency contents in the image data. As summarized in Tables 1 and 2, the two-dimensional SG filter maintained the intermediate SNR and MSE levels for the high-SNR image. Additionally, these performance indexes were improved upon in low-SNR images.

References

1. Candan, C., Inan, H.: A unified framework for derivation and implementation of Savitzky-Golay filters. Sig. Process. **104**, 203–211 (2014)
2. Xiong, G., Ding, T.H.: A filtering paradigm for signal's noise removal and feature preservation. Sig. Process. **93**, 1172–1191 (2013)
3. Krishnan, S.R., Seelamantula, C.S.: On the selection of optimum Savitzky-Golay filter. IEEE Trans. Signal Process. **61**(20), 380–391 (2013)
4. Rastogi, N., Mehra, R.: Analysis of Savitzky-Golay filter for baseline wander cancellation in ECG using wavelets. Int. J. Eng. Sci. Emerg. Technol. **6**, 15–23 (2013)
5. Birle, A., Malviya, S., Mittal, D.: Noise removal in ECG signal using Savtizky-Golay filter. Int. J. Adv. Res. Electron. Commun. Eng. **4**(5), 1331–1333 (2015)
6. Baba, K., Bahi, L., Quadif, L.: Enhancing geophysical signals through the use of Savitzky-Golay filtering method. Geo. Int. **53**(4), 399–409 (2014)
7. Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., Wang, X.: Land cover classification of landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sens. **6**, 11518–11532 (2014)
8. Patakamuri, S.K., Agrawal, S., Krishnaveni, M.: Time-series analysis of MODIS NDVI data along with ancillary data for land use/land cover mapping of Uttarakhand. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XL-8, ISPRS Technical Commission VIII Symposium, pp. 1491–1500, Hyderabad, India. 09–12 Dec 2014
9. Blanchet, G., Charbit, M.: Digital Signal and Image Processing using Matlab. ISTE Ltd., London (2006)
10. Vettereli, M., Kovacevic, J., Goyal, V.K.: Foundations of digital signal processing. Cambridge University Press, Cambridge (2014)

Research on a Credit Reservation Mechanism for Concurrent Multi-services

**Jun-Qiang Guo, Ren-Yuan Zhang, Guo-Zhao Wei
and Jin-Liang Yang**

Abstract A new credit reservation mechanism is proposed to solve the problem of online charging shared balance allocation of concurrent multi-services in a short time. In the new mechanism, the moving average based slide window algorithm is adopted to allocate the available credit for each service, and the credit allocation granularity is calculated based on the distribution of concurrent multi-service arrival, the validity time of granted credit units and the number of served sessions. An analytical model is developed, in which an appropriate reservation credit unit or the number of concurrent multi-services could be selected for various traffic conditions. The credit reservation mechanism could reduce the amount of signaling and the average interactive times between network elements and online charging systems.

Keywords Online charging · Concurrent · Multi-service · Reservation

1 Introduction

Charging and billing are important activities for operators and users in the telecommunication system. In order to construct a better charging support system in the 3G environment, and to achieve the convergence of prepaid and postpaid, the reference framework of online charging system (OCS) is proposed by 3GPP

J.-Q. Guo · R.-Y. Zhang (✉) · G.-Z. Wei · J.-L. Yang
Shandong University of Science and Technology, Qingdao, Shandong, China
e-mail: sduzry@163.com

J.-Q. Guo
e-mail: guojunqiang2008@163.com

G.-Z. Wei
e-mail: 543783719@163.com

J.-L. Yang
e-mail: yangjinliang@163.com

organization. The OCS is responsible for providing more personalized advice to users about charging and credit limit control. It also allows prepaid and postpaid sessions for the user simultaneously in real time. A major requirement imposed on online charging prepaid system is that the prepaid system should be able to cope with a concurrent multi-service environment [1]. Delivering multiple sessions simultaneously and credit reservation for multiple sessions are both important features for the OCS.

The allocation of account balance not only directly affects the expected number of interactions between network elements and OCS, but also affects the concurrent pressure of OCS and the utilization of the users' account. In [2], Prepaid Credit Distribution (PCD) mechanism and Prepaid Credit Reclaim (PCR) mechanism are proposed. The PCD algorithm does not take into account the distribution of session duration. The PCR mechanism allows all services to be equally allocated with the credit, instead of rejecting new services due to insufficient credit left. However, it does not consider the service usage of each ongoing session, and it generates an additional amount of signaling. In [3], the Recharge Threshold-based Credit Reservation (RTCR) mechanism is proposed. An analytical model is used to calculate the optimal credit threshold. However, it does not take into account the issues related to the shared balance allocation of a concurrent multi-service.

The OCS allows the shared account balance allocation of a concurrent multi-service. Due to the limited account balance, if the reserved credit units are too large for the concurrent multi-service, the unused reserved credit units will not be released until the end of services. The new credit request may be rejected because of insufficient account balance, and the utilization of the account is not high. If the credit units are reserved too small, the reserved credit units will be consumed in a short time. Then the network elements will send credit units request to OCS frequently. That means a considerable increase of the signaling messages that are generated in the network, or the network will have high overhead traffic. Therefore, using reasonable shared balance allocation granularity to reserve an appropriate amount of credit units for concurrent multi-services must be considered for operators. A concurrent multi-service credit reservation mechanism is proposed in this paper. Instead of performing credit reservation of the services independently, it will be done on a concurrent multi-service basis. In the multi-service environment, the credit allocation granularity is calculated based on the distribution of concurrent multi-session arrival, the validity time of granted credit units and the number of served sessions.

The rest of this paper is organized as follows. Section 2 provides a brief introduction to the online charging architecture and the credit control message flow based on the online charging service session. Section 3 analyses credit reservation for concurrent multi-service sessions in detail, and comes to the conclusion that appropriate reserved credit units can be selected for various traffic condition. Section 3 also develops an analytical model for the credit reservation procedure. Conclusion outlined in Sect. 4, which is followed by the reference.

2 Online Charging Architecture and the Credit Allocation Procedure

Figure 1 illustrates how Universal Mobile Telecommunications System (UMTS) integrates with the OCS [3, 4]. When User Equipment (UE) originates a prepaid session through the General Packet Radio Service (GPRS), the Gateway GPRS Support Node (GGSN) requests the prepaid credit from the OCS through the Gy interface. The OCS consists of four functional elements, and is responsible for charging the received events. The OCF consists of two modules: Session Based Charging Function (SBCF) and Event Based Charging Function (EBCF). The SBCF module is responsible for online charging of user sessions. The SBCF interacts with the Account Balance Management Function (ABMF) for credit control. The ABMF is responsible for keeping user's balance and other account data. The SBCF also interacts with the RF to determine the tariff of the requested session [4]. The RF contains prices or tariffs of all variable services and maintains the end user context information. The OCF makes full use of the received information to generate
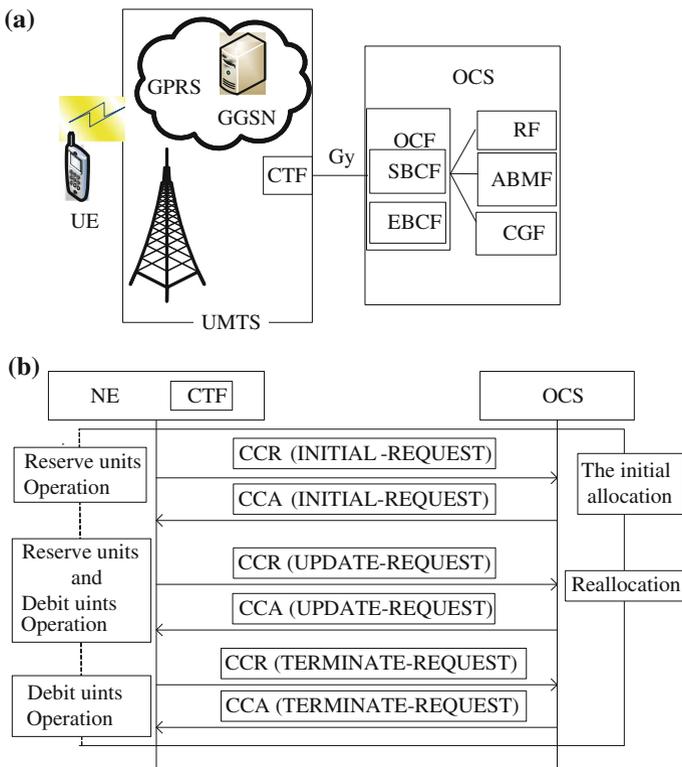

Fig. 1 The OCS architecture and diameter credit control message flow

charging data records (CDRs), which are sent to the charging gateway function (CGF) [5–7]. The Charging Trigger Function (CTF) is integrated in the network element. It also monitors the charging service events, and tracks and controls the use of credit units. OCS credit control is achieved by exchanging Credit Control Request (CCR) and Credit Control Answer (CCA) messages.

There are three types of credit control message: INITIAL_REQUEST, UPDATE_ REQUEST, and TERMINATION_REQUEST.

We note that each packet transmitted through the 3GPP gateway triggers an online charging reservation procedure with the OCS [3, 4]. The credit reservation procedure for a session-based online charging includes three types of credit control operations: Reserve Units operation, Reserve Units and Debit Units operation. The basic operation of prepaid charging is shown in Fig. 1b. In concurrent multi-service online charging, CDRs are generated per service to count the chargeable events per network element in real-time. If some specific events triggered by charging occur, a very large quantity of network elements generate a large number of CDRs that are handled by 3GPP gateways and the OCS. If event charging with credit reservation is applied in these situations, traffic signaling will be generated for the network in a short time [8, 9]. To reduce signaling cost, a new concurrent multi-service credit reservation mechanism is proposed.

3   The Analysis of the Concurrent Multi-service Credit Reservation

In order to facilitate analysis, let n denotes the total number of sessions that are created by concurrent multi-service. That is, the concurrent multi-service consists of n sessions in parallel $s_1, s_2,\ldots, s_n$ with a prepaid charging option. Instead of performing credit reservation of the services independently, the new method will be done on a group of concurrent multi-service sessions. Therefore, each group is associated with a time interval T_V, which is also the parameter to be monitored per group. Each session process is actually to be completed by delivering a group of transmission packets. Let c_i be the credit units charged for delivering s_i packets of transmission. The value of c_i is based on the user's historical service usages. We can adopt the moving average based slide window algorithm to calculate c_i. The recent service usage could better reflect the users' habits in general. Thus the recent credit usage for s_i in the average calculation should occupy a higher weight. The sliding window technique is used to optimize the average of the historical credit usages. We simply consider the several most recent credit usages. The credit usages of s_i are placed in chronological order $u_0, u_1, u_2,\ldots, u_{i-1}$ (i is the usage count). The ith value of credit usage $E[\bar{u}]$ is the corresponding credit unit c_i. Suppose that θ credit units are consumed for providing n sessions packets delivery transmission. The OCS reserves $\theta = \sum_{i=1}^{n} c_i$ credit units with the validity time T_v. In order to reduce the signaling cost and the average interactive times, the reserve units

operation is achieved in one reservation. Suppose that each session packet arrives at $t_1, t_2, \ldots, t_i, t_{i+1}$. Let τ_i denote the inter-arrival time between the s_i and the s_{i+1} session packets. The OCS reserves no credit units before the s_i packets arrive. The Reserve Units Operation is executed at t_1, and the Debit Units Operation is executed at t_d. T_R denotes the concurrent multi-service credit reservation procedure holding time. We consider each group of packets' arrival to follow a Poisson distribution with rate λ. Also, the inter-arrival time τ_i follows an exponential distribution with mean $1/\lambda$. Let $t_e = \sum_{i=1}^{m} \tau_i$ denote the elapsed validity time between s_1 and s_{m+1}. According to the Poisson process property [10], the t_e follows an Erlang distribution.

When the s_1 arrives at t_1, the OCS reserves θ credit units and sets the validity time T_V to support at most n group of packets delivery transmissions. The reserved credit units may be depleted within the validity time, and in this case, $T_R = t_e$ can be derived. In another case, the validity timer expires before the reserved credit units are consumed, so $T_R = T_V$ can be derived. According to the Little's Theorem [10], the expected number of the sessions and the expected number of the unused credit units is expressed as:

$$E[N] = 1 + \lambda E[T_R] = 1 + (n-1)\left[1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda T_V}(\lambda T_V)^k}{k!}\right] + \lambda T_V\left[\sum_{k=0}^{n-2} \frac{e^{-\lambda T_V}(\lambda T_V)^k}{k!}\right]$$
(1)

$$E[C_{un}] = \theta - E[N]E\left[\sum_{i=1}^{n} c_i\right]$$
$$= \theta - E\left[\sum_{i=1}^{n} c_i\right]\left\{\left[2 - n - \sum_{k=0}^{n-1} \frac{e^{-\lambda T_V}(\lambda T_V)^k}{k!}\right] - \left[\sum_{k=0}^{n-2} \frac{e^{-\lambda T_V}T_V(\lambda T_V)^k}{k!}\right]\right\}$$
(2)

Equations 1 and 2 validate against the analytical model as illustrated in Figs. 2, 3 and 4. In order to facilitate the calculation, let $\sum_{i=1}^{n} c_i = n$. In these figures, the $E[N]$ and the $E[C_{un}]$ represent the expected numbers of the sessions served and the expected number of the unused credit units in one reservation, respectively. Let $P_C = (E[N] - 1)/E[N]$ represent the average interactive time reduction.

Figure 2 plots E[N], $E[C_{un}]$ and P_C against the session packets arrival rate λ, where the validity time $T_V = 50$ time units. The E[N] and P_C are increasing function of λ. The maximum number of E[N] is close to n. When $\lambda > 0.5$, the average interactive times reduction is larger than 0.9. Figure 3 plots E[N], $E[C_{un}]$ and P_C against the validity time T_V, where reserved credit units $\theta = 50$. The E[N] and P_C are increasing functions of T_V. On the contrary, the $E[C_{un}]$ is a decreasing function of λ in Figs. 2 and 3. Fewer credit units are left and are returned to the OCS at the end of the new procedure. Figure 4 plots the E[N], $E[C_{un}]$ and P_C against reserved credit units θ, where validity time $T_V = 50$ time units. All of them (E[N], P_C and $E[C_{un}]$) are increasing functions of θ. When $10 \leq \theta \leq 20$, and $\lambda > 0.5$, $E[C_{un}]$ is close to
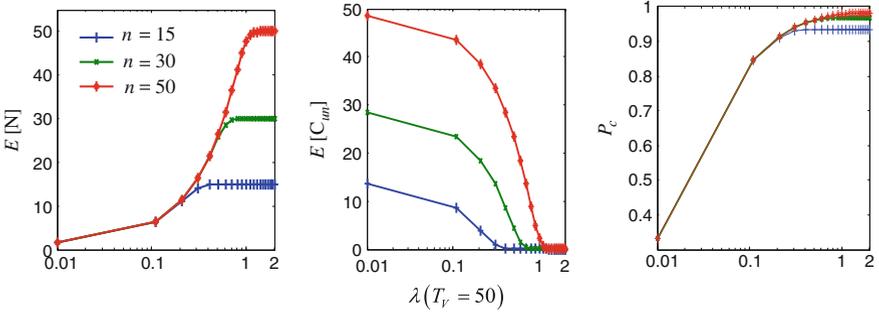
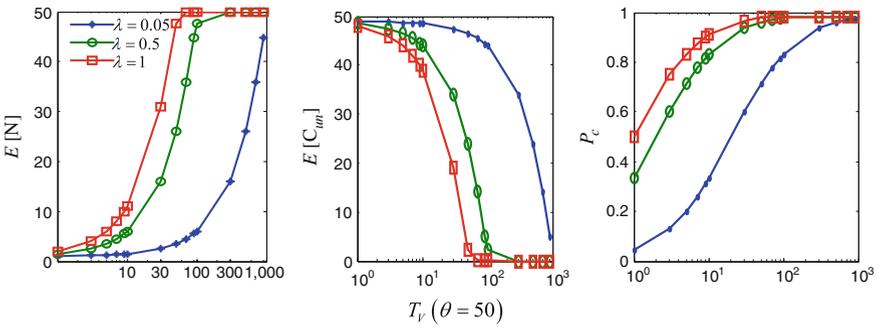**Fig. 2** Effects of λ and n
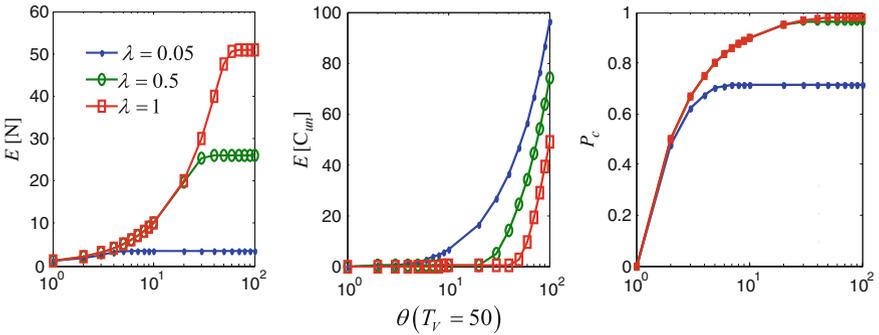

Fig. 3 Effects of λ and T_V

Fig. 4 Effects of λ and θ

zero, and E[N] and P_C are relatively high. When θ increases to a certain extent, the E[N] and P_C will converge to a maximum value. That is to say we need not reserve too many credit units to execute a concurrent multi-service.

4   Conclusion

Online charging system (OCS) allows the shared account balance allocation of concurrent multi-services. Credit reservation for the concurrent multi-service is a very important factor for optimizing signaling cost and reducing the average interactive time. A new credit reservation for concurrent multi-service procedure is proposed to reduce the signaling cost and the average interactive time. A moving average based slide window algorithm is proposed based on the user's historical service usages, which can be used in the assignment of the customer's available service usage.

An analytical model is developed to compute the E[N], $E[C_{un}]$ and P_C for further analysis of the relationship between the input parameters and the output metrics. Some conclusions to be drawn are that the average interactive times reduction is increased when the reserved credit units increases, but the unused credit units that will be returned to OCS are also increasing. The unused credit units decrease only when the reserved credit units decrease. According to the above discussion, the appropriate reserved credit unit or the number of concurrent multi-services could be selected by mobile operators for various traffic conditions.

References

1. Kurtansky, P., Stiller, B.: Time Interval-Based Prepaid Charging of QoS-enabled IP Services. Internet and Network Economics, pp. 325–335. Springer, Heidelberg (2005)
2. Lin, Y.B., Gan, C.H., Jeng, J.Y.: Credit allocation for UMTS prepaid service. In: IEEE Transactions on Vehicular Technology, vol. 55(1), pp. 306–316. IEEE Vehicular Technology Society, Piscataway (2006)
3. Sou, S.I., Hung, H.N., Lin, Y.B., Peng, N.F.: Modeling credit reservation procedure for UMTS online charging system. In: IEEE Transactions on Wireless Communications, vol. 6, pp. 4129–4135. IEEE Communications Society, New York (2007)
4. Li, X.Y., Kim, B., Cai, Y.G.: Dynamic policy and charging control framework. Bell Labs Tech. J. **17**(1), 105–123 (2012)
5. Li, H.Y., Gao, Y., Zhao, L., Li, Z.W.: Adaptive control of charging system based on online parameters identification. In: International Conference on Telecommunications Energy Conference (INTELEC), pp. 1–6. IEEE, Vancouver (2014)
6. Huang, H.S., Su, T.C., Liang, J.M., Tseng, Y.C.: A dynamic reservation scheme in online charging system for family shared plan. In: International Conference on Wireless Communications and Networking (WCNC), pp. 2279–2284. IEEE, New Orleans (2015)
7. Globa, L., Skulysh, M., Zastavenko, A.: The method of resources allocation for processing requests in online charging system. In: International Conference on Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), pp. 211–213. IEEE, Lviv (2015)

8. Sou, S.I., Tsai, D.R.: Bulk credit reservation in event-based machine type communications for 3GPP online charging. Comput. Netw. **75**(1), 75–85 (2014)
9. Larisa, G., Mariia, S., Andriy, R.: Control strategy of the input stream on the online charging system in peak load moments. In: International Conference on Microwave and Telecommunication Technology (CriMiCo), pp. 312–312. IEEE, Sevastopol (2014)
10. Kleinrock, L.: Queueing systems, Vol. I: theory. In: Theory Proceedings of the IEEE, vol. 1, pp. 990–991. IEEE Journals & Magazines, New York (2005)

Research on Data Recording and Playback Systems for a Virtual Testing Framework

Qing-Jun Qu, Yi-Ping Yao, Huai Wang and Lai-Bin Yan

Abstract The interfaces and data structure of virtual testing applications are dynamically generated according to VFL (Virtual Testing Framework Language) files, which result in changes to the source codes of data recording and playback systems. Thus, the structure and key technologies of existing recording and playback systems are not suitable for the recording and playback systems of the virtual test frame. This paper discusses the architecture of data recording and playback systems, whose features are dynamically generated and centralized according to recording policy. The architecture consists of data recording and playback tools, and their builder. The recorded replay data format and key algorithms in the architecture were investigated. Data recording and playback systems were then implemented; test results indicate that the generated data recording and playback tool can correctly record and playback data for different virtual test applications.

Keywords Virtual testing framework · Dynamically generated · Recording · Playback

Q.-J. Qu (✉) · Y.-P. Yao · H. Wang · L.-B. Yan
College of Information System and Management, National University
of Defense Technology, Changsha, Hunan, China
e-mail: qjqu2012@163.com

Y.-P. Yao
e-mail: ypyao@nudt.edu.cn

H. Wang
e-mail: huai_wang@163.com

L.-B. Yan
e-mail: lbyan2010@163.com

1   Introduction

A Virtual Testing Framework (VTF) is a typical type of general supporting frame, similar to TENA (Test and Training Enabling Architecture). VTF provides integrated facilities for design prior to testing, management during testing and analysis after testing, and can meet requirements of interoperability and compositionality among logical range, public infrastructure and model resources in virtual tests [1–3]. It is driven by the demand for system-level virtual test verification technology, intended to form system-level virtual testing and verification platforms for equipment systems. Constructed according to the concept of virtual-real integration and hierarchical structure, VTF can enhance the synchronization ability between industrial product design and virtual testing verification, and provide technical support for the standardization of a testing verification process.

Data recording and playback systems provide important support to a Virtual Testing Framework to achieve the design prior to testing, management during testing and analysis after testing. VTF uses VFL (Virtual Testing Framework Language) files to describe model attribute data and model operation. In VTF simulation applications, data interfaces are dynamically generated based on VFL files. Currently, there is no reference associated with data recording and playback systems for VTF at home or abroad.

One purpose of data recording and playback systems is to collect, monitor, store and play back test process data generated on VTF middleware; another purpose is to provide data support for the analysis of test data, correct verification of simulation models and visualization of testing processes. Different testing systems have different data recording and playback strategies. Therefore, it is important to study data recording and playback systems for VTF to meet the recording and playback needs of VTF test processes. This research is based on actual demands from a research institution.

2   Implementation Strategies of Data Recording and Playback Systems

In the construction process of Chinese simulation application systems, HLA-based distributed-simulation system construction has played a significant role. Therefore, data recording and playback technology research for HLA (High Level Architecture) simulation application systems is more mature [4–6]. Many units have developed commercial products, such as MAK's Data Logger, Pitch's Data Recorder, KD-DCT and StarLogger of the National University of Defense Technology, etc. These systems have different data recording strategies and similar data playback strategies. In summation, there are two types of recording strategies among these products: one is a centralized strategy, and the other is a distributed

strategy. Moreover, in a large application-specific simulation system, a direct strategy is typically used.

2.1 The Centralized Strategy

The centralized strategy is implemented by a common data recording program, which subscribes desired object class attributes and interactive classes through the description of FOM (Federation Object Model). With the execution of simulation applications, after RTI (Run-Time Infrastructure) pushes subscribed data to the data recording program, the program writes the received data onto the storage medium (the record file or the database). The advantages of the centralized strategy are its relatively simple achievement, independent development of programs, and good versatility for applications which adhere to HLA specifications. Because one recording program must subscribe to large amounts of data produced by the entire simulation applications, there is a risk of performance loss [7].

2.2 The Distributed Strategy

The design of the distributed strategy is implemented by adding a local record interface between the program and RTI, which utilizes the recording system to record one simulation program data at each program location [2]. This strategy can effectively reduce data traffic on the RTI. However, to implement the distributed strategy, each simulation program's recorder should coordinate interfaces with RTI. Furthermore, all distributed recording files must undergo sequential rearrangement, process redundant data and be integrated into a global data recording file before data analysis and playback.

2.3 The Direct Strategy

The direct strategy requires that simulation applications or objects be responsible for saving local data directly with an appropriate file format, e.g., one equipment demonstration system uses Logging Service, which provides a service of choosing some or all object attribute values and saving these values as ASCII or binary files [8]. Using a direct strategy will result in easy implementation and no redundant data. However, by reason of a very high degree of coupling with a special simulation system, the strategy demonstrates low generality.

3 Research on Data Recording and Playback System for VTF

The purpose of a data recording and playback system for VTF is to collect, monitor, store and play back test data. Its main function demands are as follows: (I) provide setting interface of visual data recording to set related parameters, (II) record specified virtual experimental data according to the setting, (III) provide a data saving function and data export function, (IV) play back recorded files and real-time display of data.

3.1 Architecture

Because data interfaces are dynamically generated based on VFL files in VTF simulation applications, data recording and playback system interfaces which subscribe and publish data to VTF Middleware must also be generated dynamically [9, 10]. Moreover, due to its simplicity, independence, and the versatility of a centralized strategy, to avoid multiple developments and to enhance generality, data recording and playback systems for VTF follow the centralized strategy and utilize a dynamic generation method to design modules and functions. The architecture is shown in Fig. 1.
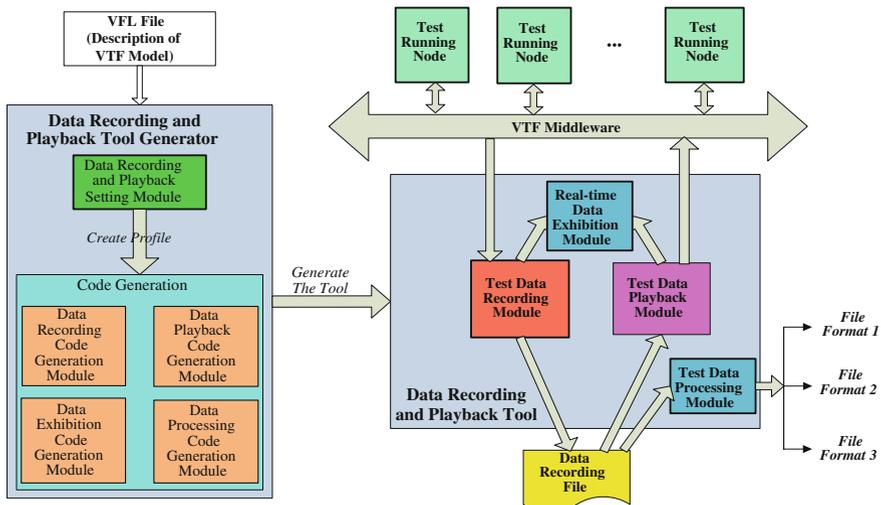

Fig. 1 Architecture of data recording and playback system for VTF

3.2 Makeup

The data recording and playback system is comprised of the data recording and playback tool generator and the data recording and playback tool. The data recording and playback tool is generated automatically according to VTF applications and user settings.

The data recording and playback tool generator is composed of five modules: the data recording and playback setting module (including the user interface), a data recording code generation module, a data playback code generation module, a data exhibition code generation module and a data processing code generation module. The data recording and playback setting module provides a visual means for the user to set the VTF application data to be recorded. Based on the user configuration, each code generation module generates a corresponding code for the data recording and playback tool.

The data recording and playback tool is composed of four modules: a test data recording module, test data playback module, real-time data exhibition module and test data processing module.

4 Key Technology

After the recording and playback tool generates the code for the data recording and playback tool, in order to record and play back data in a virtual test domain, the generated code must be compiled and linked to build an executable program based on the appropriate recording and playback data structure and the correct recording and playback algorithm.

4.1 Recording and Playback Data Structure

The design purpose of the recording and playback data structure is to conveniently record and play back data. The data structure is divided into seven parts: machine time, class ID, instance ID, logical time, message type, message length and message content. Machine time is used to control the playback speed. Class ID and instance ID are used to identify a unique object instance in the virtual test domain. Logical time is used to implement a logical simulation time control. Message type is used to identify whether the message is a discovery message, update message or delete message. Message length and message content are used to indicate the length and content of a specific message.

4.2   Data Recording Algorithm

When information is finished subscribing to VTF middleware, the data recording and playback tool starts to receive data from the VTF middleware. These data include: discovering object instance data, deleting object instance data, attribute updating data of VIO (Object in VTF), instance message (similar to the interaction of HLA) data, TSO/RO (Time Stamp Order/Receive Order) simulation time (logical time) and the wall time (machine time) of all data.

The data recording algorithm flow is described as follows. Step 1: Receive one object instance message from the VTF middleware. Step 2: Determine the type of received message. Step 3: If the type is an object instance discovering message, form a recording of the discovering instance format. Step 4: If the type is an object instance deleting message, form a recording of the deleting instance format. Step 5: If the type is an updating message of VIO/Message attributes, form a recording of the updating instance format. Step 6: Write the recording data to the record data file in various formats, and simultaneously send the data to the real-time data exhibition module to display. Step 7: Restart the reception of a new message, and return to step 1.

4.3   Data Playback Algorithm

The data playback module must implement the control to the playback process and the playback rate. Its main functions include: playback start, playback stop, playback pause, go to beginning of the recording, go to the end of the recording, reduce the playback rate, increase the playback rate, and display the current playback rate.

The data playback algorithm flow is described as follows. Step 1: Receive the playback command and determine which type of command it is. Step 2: If the command is a start command, calculate when the recording data should be played back based on the total length of playback time and the playback rate set by the user; start playback. Step 3: If the command is a pause command, set the playback state to the pause state, save the playback pointer and pause the playback of the recording file. Step 4: If the command is a continue command, resume the playback pointer, and continue playing back the data file. Step 5: If the command is a rate increase command, reset the playback rate. Step 6: If the command is a rate decrease command, reset the playback rate. Step 7: If the command is to go to the beginning of the recording file, set the playback pointer to the file beginning, and play back data from scratch. Step 8: If the command is to go to the end of the recording file, set the playback pointer to the file end, and stop playback. Step 9: If the command is to end the recording file, set the playback pointer to the file end, and stop playback. Step 10: If the command is a stop command, clear the publishing information, delete object instances, exit the test domain, and stop playback.

5 System Testing

The data recording and playback system is comprised of the data recording and playback tool generator and the data recording and playback tool. The data recording and playback tool generator is based on the Windows XP operating system and was developed in the development environment of Visual Studio 2010 and QT4.7.4 add-in.

After invoking VS2010 builder to compile and link the code generated by the data recording and playback tool generator, the data recording and playback tool generator will automatically generate the data recording and playback tool.

To verify the functions and key technologies of the data recording and playback system, a test environment with 12 interconnected PCs was built. Each PC configuration is as follows: Intel® Core Duo CPU 2.93 GHz, 2.0 GB RAM, 100 M Ethernet. Two machines were used to run VTF middleware and the data recording and playback tool (record status). In the remaining ten machines, each machine started a test program which published and subscribed to test classes and registered one instance. The first five of the ten remaining machines (Group A) updated attributes of object instance, and the other five machines (Group B) updated their own object instance attributes after receiving attribute updates sent by Group A. Experimental results indicate that the data recording and playback tool can properly record data.

In the same test environment, two machines were used to run VTF middleware and the data recording and playback tool (playback status). Five of the remaining ten machines ran test programs to subscribe to instance data which were sent by Group A in record tests, while the other five subscribed data which were sent by Group B in record tests. Experiment results indicate that the data recording and playback tool can correctly play back data.

References

1. Zhao, W., Peng, J.: Research and development on the technology of virtual test and evaluation of military industrial products. Comput. Measur. Control **19**(6), 1257–1277 (2011)
2. Feng, R.M., Wang, G.Y., Huang, K.D.: TENA and its comparison with the HLA. Syst. Eng. Electron. **27**(2), 289–291 (2005)
3. Montaola-Sales, C., Casanovas-Garciam, J., Onggo, B.S., Li, Z.: A user interface for large-scale demographic simulation. In Proceedings of International 2014 Conference on Cloud Computing Technology and Science (CLOUDCOM), pp. 723–726. IEEE press, Washington D.C. (2014)

4. Kim, K., Park, T., Pastrana, J., Marin, M., Cortes, E.A., Rabelo, L.C., et al.: Modeling of complex scenarios using LVC simulation. In: 2014 Winter Simulation Conference, pp. 2931–2941. IEEE press, Piscataway (2014)

5. Mirzoyan, D., Akesson, B., Goossens, K.: Process-variation-aware mapping of best-effort and real-time streaming applications to MPSoCs. Acm Trans. Embed. Comput. Syst. **13**(2), 1–24 (2014)

6. Gervais, C., Chaudron, J., Siron, P., Leconte, R., Saussie, D.: Real-time distributed aircraft simulation through HLA. In: Proceedings of the IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications, pp. 251–254. IEEE press, Washington D.C. (2012)

7. Liu, B.Q., Wang, H.M.: A novel compact simulation interface specification. Sci. China Inf. Sci. **57**(6), 1–18 (2014)

8. Gurunule, D., Nashipudimath, M.: A review: analysis of aspect orientation and model driven engineering for code generation. Procedia Comput. Sci. **45**, 852–861 (2015)

9. Tan, X.L., Yao, Y.P., Yan, L.B., Cai, F.H.: Code generation for VTF gateway. In: The 2013 International Conference on Information System and Engineering Management, pp. 468–473. IEEE press, Washington D.C. (2013)

10. Hu, T.F., Yao, Y.P., Jiang, Z.W., Cai, F.H.: Research on technology of time syn-chronization between heterogeneous systems HLA and VITA. In: 2013 International Conference on Information System and Engineering Management, pp. 478–482. IEEE Press, Piscataway (2013)

Robust Generalized Fault Diagnosis Observer Design for Nonlinear Systems with External Disturbances

Zhen-Duo Fu, Rui Wang, Dian-Xi Jiang and Yu Song

Abstract This study focuses on the robust generalized fault diagnosis observer design for a nonlinear system with external disturbances and an actuator fault. The robust generalized fault diagnosis observer (RGFDO) with the defined performance index, which can estimate system states and fault signal simultaneously, is designed. Finally, simulations are performed on flexible joint robotic systems to illustrate the effectiveness and applicability of the algorithm proposed.

Keywords Nonlinear system · Fault estimation · Robust

1 Introduction

Reliability and security of practical control systems are as crucial as performance defined. However, enhancing the dependability of systems has become a challenging task as a result of increasing complexity of automatic systems at present. Fault diagnosis (FD) and fault tolerant control (FTC) technologies are effective strategies designed with a certain degree of fault tolerant performance by means of the fault estimation information from fault estimation observer to guarantee reliable and desired performance of control systems not only under the fault free case but also the fault case. Fruitful FD and FTC approaches for linear systems have been proposed in both frequency and time domains [1–6]. However, because of the complexity and associated difficulties in nonlinear systems, much less work has

Z.-D. Fu (✉) · R. Wang · D.-X. Jiang · Y. Song
Beijing Institute of Mechanical Equipment, Beijing 100854, China
e-mail: fuzhenduo@126.com

R. Wang
e-mail: wr0183@sohu.com

D.-X. Jiang
e-mail: 13501250977@139.com

Y. Song
e-mail: 240373777@qq.com

been done. In this paper, we focus on the design of RGFDO in nonlinear systems with external disturbances and actuator fault.

Various approaches have been used for the design of FTC systems. Generally speaking, they can be broadly classified into: passive fault tolerant control systems (PFTCS) [7–9] and active fault tolerant control systems (AFTCS) [10–18]. The PFTCS apply unchangeable controllers throughout the systems and have the drawback of being reliable only for the class of faults expected and taken into account in the design. By contrast with the PFTCS, the AFTCS are designed on-line which the methods involve the fault information from the fault diagnosis observer so that the reliability and desired performance of the systems can be maintained [19–22].

2 Preliminaries and Problem Statement

Consider the following nonlinear system with the external disturbances and actuator fault

$$\dot{x}(t) = Ax(t) + Bu(t) + Dd(t) + Ef(t) + g(x(t)) \tag{1}$$

$$y(t) = Cx(t) \tag{2}$$

where, $x(t)$ is the state vector, $u(t)$ is the control input vector, $y(t)$ is the measured output vector, $d(t)$ is the external disturbances which can also represent modeling errors and uncertainties. $f(t)$ is actuator fault to be estimated. $g(u(t))$ is a continuous vector function. A, B, C, D and E are known constant matrices of appropriate dimensions.

Assumption 1 $g(x(t))$ is supposed to satisfy $g(0) = 0$ the following norm condition $\|g(x_1(t)) - g(x_2(t))\| \leq L_g \|x_1(t) - x_2(t)\|$ for any $x_1(t)$ and $x_2(t)$. So as $\|g(x(t))\| \leq L_g \|x(t)\|$, where $L_g > 0$ is the Lipschitz constant.

3 Robust Generalized Fault Diagnosis Observer Design

In this section, we focus on the robust generalized fault diagnosis observer design. The RGFDO which can estimate the states and fault simultaneously is proposed with the performance index defined.

Consider the RGFDO for the system (1)–(2) as follows

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + D\hat{d}(t) + E\hat{f}(t) + g(\hat{x}(t)) + L(y(t) - \hat{y}(t)) \tag{3}$$

$$\hat{y}(t) = C\hat{x}(t) \tag{4}$$

where, $\hat{f}(t)$ is the estimate for fault $f(t)$. Define $r(t) = y(t) - \hat{y}(t)$ and $e_f(t) = f(t) - \hat{f}(t)$ as the residual and fault estimation error respectively. The 1-th derivative of fault estimation is denoted as $\dot{\hat{f}}(t) = R_1\hat{f}(t) - R_2r(t)$, R_2 is the matrix to be designed, R_1 is the given constant. Therefore, the state estimation error and fault estimation error are described by

$$
\begin{aligned}
\dot{e}_x(t) &= \dot{x}(t) - \dot{\hat{x}}(t) \\
&= (A - LC)e_x(t) + Ee_f(t) + De_d(t) + (g(x) - g(\hat{x}))
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
\dot{e}_f(t) &= \dot{f}(t) - \dot{\hat{f}}(t) \\
&= R_2Ce_x(t) + R_1e_f(t) + \dot{f}(t) - R_1f(t)
\end{aligned}
\tag{6}
$$

Furthermore, the RGFDO realization is given in Theorem 1.

Theorem 1 *Consider the RGFDO designed as (3)–(4) with the fault estimation law $\dot{\hat{f}}(t) = R_1\hat{f}(t) - R_2r(t)$. The RGFDO is asymptotically convergence with the defined performance index $\|r(t)\| \leq \sigma_2\|v_2(t)\|$ if there exists symmetric positive definite matrix P_2 and a given constant matrix $(\prod)_{2\times4}$ such that the following constraint condition holds*

$$
\begin{bmatrix}
-[(A - LC)^T P_2 + P_2(A - LC)] & -P_2E & -P_2D + C^T\prod D & C^T \\
-E^TP_2 & -(R_1^T + R_1) - R_2R_2^T & -R_2\prod D & -R_2 \\
-D^TP_2 + D^T\prod^TC & -D^T\prod^TR_2^T & -\sigma_2^2D^TD - D^T\prod^T\prod D & -D^T\prod^T \\
C & -R_2^T & -\prod D & -I
\end{bmatrix} < 0
$$

$$
\begin{bmatrix}
-\sigma_2^2I & \sigma_2^2R_1 & -I \\
\sigma_2^2R_1^T & -\sigma_2^2R_1^TR_1 & R_1^T \\
-I & R_1 & 0
\end{bmatrix} < 0
$$

Proof It can be obtained from Eqs. (5)–(6) that the generalized state-space description is described by

$$
\dot{\bar{e}}_2(t) = \begin{bmatrix} A - LC & I \\ R_2C & R_1 \end{bmatrix}\bar{e}_2(t) + v_2(t)
\tag{7}
$$

with definition: $\bar{e}_2(t) = \begin{bmatrix} e_x(t) \\ e_f(t) \end{bmatrix}$, $v_2(t) = \begin{bmatrix} g(x) - g(\hat{x}) + De_d(t) \\ \dot{f}(t) - R_1f(t) \end{bmatrix}$. And then, the RGFDO performance index is defined as $\|r(t)\| \leq \sigma_2\|v_2(t)\|$, and furthermore it can be rewritten as

$$
\|r(t)\| \leq \sigma_2\|v_2(t)\| \Rightarrow \int_0^t [r^T(\tau)r(\tau) - \sigma_2^2v_2^T(\tau)v_2(\tau) - \dot{V}_2(\tau)]d\tau + V_2(t) \leq 0
\tag{8}
$$

Then, a candidate Lyapunov function is chosen as

$$V_2(t) = e_x^T(t)P_2e_x(t) + e_f^T(t)e_f(t)$$

Since, the time derivative of $V_2(t)$ is

$$
\begin{aligned}
\dot{V}_2(t) =\ & \dot{e}_x^T(t)P_2e_x(t) + e_x^T(t)P_2\dot{e}_x(t) + \dot{e}_f^T(t)e_f(t) + e_f^T(t)\dot{e}_f(t) \\
=\ & [(A-LC)e_x(t) + Ee_f(t) + De_d(t) + (g(x)-g(\hat{x}))]^T P_2e_x(t) \\
& + e_x^T(t)P_2[(A-LC)e_x(t) + Ee_f(t) + De_d(t) + (g(x)-g(\hat{x}))] \\
& + [R_2Ce_x(t) + R_1e_f(t) + \dot{f}(t) - R_1f(t)]^T e_f(t) + e_f^T(t)[R_2Ce_x(t) + R_1e_f(t) + \dot{f}(t) - R_1f(t)] \\
=\ & e_x^T(t)[(A-LC)^T P_2 + P_2(A-LC)]e_x(t) + e_f^T(t)E^T P_2e_x(t) + e_d^T(t)D^T P_2e_x(t) \\
& + e_x^T(t)P_2Ee_f(t) + e_x^T(t)P_2De_d(t) + e_x^T(t)C^T R_2^T e_f(t) + e_f^T(t)R_1^T e_f(t) + e_f^T(t)R_2Ce_x(t) + e_f^T(t)R_1e_f(t) \\
& + 2(g(x)-g(\hat{x}))^T P_2e_x(t) + \dot{f}^T(t)e_f(t) - f^T(t)R_1^T e_f(t) + e_f^T(t)\dot{f}(t) - e_f^T(t)R_1f(t)
\end{aligned}
\tag{9}
$$

For simplification, denote as: $\vartheta = \vartheta_1 + \vartheta_2 + \vartheta_3 = r^T(\tau)r(\tau) - \sigma_2^2 v_2^T(\tau)v_2(\tau) - \dot{V}_2(\tau)$.

Substituting Eq. (9) into Eq. (8) and it follows from that

$$
\vartheta_1 = \begin{bmatrix} e_x^T(t) & e_f^T(t) & e_d^T(t) \end{bmatrix}
$$
$$
\begin{bmatrix}
C^T C - [(A-LC)^T P_2 + P_2(A-LC)] & -(P_2E + C^T R_2^T) & -P_2D \\
-(E^T P_2 + R_2C) & -(R_1^T + R_1) & 0 \\
-D^T P_2 & 0 & -\sigma_2^2 D^T D
\end{bmatrix}
\begin{bmatrix} e_x(t) \\ e_f(t) \\ e_d(t) \end{bmatrix}
$$

$$
\vartheta_2 = \begin{bmatrix} \dot{f}^T(t) & f^T(t) & e_f^T(t) \end{bmatrix}
\begin{bmatrix}
-\sigma_2^2 I & \sigma_2^2 R_1 & -I \\
\sigma_2^2 R_1^T & -\sigma_2^2 R_1^T R_1 & R_1^T \\
-I & R_1 & 0
\end{bmatrix}
\begin{bmatrix} \dot{f}(t) \\ f(t) \\ e_f(t) \end{bmatrix}
$$

$$
\begin{aligned}
\vartheta_3 =\ & -\sigma_2^2(g(x)-g(\hat{x}))^T(g(x)-g(\hat{x})) - 2\sigma_2^2(g(x)-g(\hat{x}))^T De_d(t) \\
& - 2(g(x)-g(\hat{x}))^T P_2e_x(t) \\
\leq\ & -\sigma_2^2 L_g^2 \|e_x(t)\|^2 - \sigma_2^2 \varepsilon_3 L_g^2 \|e_x(t)\|^2 - \varepsilon_3^{-1}\sigma_2^2 \lambda_{\max}(D^T D)\|e_d(t)\|^2 \\
& - \varepsilon_4 L_g^2 \|e_x(t)\|^2 - \varepsilon_4^{-1}\lambda_{\max}(P_2^T P_2)\|e_x(t)\|^2 \\
\leq\ & 0
\end{aligned}
$$

Hence, when $\vartheta_1 < 0$, $\vartheta_2 < 0$, the performance index (8) holds.

For the dimension matching motivation, a given constant matrix $(\prod)_{2 \times 4}$ is introduced, as a result,

$$\begin{bmatrix} -\left[(A - LC)^T P_2 + P_2 (A - LC) \right] & -P_2 E & -P_2 D + C^T \Pi D & C^T \\ -E^T P_2 & -(R_1^T + R_1) - R_2 R_2^T & -R_2 \Pi D & -R_2 \\ -D^T P_2 + D^T \Pi^T C & -D^T \Pi^T R_2^T & -\sigma_2^2 D^T D - D^T \Pi^T \Pi D & -D^T \Pi^T \\ C & -R_2^T & -\Pi D & -I \end{bmatrix} < 0$$

(10)

$$\begin{bmatrix} -\sigma_2^2 I & \sigma_2^2 R_1 & -I \\ \sigma_2^2 R_1^T & -\sigma_2^2 R_1^T R_1 & R_1^T \\ -I & R_1 & 0 \end{bmatrix} < 0$$

(11)

Therefore, the RGFDO is asymptotically convergence with the defined performance index when the parameters satisfy the constraint condition (10) and (11).

4 Simulation Results

The proposed scheme has been performed on flexible joint robotic system with a one-link manipulator actuated by a DC motor to evaluate the performance of the algorithm proposed and the simulation results are implemented by Simulink in Matlab.

4.1 Simulation Parameters

Consider the system described as Eqs. (1)–(2), the state-space model of flexible joint robotic system is [23]

$$\begin{cases} \dot{\theta}_m = \omega_m \\ \dot{\omega}_m = \frac{k}{J_m}(\theta_1 - \theta_m) - \frac{b}{J_m}\omega_m + \frac{\kappa_\tau}{J_m}u \\ \dot{\theta}_1 = \omega_1 \\ \dot{\omega}_1 = \frac{k}{J_1}(\theta_1 - \theta_m) - \frac{mgh}{J_1}\sin(\theta_1) \end{cases}$$

(12)

where, system states are $x^T(t) = (x_1(t), x_2(t), x_3(t), x_4(t)) = (\theta_m, \omega_m, \theta_1, \omega_1)$, J_1 and J_m are inertia of the motor and link. The system parameters are as follows

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -48.6 & -1.25 & 48.6 & 0 \\ 0 & 0 & 0 & 10 \\ 1.95 & 0 & -1.95 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 21.6 \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, D = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 0 \\ 12.5 \\ 0 \\ 0 \end{bmatrix}, g(x(t)) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -0.333\sin(x_3(t)) \end{bmatrix}$$

In the simulation, the simulation time is T = 100 s, control input pre-designed is assumed to be $u(t) = 0.09$ and the system with the external disturbances $d(t) = -0.009 \times \sin (0.6 \times t)$. Response of state under the fault-free case and the faulty case are shown in Figs. 1 and 2 respectively. Besides, the system with the actuator fault as follows

$$f(t) = \begin{cases} 0.235 + 0.26 \times \cos(1.9 \times t) & t \in [30, 60) \\ 0 & t \in other \end{cases} \qquad (13)$$

4.2   RGFDO Implementation

In this subsection, the parameters R_2 and L in Theorem 1 are solved by the LMITOOL in Matlab. Detailed implementations of RGFDO proposed are as follows.

The parameters results of RGFDO is

$$R_1 = -120, \quad R_2 = [35.2240 \quad 27.7240], \quad L = \begin{bmatrix} 0.9213 & 0.3602 \\ 0.6369 & 5.9850 \\ 1.0231 & 0.9826 \\ 0.0611 & 0.3861 \end{bmatrix}$$

As a result, the RGFDO can be implemented by the parameters solved. Figures 3, 4 and 5 show the state estimation, fault estimation and state estimation
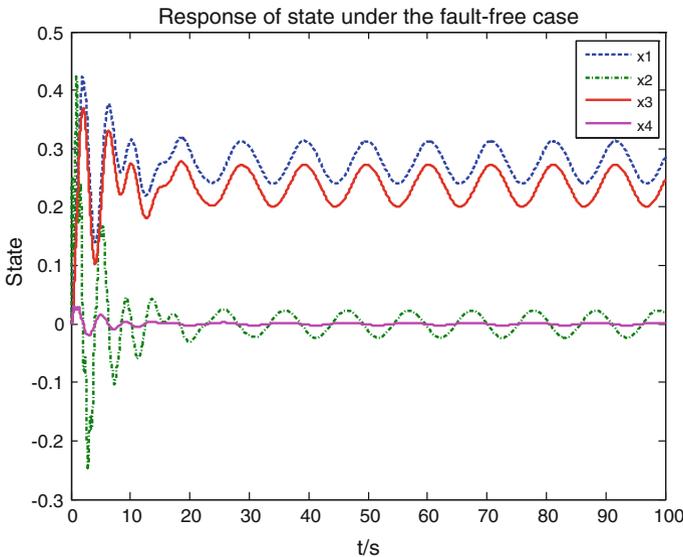
Fig. 1 Response of state under the fault-free case
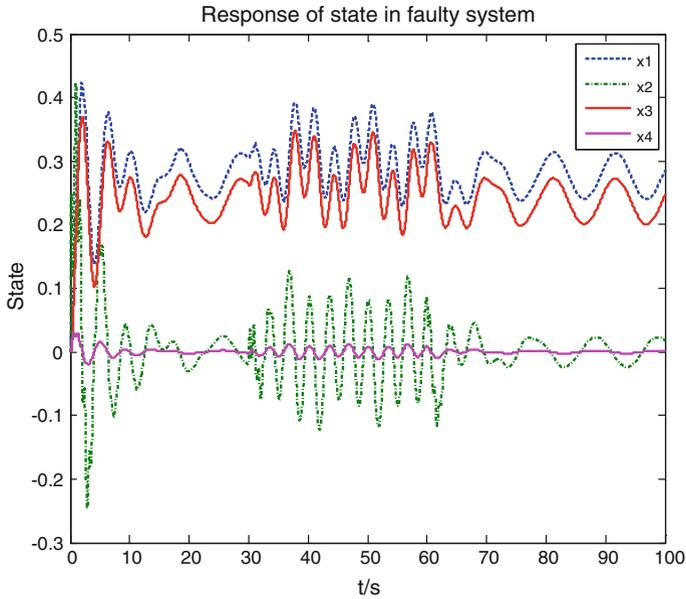
Fig. 2 State response under the faulty case
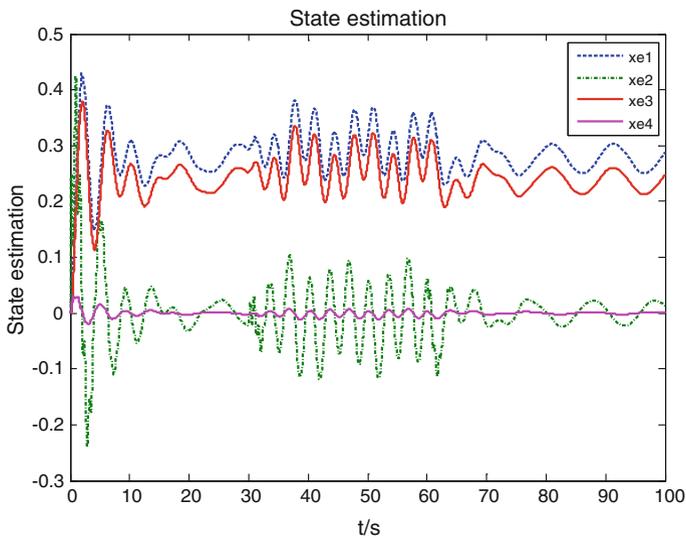
Fig. 3 State estimation

errors respectively. It can be seen that the desired fault estimation result is obtained from Fig. 4. As shown in Fig. 5, we can see that the state estimation errors are asymptotically convergence; besides the good tracking performance has been
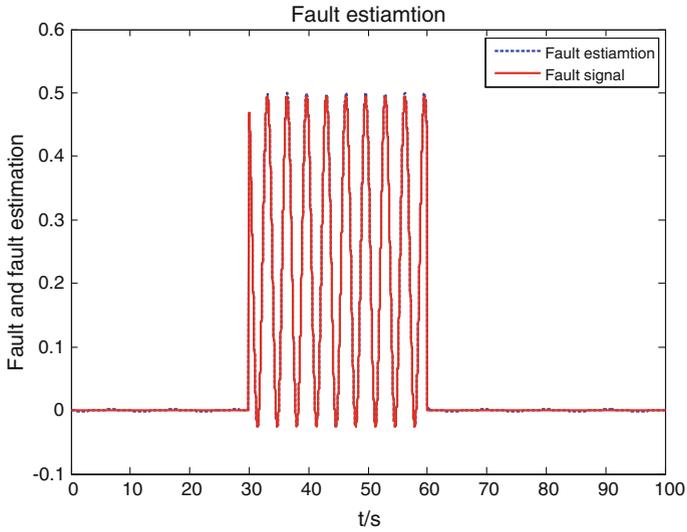
Fig. 4 Fault signal and fault estimation
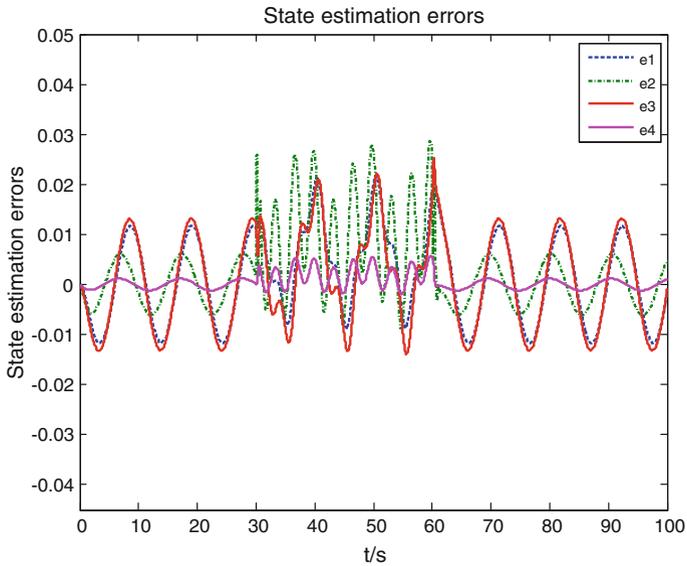
Fig. 5 State estimation errors

achieved although there is a significant jump at the actuator failure instant. And then, the statistics of state estimation errors are concluded in Table 1, obviously, the desired performance of RGFDO is also can be shown by it.

Table 1 Data statistics of state estimation errors

	e_1	e_2	e_3	e_4
Mean	0.0001765	0.0012480	−0.0004033	0.0002008
Std	0.0080960	0.0065530	0.0090480	0.0012810

5   Conclusions

In this paper, robust generalized fault diagnosis observer is proposed for a class of Lipschitz nonlinear system with external disturbances and actuator fault. As shown in our results, the desired fault signal estimation is obtained. However, expanding the algorithm to more general nonlinear systems is included in our further research works.

References

1. Yang, H., Jiang, B., Zhang, Y.: Tolerance of intermittent faults in spacecraft attitude control: switched system approach. IET Control Theory Appl. **6**, 2049–2056 (2012)
2. Yao, L.N., Qin, J.F., Wang, A.P., Wang, H.: Fault diagnosis and fault-tolerant control for non-gaussian non-linear stochastic systems using a rational square-root approximation model. IET Control Theory Appl. **7**, 116–124 (2013)
3. Ma, H.J., Yang, G.H.: Detection and adaptive accommodation for actuator faults of a class of non-linear systems. IET Control Theory Appl. **6**, 2292–2307 (2012)
4. Hu, Q., Xiao, B., Friswell, M.I.: Fault tolerant control with H∞ performance for attitude tracking of flexible spacecraft. IET Control Theory Appl. **6**, 1388–1399 (2012)
5. Cieslak, J., Henry, D., Zolghadri, A.: Fault tolerant flight control: from theory to piloted flight simulator experiments. IET Control Theory Appl. **4**, 1451–1464 (2010)
6. Lee, H., Kim, Y.: Fault-tolerant control scheme for satellite attitude control system. IET Control Theory Appl. **4**, 1436–1450 (2010)
7. Khosrowjerdi, M.J., Nikoukhah, R., Safari-Shad, N.: A mixed H2/H∞ approach to simultaneous fault detection and control. Automatica **40**, 261–267 (2004)
8. Niemenn, H., Stoustrup, J.: Passive fault tolerant control of double inverted pendulum—a case study. Control Eng. Pract. **13**, 1047–1059 (2005)
9. Benosman, M., Lum, K.Y.: Passive actuators' fault-tolerant control for affine nonlinear systems. IEEE Trans. Control Syst. Technol. **18**, 152–163 (2010)
10. Polycapou, M.M., Trunov, A.B.: Learning approach to nonlinear fault diagnosis: detectability analysis. IEEE Trans. Autom. Control **45**, 806–812 (2000)
11. Mendonça, L.F., Sousa, J.M.C., Sá da Costa, J.M.G.: Fault tolerant control using a fuzzy predictive approach. Expert Syst. Appl. **39**, 10630–10638 (2012)
12. Khalil, H.K.: A note on the robustness of high-gain-observer-based controller to unmodeled actuator and sensor dynamics. Automatica **41**, 1224–1821 (2005)
13. Escobar, R.F., Astorga-Zaragoza, C.M., Téllez-Anguianoc, A.C., Juárez-Romerob, D., Hernández, J.A., Guerrero-Ramírez, G.V.: Sensor fault detection and isolation via high-gain observers: application to a double-pipe heat exchanger. ISA Trans. **50**, 480–486 (2011)
14. Zarei, J., Poshtan, J.: Design of nonlinear unknown input observer for process fault detection. Ind. Eng. Chem. Res. **49**, 11443–11452 (2010)
15. Xu, J., Lum, K.Y., Loh, A.P.: A gain-varying UIO approach with adaptive threshold for FDI of nonlinear F16 Systems. J. Control Theory Appl. **8**, 317–325 (2010)
16. Wan, D., Lum, K.Y.: Adaptive unknown input observer approach for aircraft actuator fault detection and isolation. Int. J. Adapt. Control Signal Process. **21**, 31–48 (2007)

17. Benosman, M., Lum, K.Y.: Application of absolute stability theory to robust control against loss of actuator effectiveness. IET Control Theory Appl. **3**, 772–788 (2009)
18. Yang, H., Cocquempot, V., Jiang, B.: Robust fault tolerant tracking control with application to hybrid nonlinear systems. IET Control Theory Appl. **3**, 211–224 (2009)
19. Yin, S., Zhu, X.P., Kaynak, O.: Improved PLS focused on key-performance-indicator-related fault diagnosis. IEEE Trans. Industr. Electron. **62**(3), 1651–1658 (2015)
20. Dunnett, S.J., Mao, L., Jackson, L.M.: Fault Diagnosis of Practical Proton Exchange Membrane Fuel Cell System Using Signal-Based Techniques. Institut National Polytechnique De Toulouse, Toulouse (2015)
21. Gao, Z., Ding, S.X., Cecati, C.: Real-time fault diagnosis and fault-tolerant control. IEEE Trans. Industr. Electron. **62**(6), 3752–3756 (2015)
22. Basile, F., Cabasino, M.P., Seatzu, C.: State estimation and fault diagnosis of labeled time petri net systems with unobservable transitions. IEEE Trans. Autom. Control **60**(4), 997–1009 (2015)
23. Spong, M.: Modeling and control of elastic joint robots. ASME J. Dyna. Syst. Measure Control **109**, 310–319 (1987)

Selective Weighting PTS PAPR Reduction Scheme in OFDM Systems

Ling-Yin Wang, Hua Yuan and Li-Guo Liu

Abstract Partial Transmit Sequence (PTS) is a promising scheme for improving peak-to-average power ratio (PAPR) performance of orthogonal frequency division multiplexing (OFDM), with no signal distortion. Unfortunately, a complicated phase weighting process is involved in the original PTS (O-PTS), which induces high complexity. In this paper, a selective weighting PTS (SW-PTS) PAPR reduction scheme is proposed. In SW-PTS, selective weighting is firstly employed for reducing computational complexity. Then, unlike O-PTS, some specialized phase weighting sequences are generated for weighting the first subblock sequence to improve PAPR reduction performance. With massive computer simulations being done, proposed SW-PTS can reduce computational complexity clearly and achieve similar PAPR reduction performance compared to O-PTS.

Keywords OFDM · PAPR · PTS · Selective weighting

1 Introduction

As an attractive multicarrier modulation scheme, orthogonal frequency division multiplexing (OFDM) provides high-speed data transmission and the ability to resist frequency-selective fading channels [1]. As a result, it has been adopted in many fields of wireless communications including asymmetric digital subscriber line (ADSL), wireless local area networks (WLANs), etc.

L.-Y. Wang (✉) · H. Yuan · L.-G. Liu
School of Information Science and Engineering, University of Jinan,
Jinan 250022, China
e-mail: andrewandpipi@hotmail.com

H. Yuan
e-mail: yuanh@ujn.edu.cn

L.-G. Liu
e-mail: liulg@ujn.edu.cn

Unfortunately, one of the major shortcomings in OFDM systems is high peak power of transmitted signals, which induces large peak-to-average power ratio (PAPR). The large PAPR puts forward higher requirements for high power amplifiers. Otherwise, it will bring severe system performance degradation. To improve this problem, several methods for PAPR reduction have been presented [2, 3]. Among these methods, a category of schemes are focused on reducing the probability of OFDM signals with large PAPR, such as selected mapping (SLM) [4], interleaving [5] and partial transmit sequence (PTS) [6–8]. Without any signal distortion, this category can provide good PAPR property. But for original PTS (O-PTS) [6], finding OFDM candidate sequence with the best PAPR is a highly complicated problem, which usually induces large complexity.

In this paper, a selective weighting PTS (SW-PTS) PAPR reduction scheme in OFDM is proposed. In SW-PTS, selective weighting for subblock sequences is firstly adopted for reducing computational complexity. Then, to increase the number of OFDM candidate sequences, some specialized phase weighting sequences are generated for weighting the first subblock sequence, which equivalently makes PAPR reduction performance better. Massive computer simulations show that proposed SW-PTS can achieve dramatic computational complexity reduction and similar PAPR reduction performance compared with O-PTS.

2 OFDM System Model and Original PTS

Assume that there are N subcarriers in an OFDM system. The nth sample x_n of a discrete-time transmitted OFDM signal can be given by

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi kn/N}, 0 \le n \le N-1 \tag{1}$$

where X_k is the complex data symbol carried by the kth subcarrier.

For a discrete-time OFDM signal, its PAPR can be explained by the ratio of the maximum power to the average power of this signal, given by

$$\text{PAPR}(\boldsymbol{x}) = 10 \log_{10} \frac{\max_{0 \le n \le N-1}\{|x_n|^2\}}{E\{|\boldsymbol{x}|^2\}} (\text{dB}) \tag{2}$$

where $E\{\cdot\}$ is mathematical expectation operator and $\boldsymbol{x} = [x_0, x_1, \ldots, x_{N-1}]$.

To evaluate PAPR reduction performance of OFDM systems, complementary cumulative distribution function (CCDF) [9] is usually used, expressed by

$$\text{CCDF}(N, \text{PAPR}_0) = \Pr\{\text{PAPR} > \text{PAPR}_0\} = 1 - \left(1 - e^{-\text{PAPR}_0}\right)^N \tag{3}$$

where PAPR_0 is a given threshold of PAPR.

Specially, to approximate the real PAPR of continuous-time OFDM signals, the oversampling is usually adopted. A discrete-time OFDM signal with four times oversampling (i.e., $L = 4$) can be effective to reach the real PAPR [10], where L is the oversampling factor.

For O-PTS scheme, the complex data sequence $\boldsymbol{X} = [X_0, X_1, \ldots, X_{N-1}]$ is firstly partitioned into V disjoint subblock sequences $\boldsymbol{X}^v, v = 1, 2, \ldots, V$. The elements of each subblock sequence can be expressed by

$$X_k^v = \begin{cases} X_k & if \ X_k \subset \boldsymbol{X}^v \\ 0 & if \ X_k \not\subset \boldsymbol{X}^v \end{cases} \tag{4}$$

where X_k^v represents the kth element of the vth subblock sequence and X_k denotes the kth complex data symbol of $\boldsymbol{X}$. Thereupon, the complex data sequence $\boldsymbol{X}$ can be shown by $\boldsymbol{X} = \sum\limits_{v=1}^{V} \boldsymbol{X}^v$.

By employing inverse fast Fourier transform (IFFT) operations, the frequency-domain subblock sequence $\boldsymbol{X}^v$ is transformed into the time-domain one $\boldsymbol{x}^v$. Then, these subblock sequences are independently weighted by phase weighting factors. All the weighted subblock sequences are added to achieve an OFDM candidate sequence, expressed by

$$\boldsymbol{x}' = \text{IFFT}\left\{ \sum_{v=1}^{V} b_v \boldsymbol{X}^v \right\} = \sum_{v=1}^{V} b_v \cdot \text{IFFT}\{\boldsymbol{X}^v\} = \sum_{v=1}^{V} b_v \boldsymbol{x}^v \tag{5}$$

where b_v is phase weighting factor for the vth subblock sequence.

Finally, by employing different phase weighting factors, a number of OFDM candidate sequences can be obtained, and the one with the lowest PAPR can be found for transmitting. Moreover, the side information is required to help the receiver recover the input data correctly.

3   Selective Weighting PTS

3.1   Ideas of SW-PTS

For O-PTS, its main shortcoming is large computational complexity. Thereupon, in proposed SW-PTS, the selective weighting is firstly adopted for simplifying phase weighting process, which means that only a part of subblock sequences are weighted by the allowed phase weighting factors. For easily understanding selective weighting, we assume that the allowed phase weighting factors are used for all the even subblock sequences, and the odd subblock sequences are kept unchanged. After the selective weighting is completed, the number of OFDM candidate sequences generated at this stage can be expressed by

$$\begin{cases} W^{\frac{V}{2}} & V \text{ is even} \\ W^{\frac{V-1}{2}} & V \text{ is odd} \end{cases} \tag{6}$$

Based on Eq. (6), if the values of V and W are given, the number of OFDM candidate sequences in proposed SW-PTS is less than that in O-PTS. In other words, after selective weighting is performed in proposed SW-PTS, its PAPR reduction performance will degrade due to the fact that the number of OFDM candidate sequences is decreased.

To improve PAPR reduction performance of proposed SW-PTS, some specialized phase weighting sequences are generated for weighting the first subblock sequence. For example, if the number of subblock sequences V is four and the adjacent subblock partition is employed, the phase weighting sequences for the first subblock sequence can be given by

$$D_i = \left[d_{i,1}, d_{i,2}, \ldots, d_{i,\frac{LN}{4}}, \underbrace{0, \ldots, 0}_{\frac{3LN}{4}} \right], \quad i = 1, 2, \ldots, S \tag{7}$$

where $d_{i,j}$ denotes the jth element of the ith phase weighting sequence, S is the number of phase weighting sequences, N is the number of subcarriers in an OFDM system and L is the oversampling factor.

By employing these phase weighting sequences, the first subblock sequence can be changed and all the weighted even subblock sequences can be used once more to generate additional OFDM candidate sequences. Since the number of OFDM candidate sequences is increased, PAPR reduction performance of proposed SW-PTS can be improved. Significantly, these additional candidate sequences can be easily obtained by

$$y' = y - x^1 + x^{1'} \tag{8}$$

where y', y, x^1 and $x^{1'} = D_i x^1$ denote an additional OFDM candidate sequence, an OFDM candidate sequence generated at the stage of selective weighting, the original first subblock sequence and the weighted first subblock sequence respectively. It is noteworthy that compared with the OFDM candidate sequences generated at the stage of selective weighting, the generation of an additional one needs only $LN/4$ complex multiplications and $2LN$ complex additions. As for proposed SW-PTS and O-PTS, the detailed computational complexity analysis will be given in the following section.

Finally, among all the OFDM candidate sequences generated in two stages, the one with the lowest PAPR is chosen for transmitting. Just like O-PTS, the side information is also required for making the receiver recover the input data sequence successfully in proposed SW-PTS.

3.2 Computational Complexity Analysis

For a fair comparison, assume that the same number of candidate sequences is achieved in both proposed SW-PTS and O-PTS. Because the number of IFFT operations needed in proposed SW-PTS is same as that in O-PTS, the computational complexity from phase weighting process and the combination of all the subblock sequences is only taken into account.

For O-PTS, if the number of allowed phase weighting factors W and the number of subblock sequences V are given, W^{V-1} candidate sequences can be generated. One OFDM candidate sequence can result in $LN(V-1)$ complex multiplications and $LN(V-1)$ complex additions. Thus, computational complexity of O-PTS can be expressed by

$$Complex\ Mul.: LN(V-1)W^{V-1} \tag{9}$$

$$Complex\ Add.: LN(V-1)W^{V-1} \tag{10}$$

In proposed SW-PTS, since the OFDM candidate sequences are generated in two stages, its computational complexity consists of two parts. The first part comes from the generation of OFDM candidate sequences at the stage of selective weighting. The second part is caused by the generation of additional OFDM candidate sequences at the stage of weighting the first subblock sequence. Therefore, computational complexity of proposed SW-PTS can be given by

$$Complex\ Mul.: \begin{cases} \frac{LN}{4}\left(2VW^{\frac{V}{2}} + W^{\frac{V}{2}-1} - 1\right) & V \text{ is even} \\ \frac{LN}{4}\left[(2V-1)W^{\frac{V-1}{2}} - 1\right] & V \text{ is odd} \end{cases} \tag{11}$$

$$Complex\ Add.: \begin{cases} LN\left[(V-3)W^{\frac{V}{2}} + 2W^{V-1}\right] & V \text{ is even} \\ LN\left[(V-3)W^{\frac{V-1}{2}} + 2W^{V-1}\right] & V \text{ is odd} \end{cases} \tag{12}$$

4 Simulation Results

To show the performance of proposed SW-PTS, massive computer simulations are done. In simulations, the OFDM system parameters are set to the number of subcarriers $N = 512$ and the oversampling factor $L = 4$. For a fair comparison, the same number of candidate sequences is generated in both proposed SW-PTS and O-PTS. Figure 1 gives CCDFs of proposed SW-PTS, O-PTS and original OFDM with quadrature phase shift keying (QPSK).

As we can see in Fig. 1, PAPR reduction performance of proposed SW-PTS is a little worse than that of O-PTS. For CCDF = 0.1 %, when $V = 4$ and $W = 4$, PAPR
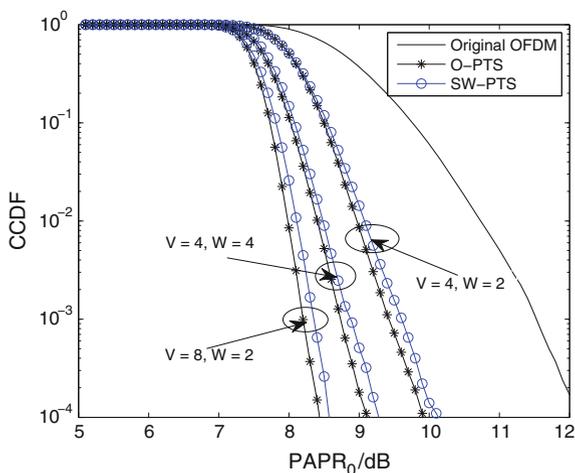
value of original OFDM is 11.52 dB and those of proposed SW-PTS and O-PTS
are 8.87 and 8.74 dB. Compared with O-PTS, the proposed SW-PTS shows a very
small performance loss. That is to say, proposed SW-PTS can obtain similar PAPR
reduction performance compared with O-PTS.

For computational complexity, the proposed SW-PTS can achieve dramatic
computational complexity reduction. For instance, if $V = 4$ and $W = 4$, the pro-
posed SW-PTS can result in 45 % complex addition reduction and 82.9 % complex
multiplication reduction; when $V = 8$ and $W = 2$, 62.5 % complex addition
reduction and 92.7 % complex multiplication reduction can be achieved. From this
we can see that, with the number of OFDM candidate sequences increasing, pro-
posed SW-PTS can achieve more computational complexity reduction.

5 Conclusion

In this paper, a selective weighting PTS (SW-PTS) scheme is proposed for
achieving computational complexity reduction and similar PAPR reduction per-
formance compared with O-PTS. In proposed SW-PTS, the selective weighting is
employed for reducing computational complexity. Then, some specialized phase
weighting sequences are generated for weighting the first subblock sequence to
improve PAPR reduction performance. Computer simulations and analysis show
that the proposed SW-PTS can obtain dramatic computational complexity reduction
and similar PAPR reduction performance compared with O-PTS.

References

1. Nee, R.V., Prasad, R.: OFDM for Wireless Multimedia Communications. Artech House Publishers, Boston (2004)
2. Han, S.H., Lee, J.H.: An overview of peak-to-average power ratio reduction techniques for multicarrier transmission. IEEE Wireless Commun. **12**(2), 56–65 (2005)
3. Jiang, T., Wu, Y.: An overview: peak-to-average power ratio reduction techniques for OFDM signals. IEEE Trans. Broadcast. **54**(2), 257–268 (2008)
4. Ji, J., Ren, G., Zhang, H.: A semi-blind SLM scheme for PAPR reduction in OFDM systems with low-complexity transceiver. IEEE Trans. Veh. Technol. **64**(6), 2698–2703 (2015)
5. Jayalath, A.D.S., Tellambura, C.: The use of interleaving to reduce the peak-to-average power ratio of an OFDM signal. In: IEEE Global Telecommunications Conference (GLOBECOM), pp. 82–86. IEEE, Piscataway (2000)
6. Muller, S.H., Huber, J.B.: OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences. IET Electron. Lett. **33**(5), 368–369 (1997)
7. Ye, C., Li, Z., Jiang, T., Ni, C.: PAPR reduction of OQAM-OFDM signals using segmental PTS scheme with low complexity. IEEE Trans. Broadcast. **60**(1), 141–147 (2014)
8. Li, L., Qu, D., Jiang, T.: Partition optimization in LDPC-coded OFDM systems with PTS PAPR reduction. IEEE Trans. Veh. Technol. **63**(8), 4108–4113 (2014)
9. Sohn, I.: A low complexity PAPR reduction scheme for OFDM systems via neural networks. IEEE Commun. Lett. **18**(2), 225–228 (2014)
10. Ochiai, H., Imai, H.: On the distribution of the peak-to-average ratio in OFDM signal. IEEE Trans. Commun. **49**(23), 282–289 (2001)

Single-User MIMO Systems for Simultaneous Wireless Information and Power Transfer

Jun Zhou, Jian-Xin Dai, Jing-Wei Liang, Jie Qi and Shuai Liu

Abstract This paper studies single-user Multiple-Input Multiple-Output (MIMO) systems for simultaneous wireless information and power transfer. The paper considered the broadcast system composing of three nodes, where one receiver decodes information and the other receiver harvests energy separately from the signals received from the transmitter. The transmitter and the receivers are equipped with multiple antennas. The problems of achieving the maximal information rate or harvested energy are raised. We obtained the most ideal transmission strategy to manage tradeoffs for getting maximal energy transfer versus information rate, which can be characterized by the boundary of a rate-energy (R-E) region. Simulation results show the feasible trade-off region for improving the system performance according to achievable data rate and harvested energy and the relationship of the energy to the power of the signal. A major contribution of this paper is the introduction of information transfer and the wireless energy tradeoff, the latter of which is derived using a mathematical method.

Keywords Simultaneous wireless information and power transfer (SWIPT) · Rate-energy tradeoff · Energy harvesting · MIMO system

J. Zhou · J.-W. Liang · J. Qi · S. Liu
College of Telecommunications and Information Engineering,
Nanjing University of Posts and Telecommunications, Nanjing 210003, China
e-mail: 15601582303@163.com

J.-W. Liang
e-mail: for.crx@qq.com

J. Qi
e-mail: qijieaka@163.com

S. Liu
e-mail: liu977803265@163.com

J.-X. Dai (✉)
School of Science, Nanjing University of Posts and Telecommunications,
Nanjing 210023, China
e-mail: daijx@njupt.edu.cn

1 Introduction

With the development of wireless communication technology, the requirements of wireless device batteries have become increasingly high. In the vast majority of cases, we can recharge or replace batteries to prolong the lifetime of a wireless device, but in some special instances, this is not feasible. Since radio-frequency (RF) signals that transport information can also carry energy, simultaneous wireless information and power transfer (SWIPT) is becoming a promising research point that draws increasing attentions. Recently SWIPT in the wireless networks have attracted many researchers' attentions from the information theoretic perspective [1, 2]. The Varshney and Grover considered a multiple-input single-output system in which one access point (AP) performs a SWIPT to serve a user terminal (UT) that is not equipped with external power supply [3]. Coordinated multi-point downlink transmission with M base stations and J mobile stations was investigated [4]. Guo and Durrani [5] studied simultaneous wireless information and power transfer in the ad hoc systems. Beamforming performs an important role in wireless systems with simultaneous transfer of power and information. For fading additive white Gaussian noise (AWGN) channels which were subjected to the time-varying interference, Liu et al. [6] proposed an original principle, we called that "opportunistic energy harvesting", where the receiver switches between decoding information and harvesting energy. Various functional receiver architectures were investigated and at the same time Zhou et al. [7] studied a new integrated energy and information receiver design. Zhou [8] studied optimal power splitting at the receiver. Michalopoulos et al. [9] studied the relay selection for SWIPT. Zhang and Liang [10] studied the dynamic resource allocation in the cognitive radio networks.

In the work presented here, we address the optimization problem of the maximum value of the received energy and the maximum transmission rate through the mathematical method. Also the joint optimization problem is solved by simulation.

2 System Model and Problem Formulation

2.1 System Model

This system design is based on the single-user scenario designed by Zhang and Ho [11] and Wang and Wang [12]. The single-user three nodes model is shown in Fig. 1. For the network consisting of three nodes, the left is the transmitter and the right nodes are the energy receiver and the information receiver. The energy receiver harvests energy from the signal sent by the transmitter, then the energy can be used to recharge the information receiver to decode the information received.

This section presents some further hypotheses. First, the number of transmitted symbols in a sufficiently long period of time can be approximated as infinite, so the wireless channel between the sender and receiver is a quasi-static fading

Fig. 1 The model of
single-user MIMO systems
for SWIPT
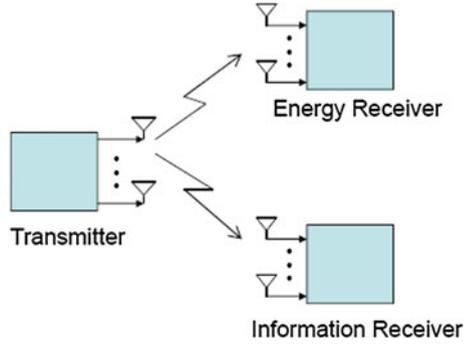
environment. Second, our research system usually works in high signal to noise ratio (SNR). Third, we assume that a multiple antennas MIMO system, and the transmitter is equipped with multi-antenna and every user terminal is equipped with at least one antenna.

We assume that the transmitter have $M > 1$ transmitting antennas, and the information decoding (ID) receiver and the energy harvesting (EH) receiver are equipped with $N_{ID} > 1$ and $N_{EH} > 1$ receiving antennas. So the channel characteristics from the transmitter to the energy receiver and the signal receiver can be represented by $G \in C^{N_{EH} \times M}$, and $H \in C^{N_{ID} \times M}$.

$x(n)$ is the base band signal sent by the transmitter, so the signal received by the receiver is represented in Eq. (1)

$$y_{ID}(n) = Hx(n) + z(n), \tag{1}$$

and $z(n) \in C^{N_{ID} \times 1}$ is interference noise, the mean value is 0, and the variance is σ^2. The signal received by the energy receiver is represented in Eq. (2)

$$y_{EH}(n) = Gx(n). \tag{2}$$

The covariance matrix of $x(n)$ can be represented by $S = E[x(n)^H x(n)]$. We can derive the total harvested RF-band power from (2), from the receiving antennas equipped by EH receiver, denoted by Q,

$$Q = \rho E\left[\|Gx(n)\|^2\right]. \tag{3}$$

$\rho \in (0,1]$ means energy conversion efficiency. We also can derive information rate from Eq. (1), from the receiving antennas equipped by ID receiver, denoted by R,

$$R = B \, \log\left|I + \frac{HSH^H}{\sigma^2}\right|. \tag{4}$$

Finally, we assume that the average power is constrained and denoted by

$$E\left[\|x(n)\|^2\right] = tr(S) \leq P. \tag{5}$$

2.2 Problem Formulation

Firstly, we consider the channel from the transmitter to the ID receiver when the value of energy harvested has a minimum value. Under the circumstance, the design goal for S is to get the maximal information rate. Considering Eqs. (3)–(5), the problem could be expressed as

$$\begin{aligned} \max_{S} \quad & C = B \, \log\left|I + \frac{HSH^H}{\sigma^2}\right| \\ \text{(P1)}: s.t. \quad & tr(S) \leq P, S \succeq 0 \\ & Q = \rho tr(GSG^H) \geq Q_o \end{aligned}, \tag{6}$$

$S \succeq 0$ means that signal is physically present. We assume that $\sigma^2 = 1$, $\rho = 1$. In addition, since B is constant, we can ignore it. So the problem can be simplified:

$$\begin{aligned} \max_{S} \quad & C = \log|I + HSH^H| \\ \text{(P1)}: s.t. \quad & tr(S) \leq P, S \succeq 0 \\ & Q := tr(GSG^H) \geq Q_o \end{aligned}. \tag{7}$$

The Lagrangian of (P1) can be written as

$$L(s, \lambda, u) = \log|I + HSH^H| + \lambda\left[tr(GSG^H) - Q_o\right] - u[tr(S) - P]. \tag{8}$$

The Lagrange dual function of (P1) is

$$g(\lambda, u) = \max_{S \succeq 0} L(S, \lambda, u). \tag{9}$$

We can derive that

$$\begin{aligned} g(\lambda, u) &= \max_{S \succeq 0} \log|I + HSH^H| + \lambda tr(GSG^H) - utr(S) \\ &= \max_{S \succeq 0} \log|I + HSH^H| - tr\left[(uI - G^H G)S\right]. \end{aligned} \tag{10}$$

Let g_1 be the maximum eigenvalue of matrix, we know that $u > g_1$ must be true. Make $A = uI - G^H G$, so $A > 0$ ($u > g_1$), and thus A^{-1} exists. So

$$g(\lambda, u) = \max_{S \succeq 0} \log|I + HSH^H| - tr(AS). \tag{11}$$

The dual solution of problem (P1) can be described as the solution of $\min_{\lambda \geq 0, \, u \geq 0}$ $g(\lambda, u)$.

The SVD of matrix $HA^{-1/2}$ is solved by $HA^{-1/2} = U_0\Lambda_0^{1/2}V_0^H$, where $U_0 \in C^{M \times T_1}$, $V_0 \in C^{M \times T_1}$, $\Lambda_0 = diag(h_1', \ldots, h_{T_1}')$, $h_1' \geq h_2' \geq \cdots \geq h_{T_1}' \geq 0$. For the given value of λ and u, it has been shown that the optimal solution of the problem [10]

$$S = A^{-1/2}V_0\Lambda_0V_0^HA^{-1/2}, \tag{12}$$

and the subgradient of $g(\lambda, u)$ at the point $[\lambda, u]$ is $[tr(GSG^H) - Q_0, P - tr(S)]$, after using the ellipsoid method to get the λ^* and u^*, we can get the optimal S^*. The above approach for settling (P1) is summed up in Box 1.

Box. 1 Algorithm for Settling Problem (P1)
Initialize $u > 0$, $\lambda \geq 0$, $u > \lambda g_1$

Repeat
Compute S using (11) with the given λ and u
Compute the subgradient of $g(\lambda, u)$
Update λ and u using the ellipsoid method subject to $u > \lambda g_1 \geq 0$
Until λ and u converge to the prescribed accuracy
Set $S^* = S$

Secondly, we consider the channel from the transmitter to the EH receiver when the value of information rate has a minimum value. Under the circumstance, the design goal for S is to get the maximal energy harvested by the EH receiver. Considering formulas (3)–(5), and considering that $\sigma^2 = 1$, $\rho = 1$ and B is constant, the problem can be expressed as

$$\begin{aligned} \max_{S} \quad & Q = tr(GSG^H) \\ (\text{P2}): s.t. \quad & tr(S) \leq P, S \succeq 0 \\ & R = \log|I + HSH^H| \geq R_o \end{aligned} \tag{13}$$

We consider the special case $R_0 = 0$, in other words, we don't consider the information rate, and we can find the optimal solution [5]

$$S_{EH} = Pv_1v_1^H. \tag{14}$$

And the optimal value is given by

$$Q_{\max} = g_1 P. \tag{15}$$

Let $T_2 = \min(M, N_{EH})$, and the SVD of matrix G is $G = U_G \Gamma_G^{1/2} V_G^H$, where $U_G \in C^{N_{EH} \times T_2}$, and $V_G \in C^{M \times T_2}$, each of which consists of orthogonal columns with unit norm, $\Gamma_G = diag(g_1, \ldots, g_{T_2})$, $g_1 \geq g_2 \geq \cdots \geq g_{T_2}$, meanwhile let v_1 express the first column of V_G. We can get the following conclusion $S_{EH} = P v_1 v_1^H$ and $Q_{\max} = g_1 P$.

Lastly, we consider the MIMO link without any minimum value. Here, the design goal for S is to get the best tradeoff between information rate and energy harvested. Considering formulas (3)–(5), the problem can be formulated as

$$(P3): f(R, Q) \underset{S}{=} \left\{ (R, Q) : R \leq \log|I + HSH^H|, \quad Q \leq tr(GSG^H), tr(S) \leq P, S \succeq 0 \right\}. \tag{16}$$

We solve this problem to get the rate-energy trade-off by simulation.

3 Simulation

In the part, we will show some mathematical results to illustrate and verify the above hypothetical analyses.
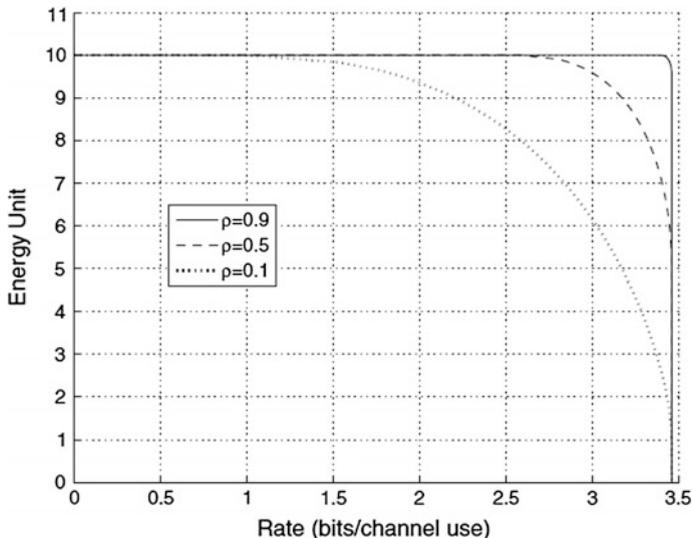
Fig. 2 Rate-energy tradeoff for a MIMO system to the (separated) EH and ID receivers
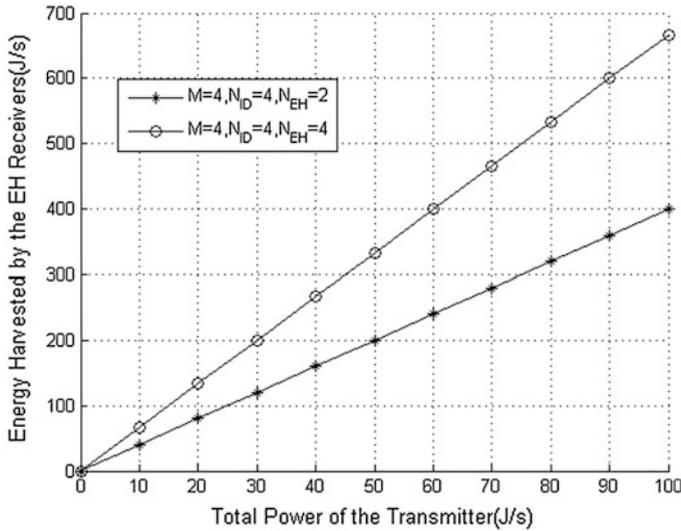
Fig. 3 Comparison of harvested energy as a function of P_ψ in MIMO system

Figure 2 shows the curves of the rate versus the energy under different value of energy conversion efficiency. With the increase of ρ, the curve will move up. Meanwhile, the intersection of the curve with the X axis and the Y axis remain constant. For a given ρ, as the rate increases, at first the energy remain unchanged, and then decreases, at last falls sharply.

Figure 3 shows the curves of the energy versus the power of signal sent by the transmitter, with the increase of power, the energy increase proportionally. This is consistent with our intuition.

4 Conclusion

In the paper, we first solve the optimization problem of the maximum value of the received energy and the maximum transmission rate subjected to the minimum rate and minimum energy, respectively. The optimization problem of the maximum transmission rate is solved by mathematical method and the method is different from previous works. Finally, the joint optimization problem is solved by simulation. While we did not determine an optimal solution to the general situation, and this is the future work direction. Furthermore, we should also try to extend the conclusion to the case of multi-users. A major contribution of this paper is the introduction of the wireless energy and information transfer tradeoff, so as to derive the tradeoff by mathematical method.

References

 1. Varshney, L.R.: Transporting information and energy simultaneously. In: IEEE International Symposium on Information Theory, pp. 1612–1616. IEEE Press, Piscataway (2008)
 2. Grover, P., Sahai, A.: Shannon meets Tesla: wireless information and power transfer. In: IEEE International Symposium on Information Theory, vol. 41(3), pp. 2363–2367. IEEE Press, Piscataway (2010)
 3. Liu, C.F., Lee, C.H.: Simultaneous wireless information and power transfer under different CSI acquisition schemes. IEEE Trans. Wireless Commun. **14**(4), 1911–1926 (2015)
 4. Liang, K., Zhao, L.Q., Yang, K.: A fair power splitting algorithm for simultaneous wireless information and energy transfer in CoMP downlink transmission. Wireless Pers. Commun. 1–24 (2015)
 5. Guo, J., Durrani, S.: Outage probability of ad hoc networks with wireless information and power transfer. IEEE Wireless Commun. Lett. **4**(4), 1 (2015)
 6. Liu, L., Zhang, R., Chua, K.C.: Wireless information transfer with opportunistic energy harvesting. IEEE Trans. Wireless Commun. **12**(1), 288–300 (2013)
 7. Zhou, X., Zhang, R., Ho, C.K.: Wireless information and power transfer: architecture design and rate-energy tradeoff. Commun. IEEE Trans. **61**(11), 4754–4767 (2013)
 8. Zhou, X.Y.: Training-based SWIPT: optimal power splitting at the receiver. IEEE Trans. Veh. Technol. **64**(9), 1–6 (2015)
 9. Michalopoulos, D.S., Suraweera, H.A., Schober, R.: Relay selection for simultaneous information transmission and wireless energy transfer: a tradeoff perspective. IEEE J. Sel. Areas Commun. **33**(8), 1578–1594 (2015)
10. Zhang, R.Y., Liang, C., Cui, S.: Dynamic resource allocation in cognitive radio networks. IEEE Signal Process. Mag. **27**(3), 102–114 (2010)
11. Zhang, R., Ho, C.K.: MIMO broadcasting for simultaneous wireless information and power transfer. Wireless Commun. IEEE Trans. **12**(5), 1989–2001 (2013)
12. Wang, S.H., Wang, B.Y.: Robust secure transmit design in MIMO channels with simultaneous wireless information and power transfer. IEEE Signal Process. Lett. **22**(11), 2147–2151 (2015)

Study on Nonlinear Dynamic Behavior in a Linear-Mode Switched-Capacitor DC-DC Converter

Xian-Rui Li, Yong-Gang Zhao, Yan-Li Zhu and Rui Shen

Abstract This paper studies the nonlinear dynamic behavior exhibited by a switched-capacitor DC-DC regulator in linear-mode. Numerical simulations are presented not only to illustrate our theoretical results, but also to exhibit the system's complex dynamical behavior, such as the cascade of period-doubling bifurcation in orbits of period 2, 4, 8, 16 quasi-periodic orbits, and the chaotic sets. Lyapunov exponents are numerically computed to further confirm the complexity of the dynamical behavior. Finally, a chaotic attractor in the converter is obtained by simulating the circuit with PSPICE program.

Keywords Switched-Capacitor DC-DC converter · Bifurcation · Chaos

1 Introduction

In Switched-Capacitor DC-DC design, power electronics engineers frequently observed electromagnetic interference problems [1] and some noise [2]. Profound studies about the nonlinear phenomena in the field of power electronics are profitable and mandatory. However, studying bifurcation and chaos is necessary to understand the behavior in power electronics varies with the change in parameters.

X.-R. Li (✉) · Y.-G. Zhao · Y.-L. Zhu · R. Shen
Institute of Electronic CAD, Xidian University, Xi'an, China
e-mail: lixianrui4213@126.com

Y.-G. Zhao
e-mail: zhaoyonggang4213@126.com

Y.-L. Zhu
e-mail: zhuyanli4213@126.com

R. Shen
e-mail: shenrui4213@126.com

Some paper have discussed the bifurcation and chaos phenomena in Buck converter [3–7], Boost converter [8, 9] and Cuk converter [10–13]. However, Switched-Capacitor converter, has not been reported. This paper analyses the bifurcation and chaos in the Linear-mode Switched-Capacitor DC-DC converter.

2 Linear-Mode Switched-Capacitor DC-DC Converter

A linear-mode circuit comprises an adjustable resistor R_C, an operational amplifier AMP, an oscillator OSC, a load resistance R_L, a capacitor C_1, an output capacitor C_0, two resistors (R_1, R_2), two switches (S_1, S_2), a reference voltage V_{REF}, a feedback voltage V_{FB}, an output voltage V_{CPO}, a clock signal CLK signal, and its non-signal $\overline{CLK}$, the diagram is shown in Fig. 1. Two switch states can be distinguished in continuous mode. State A, the charging state, switches S_1 on and switches S_2 off when CLK is high, and the input current I_{IN} flows from input voltage V_{IN} through the adjustable resistor R_C to capacitor C_1. State B, the energy transmission stage, switches S_1 off and switches S_2 on when CLK is low; and capacitor C_1 is connected in parallel between V_{CPO} and ground, which is discharged simultaneously to the load through the switch S_2. When the supply voltage or load current changes, the feedback loop will adjusts the resistance of adjustable resistor R_C to stabilize output voltage by controlling operational amplifier AMP [14].
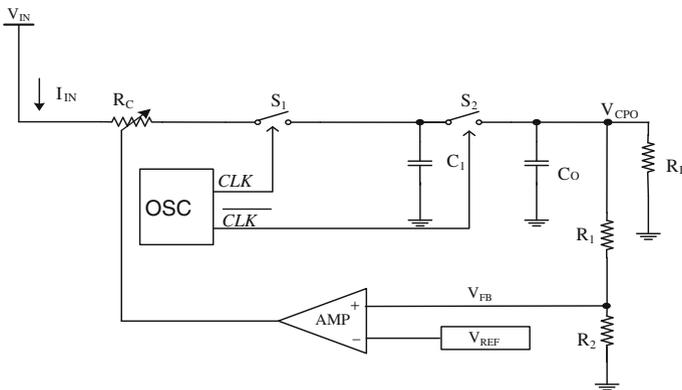
Fig. 1 Structure diagram of linear-mode switched-capacitor DC-DC converter

3 Two-Dimensional Discrete-Time Model

The simple schematic diagram of state A and state B are shown in Fig. 2 [15, 16].

As is shown in Fig. 2, x denotes the state vector of the circuit, $x[v_1, v_2]$. v_1 denotes the voltage of capacitor C_1, v_2 denotes the voltage of capacitor C_0. It can be described by the following of state equations:

$$\text{State A:} \begin{cases} \dot{v_1} R_1 C_1 + v_1 = Vin \\ \dot{v_2} C_O = -\frac{v_2}{R_L} \end{cases}, \quad \text{State B:} \begin{cases} \dot{v_1} R_2 C_1 + v_1 = v_2 \\ \dot{v_2} R_2 C_O = v_1 - \frac{R_2 + R_L}{R_L} v_2 \end{cases} \quad (1)$$

Switches are ideal switches, where r denotes equivalent series resistance of capacitors, K denotes feedback coefficient, X denotes reference voltage, T denotes a clock cycle, D denotes the duty cycle, then $R_1 = R_C + r$, $R_2 = r$, the state A's and B's matrix are given by:

$$A_1 = \begin{bmatrix} -\frac{1}{(R_C + r)C_1} & 0 \\ 0 & -\frac{1}{C_O R} \end{bmatrix}, B_1 = \begin{bmatrix} \frac{1}{(R_C + r)C_1} \\ 0 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} -\frac{1}{rC_1} & \frac{1}{rC_1} \\ \frac{1}{rC_O} & -\frac{r+R_L}{rC_O R_L} \end{bmatrix}, B_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (2)$$

where $R_C = R + K[V_2(n) - X]$. In the time $[nT, (n + 1)T]$, we can get:

$$x(n+1) = e^{\tau_2 A_2} e^{\tau_1 A_1} x(n) + e^{\tau_2 A_2} \int_{t_n}^{t'_n} e^{(t'_n - S)A_1} B_1 E ds, \quad (3)$$

where I is the second order unit matrix, $x(n) = x(t_n) = [v_1(n) \; v_2(n)]^T$, $E = V_{in}$, $\tau_1 = DT$, $\tau_2 = (1 - D)T$.

(a)

(b)
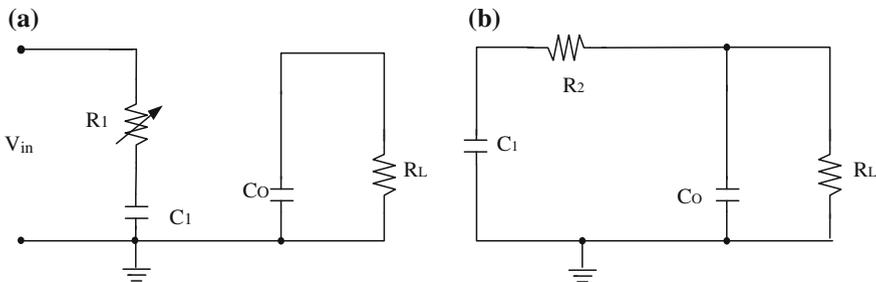
Fig. 2 **a** Schematic diagram of state A. **b** Schematic diagram of state B

Using the undetermined coefficient method $e^{\tau_1 A_1}$, $e^{\tau_2 A_2}$ can be obtained as:

$$e^{\tau_1 A_1} = \begin{bmatrix} e^{-\frac{h_1 \tau_1}{C_1}} & 0 \\ 0 & e^{-\frac{\tau_1}{C_0 R_L}} \end{bmatrix}, \quad e^{\tau_2 A_2} = \begin{bmatrix} a - b\frac{h}{C_1} & b\frac{h}{C_1} \\ b\frac{h}{C_0} & a - b(\frac{1}{C_0 R_L} + \frac{h}{C_0}) \end{bmatrix} \quad (4)$$

where $\quad h = \frac{1}{r}, \quad h_1 = \frac{1}{R_C + r}, \quad \alpha = \frac{h}{C_1} + \frac{h}{C_0} + \frac{1}{C_0 R_L}, \quad \beta = \sqrt{\alpha^2 - \frac{4h}{C_1 C_0 R_L}},$
$a = \left(\frac{1}{2} + \frac{\alpha}{2\beta}\right)e^{\frac{\beta - \alpha}{2}\tau_2} + \left(\frac{1}{2} - \frac{\alpha}{2\beta}\right)e^{\frac{-\beta - \alpha}{2}\tau_2}, \quad b = \frac{1}{\beta}\left(e^{\frac{\beta - \alpha}{2}\tau_2} - e^{\frac{-\beta - \alpha}{2}\tau_2}\right).$

From Eq. (3), the two-dimensional discrete-time model of linear-mode switched-capacitor DC-DC converter can be calculated. The equation group is written as follows:

$$\begin{cases} v_1(n+1) = \left(a - \frac{bh}{C_1}\right)\left[e^{-\frac{h_1 \tau_1}{C_1}}v1(n) + E\left(1 - e^{-\frac{h_1 \tau_1}{C_1}}\right)\right] + \frac{bh}{C_1}e^{-\frac{\tau_1}{C_0 R_L}}v_2(n) \\ v_2(n+1) = \frac{bh}{C_0}\left[e^{-\frac{h_1 \tau_1}{C_1}}v1(n) + E\left(1 - e^{-\frac{h_1 \tau_1}{C_1}}\right)\right] + e^{-\frac{\tau_1}{C_0 R_L}}\left[a - b\left(\frac{1}{C_0 R} + \frac{h}{C_0}\right)\right]v_2(n) \end{cases}$$

$$(5)$$

4 Stability Analysis

In order to study the system stability, it is necessary to calculate the system condition in stable operation and the values of the circuit parameter in unstable operation according to the Jacobian matrix on the fixed point of the discrete-time model.

The circuit parameters of Fig. 1 are set as follows: Vin = 6 V, X = 3 V, C_1 = 4.7 μF, C_0 = 4.7 μF, R_L = 20 Ω, T = 1 us, r = 0.1 Ω, D = 0.2571. Then the fixed point and the Jacobian matrix on the fixed point of discrete-time model are computed by MATLAB. In the case of the linear-mode switched-capacitor DC-DC converter, the two-dimensional discrete-time model and the Jacobian matrix can be written in the following form:

$$x(n+1) = f(x(n), K) = \begin{bmatrix} f_1(v_1(n), v_2(n), K) \\ f_2(v_1(n), v_2(n), K) \end{bmatrix},$$
$$J(x_0) = \begin{bmatrix} \partial f_1/\partial v_1(n) & \partial f_1/\partial v_2(n) \\ \partial f_2/\partial v_1(n) & \partial f_2/\partial v_2(n) \end{bmatrix}$$

$$(6)$$

The steady fixed point is $|\lambda_1| < 1$ and $|\lambda_2| < 1$. Where $\lambda_{1,2}$ are eigenvalues of $J(x_0)$. Using Eq. (6) we can calculate the Jacobian matrix at $x_0 = [3.0093\ 3]^T$, R = 5 Ω.

$$J(x_0)_{max} = \begin{bmatrix} 0.51477 & 0.47562 - 0.00324K \\ 0.47184 & 0.51653 - 0.00297K \end{bmatrix} \quad (7)$$

The system condition in stable operation can be calculated as $0 < K < 698.4$. The system will not steadily operate when K is greater than 698.4.

5 Bifurcation Diagram

To verify the theoretical analysis result in the pervious section, feedback coefficient K is as the parameter, using MATLAB to calculate Eq. (6) by using iterative calculation for several times. The step length of K is taken as 1 and 300 iterative values are removed, which are related to the initial value, the bifurcation diagram of voltage v_2 with feedback coefficient K. The bifurcation diagram is shown in Fig. 3.

As is shown in the Fig. 3, it is seen that the bifurcation diagrams of the buck converter when K is about 699. The period 2 behavior again bifurcates to period 4 when K is about 860. As K is increased above 910, the system enters chaos.

6 Largest Lyapunov Exponent

Lyapunov exponents are often used to study the chaotic systems. Lyapunov exponents for all directions is negative, zero and positive for stable periodic and subharmonic motions, quasi-periodic motions, and chaotic motions, respectively
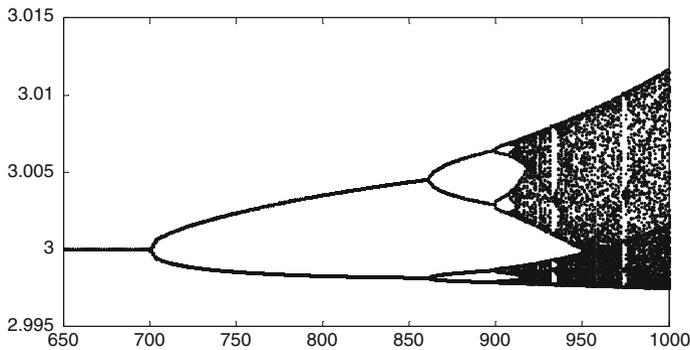
Fig. 3 Bifurcation diagram of voltage v_2 with feedback coefficient K
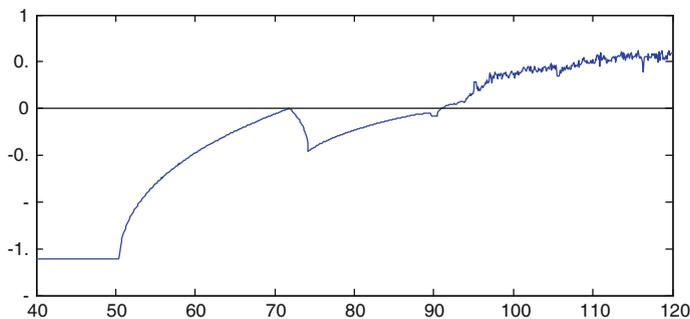
Fig. 4 Curve of largest Lyapunov exponent at different values of K

[11]. The Lyapunov exponent of n-dimensional discrete-time model and the largest Lyapunov exponent (denoted by $\lambda_{\max}$) can be obtained as follows:

$$\lambda_k = \lim_{i \to \infty} \frac{1}{i} \ln \left| \lambda_k^{(i)} \right|, (k = 1, \ldots, n), \quad \lambda_{\max} = \max(\lambda_1, \lambda_2, \ldots, \lambda_n) \qquad (8)$$

where λ_k denotes the eigenvalue of the Jacobian matrix at each iteration, i denotes the number of iterations. The largest Lyapunov exponent at different values of K is shown as Fig. 4.

7 Verification by Computer Simulations of the Actual System

An actual linear-mode switched-capacitor DC-DC converter is simulated to verify the pervious theoretical analysis. The circuit is shown in Fig. 5.

As is shown in Fig. 5, v_1 is corresponding to the oscillator OSC in Fig. 1, The circuit parameters are set as follows: the cycle of v_1 is 1 us, D = 0.2571, V_{ref} = 3 V, R2 = 9.9 k, R3 = 100 Ω, the other parameters are the same as is shown in Sect. 3, the value of K is changed by changing the resistance of R1. When k = 600, 800, 1000, the system state are in steady state, period state, and chaos state corresponding and the simulation results are shown in Figs. 6, 7 and 8 correspondingly.
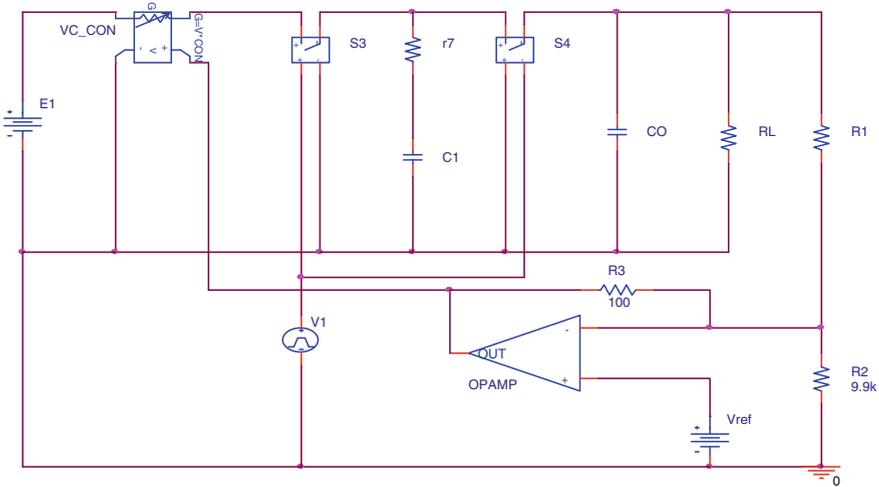
Fig. 5 Simulation circuit of linear-mode switched-capacitor DC-DC converter

(a)

(b)
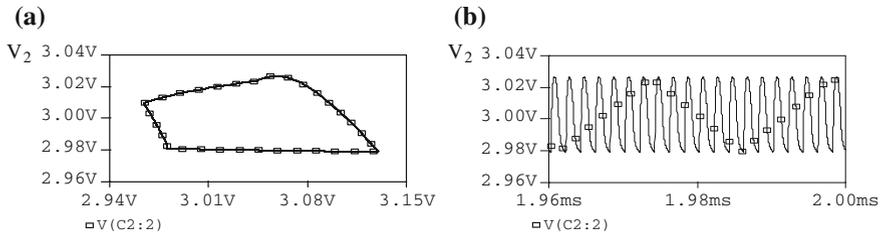
Fig. 6 Steady state, **a** time domain waveform, **b** phase portrait

(a)

(b)
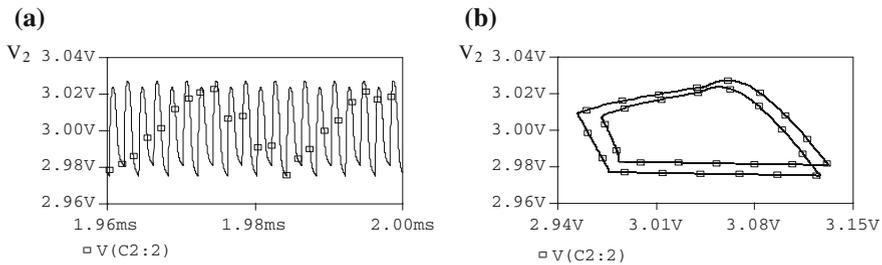
Fig. 7 Two-period state, **a** time domain waveform, **b** phase portrait

(a)

(b)
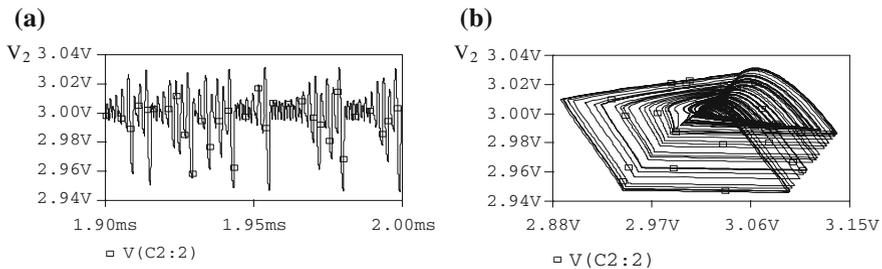
Fig. 8 Chaos state, **a** time domain waveform, **b** phase portrait

8 Conclusion

In this paper, the bifurcation and chaos phenomena in the Linear-Mode Switched-Capacitor DC-DC converter is analyzed. Using the model, the bifurcation phenomena under different circuit parameters have been analyzed. The simulation results indicate that the Switched-Capacitor converter illustrates a wide range of nonlinear behaviors. The system goes to chaos via period-doubling route with the bifurcation parameter changed.

The rich nonlinear phenomena in the Linear-Mode Switched-Capacitor DC-DC converter reveals that simple equations result in complex behaviors and different systems have similar behaviors. Choose the parameter values to obtain the proper behavior is important, So study domains of bifurcation and chaos is necessary More importantly, the bifurcation and chaos can be avoided if we understand the nonlinear phenomena.

References

1. Liu, M.L.: Demystifying Switched-Capacitor Circuits. Butterworth-Heinemann Press, Oxford (2006)
2. Liu, J., Chen, Z.M., Zhong, Y.R.: Study on switched-capacitor DC-DC converter. Acta Electronica Sinica **25**, 88–90 (1997)
3. Hsieh, F.H., Wang, H.K., Chang, P.L., Lin, S.H.: Nonlinear behaviors in a voltage-mode controlled half-bridge buck converter via varying load resistance. In: International Conference on Machine Learning and Cybernetics, vol. 2, pp. 627–631. IEEE, Piscataway (2014)
4. Basak, B., Parui, S.: Exploration of bifurcation and chaos in buck converter supplied from a rectifier. IEEE Trans. Power Electron. **25**, 556–1564 (2010)
5. He, S.Z., Xu, J.P., Zhou, G.H., Bao, B.C., Yan, T.S.: Stability control and mode shift of ramp compensation in V^2 controlled buck converter. Chin. J. Electron. **24**, 295–299 (2015)
6. Yuan, G.H., Banerjee, S., Ott, E., Yorke, J.A.: Border-collision bifurcation in the buck converter. IEEE Trans. Circ. Syst.-Fundam. Theory Appl. **45**, 07–715 (1998)
7. Angulo, F., Olivar, G., Londono, G.: Generalisation of ZAD strategy: an application to a DC-DC buck converter. Dynamics **166**, 98–107 (2001)
8. Iqbal, S., Zang, X.Z., Zhu, Y.H., Zhao, J.: Study of bifurcation and chaos in DC-DC boost converter using discrete-time map. In: International Conference on Mechatronics and Control, pp. 1813–1817. IEEE, Piscataway (2014)
9. Zamani, N., Niroomand, M., Ataei, M.: Bifurcation and chaos control in power-factor correction boost converter. In: Conference on Electrical Engineering, pp. 1307–1312. IEEE, Piscataway (2014)
10. Kavitha, A., Uma, G., Reesha, M.B.: Analysis of fast-scale instability in a power factor correction Cuk converter. IET Power Electron. **5**, 1333–1340 (2012)
11. Deivasundari, P., Uma, G., Kavitha, A., Umamaheswari, M.G.: Coexistence of fast-scale and slow-scale instability in Cuk power factor correction AC-DC pre-regulators under nonlinear current-mode control. IET Power Electron. **6**, 78–87 (2013)
12. Parvathysha, D., Gov, U.: Chaotic dynamics of a zero average dynamics controlled DC-DC Cuk converter. IET Power Electron. **7**, 289–298 (2014)
13. Tse, C.K., Lai, Y.M., Iu, H.H.C.: Holf bifurcation and chaos in a free-running current-controlled Cuk switching regulator circuits and systems I. IEEE Trans. Circ. Syst. I Fundam. Theor. Appl. **47**(4), 448–457 (2000)
14. Charge Pump DC/DC Converters with Reduced Input Noise, US Patent 6411531 B1, June 25 (2002)
15. Zhang, L.S., Wang, L.Z.: Study on nonlinear dynamics behavior in PWM switched-capacitor DC-DC converter. Acta Electronica Sinica **36**, 266–270 (2008)
16. Chan, W.C.Y.: Smdy of chaos in DC/DC switching converters. Ph.D. thesis, Hong Kong Polytechnic University, Hongkong (1998)

Synchronization Algorithms Applied to Code Division Multiplexing Link Based on the Walsh Code of Ground Based Beamforming

Wei Wan, Le Tang and Hang Tu

Abstract A synchronization algorithm that applies to the Code Division Multiplexing (CDM) link based on Walsh Code is proposed to solve problems of pre-existing algorithms based on pilot. The algorithm proposed by this paper achieved blind synchronization without pilot, and improved applicability at low SNR by incoherent accumulation. The simulation results indicate that the proposed algorithm achieves high synchronization precision and acceptable synchronization time, and can thus be applied to CDM links based on Walsh Codes of GBBF (Ground Based Beam forming).

Keywords Walsh code · GBBF · Synchronization · CDM

1 Introduction

In satellite communication system, Digital Beamforming (DBF) is widely-applied to improve the capacity and efficiency of satellites; GBBF is one of these applications. GBBF is a technology that transfers Digital Beamforming from an on-board platform to a gateway on the ground. Compared to Digital Beamforming on-board platforms, Ground Based Beamforming offers great advantages in terms of price, power and flexibility. Hence, GBBF has been attracting increasing research attention in recent years [1].

The bottleneck of GBBF technology is limitation of bandwidth of the feeder link. The arrays of on-board antenna groups may include several hundreds, which

W. Wan (✉) · L. Tang · H. Tu
School of Information and Electronics, Beijing Institute
of Technology, Beijing, China
e-mail: 540570549@qq.com

L. Tang
e-mail: 402448157@qq.com

H. Tu
e-mail: 364726566@qq.com

leads to a high dimension of feeder signals, and the total bandwidth of feeder signals may exceed the bandwidth of the feeder link. This problem may be solved by three methods. The first method is on-board signal compression. Since its basic idea is to eliminate the signal correlation between feeds, a decrease of bandwidth will be very limited if the number of the feeds is similar to the number of beams. The second method is to apply a high frequency feeder link, such as a Q or V band, in which more spectra are available. This method requires mature radio technology for the Q/V band, and must update the on-board platform. The third method is to improve the ratio of bandwidth utilization. The traditional method of channel reuse is Frequency Division Multiplexing (FDM), which is a mature technology, but guard bands waste bandwidth of the feeder link. Compared to FDM, CDM can improve bandwidth efficiency, and is more suitable for the feeder link of GBBF [2].

The application of CDM technology in GBBF feeder links requires orthogonal codes to divide channel; it can be an m sequence, gold sequence or Walsh Code. The m sequence and gold sequence have excellent autocorrelation, which is easy for signal synchronization in the receiver. However, cross-correlation of the m sequence and gold sequence is unsatisfactory, and will cause channel interference. In GBBF, channel interference is unacceptable since it will seriously impact beamforming. Walsh Code has excellent cross-correlation, which can avoid channel interference, but the autocorrelation of Walsh Code is disadvantageous; thus, synchronization based on the autocorrelation of Walsh Code in the receiver is difficult to realize.

The pre-existing synchronization schemes of CDM links based on Walsh Code are pilot-aided [3]. For example, in the second cellular mobile communication system IS-95, the synchronization of the down link is achieved by sending a high-powered pilot to the first sub-channel divided by Walsh Code. The use of a pilot will waste channel capacity, and the complexity of demodulation will increase at the receiver since the power of the pilot is high [4, 5]. This paper attempts to solve this problem by investigating the synchronization algorithm of CDM links based on Walsh Code, and achieved bind synchronization without a pilot by modifying the structure of the Delay Locked Loop (DLL).

2 Model of CDM Link Based on Walsh Code

A model of a CDM link based on Walsh Code is shown in Fig. 1.

Signals $S_1 \sim S_N$ are first modulated to signal $M_i(t)$, as follows:

$$M_i(t) = \sum_{k=-\infty}^{k=+\infty} S_{i,k} g(t - kT) \tag{1}$$

T represents the symbol period, g(t) is the gate signal of gate width T, and k is the number of symbols. To improve the randomness of the modulated signal, the m sequence is used to scramble the signals. The scrambled signals are shown as follows:
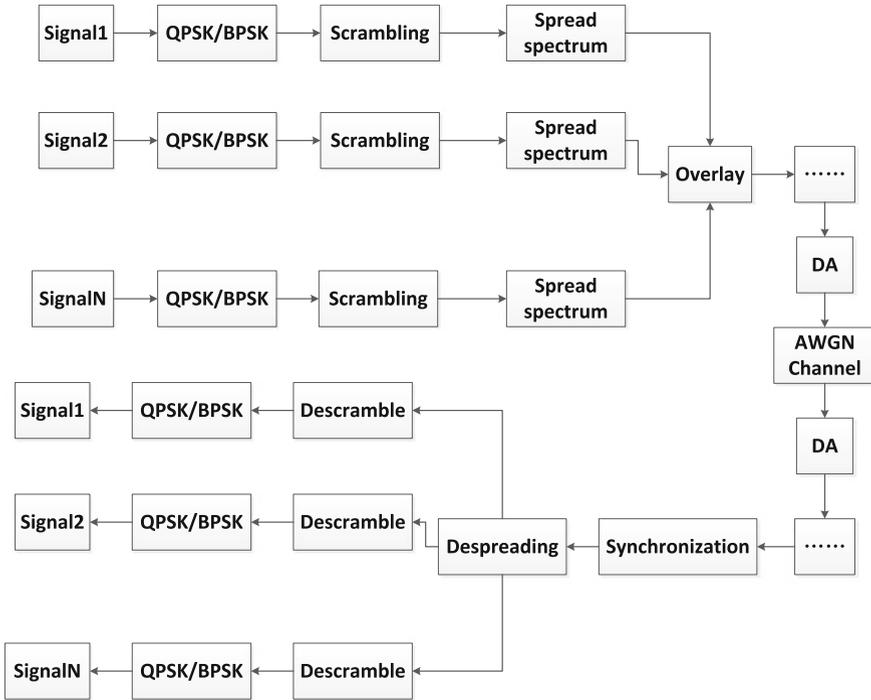
Fig. 1 Block diagram of CDM link based on Walsh Code

$$R_i = \sum_{j=-\infty}^{j=+\infty} M_i(t)m_i(t - jMT) \qquad (2)$$

where $m_i(t)$ represents the m sequence, and M is the length of $m_i(t)$. The scrambled signals are then spread spectrum by Walsh Code, and added to become one signal $F(t)$, as follows:

$$F(t) = \sum_{i=1}^{N} \sum_{n=-\infty}^{n=+\infty} R_i(t)W_i(t - nT) \qquad (3)$$

where $W_i(t)$ is the Walsh Code and $W_i(t) = \sum_{l=1}^{l=N} w_{i,l}g(Nt - lT_c)$, $T_c = \frac{1}{N}T$ is the chip period, and $w_{i,l}$ is the chip value of the Walsh Code.

Finally, the signal $F(t)$ is sent by DA after interpolation, forming and up-conversion. The receiver inverses the process to recover the original signals [6]. The difficulty of synchronization in CDM links based on Walsh Code is bit synchronization; therefore, the capture and carrier synchronization is assumed to be completed.

3   Synchronization of CDM Link Based on Walsh Code

The sending signal F(t) of CDM links based on Walsh Code is the synchronized sum of chips, which appears as the increase or decrease of chip amplitude. Therefore, F(t) can be regarded as a single-channel signal of which the chip amplitude changes continuously. The structure of traditional synchronization algorithms for single-channel signals, (e.g., DLL), is shown in Fig. 2. The principle of the algorithm is constantly adjusts interpolating points according to the phase discriminator until DLL reaches a steady state; hence, the phase discriminator is key for DLL to synchronize the signal [7, 8]. The principle of a phase discriminator (e.g., Gardner) is shown as Fig. 3. The Gardner error detection requires two over-sampling rates, which calculate error according to the value of (k − 1), y(k) and y(k − 0.5). The output of the phase discriminator is as follows:

$$u(k) = y(k - 0.5)(y(k - 1) - y(k)) \tag{4}$$

As shown in Fig. 3, the phase of interpolating points in Fig. 3a is identical to that of the optimum points, so that y(k − 0.5) is near zero and the phase discriminator output is zero. As shown in Fig. 3b, the phase of interpolating points is delayed compared to that of the optimum points, and the phase discriminator output is a positive number. As shown in Fig. 3c, the phase of interpolating points is ahead of that of the optimum
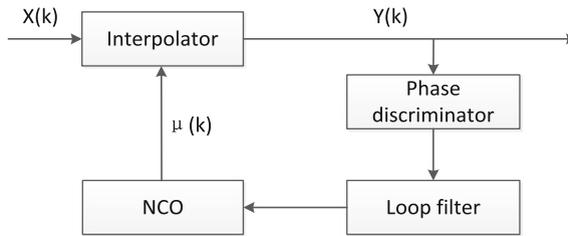
Fig. 2 Structure of DLL
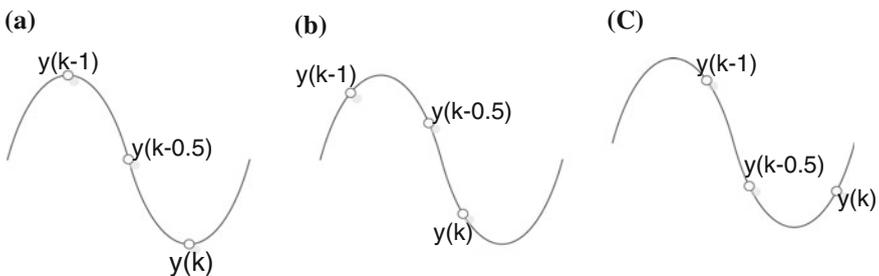
Fig. 3 Principle of Gardner

Fig. 4 Wave of received signal in CDM link based on Walsh Code
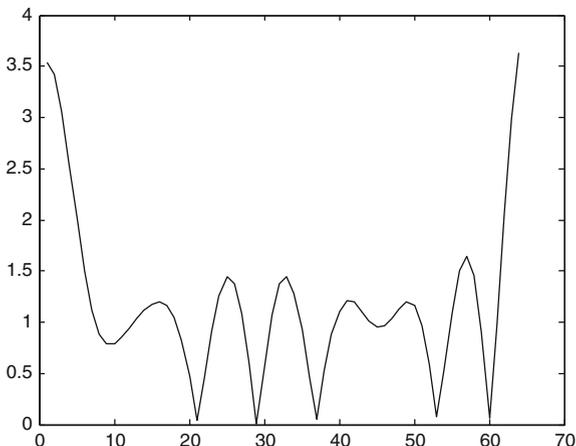
points, and the discriminator output is a negative number. DLL adjusts the interpolating points according to the polarity of the phase discriminator output.

Results indicate that the phase discriminator requires constant chip amplitude to obtain the correct outcome, indicating that DLL is unable to synchronize signals in which chip amplitude changes continuously, as shown in Fig. 4. To solve this problem, a Modified Delay Locked Loop (MDLL) is proposed.

The frame of a Modified Delay Locked Loop is shown in Fig. 5, which utilizes two increased interpolators in the Delay Locked Loop. The three interpolator output points A, B and C (with time interval d) and the phase discriminator outputs are determined by analyzing the values of A, B, and C. Points A, B and C must be in one chip, indicating that d < 0.5 chip; typically, d = 0.2 chip ∼ 0.3 chip.

The output of the phase discriminator is given as follows:

$$u_\tau(n) = K \times \frac{A(n) - C(n)}{A(n) + C(n)} \tag{5}$$

Fig. 5 The structure of MDLL
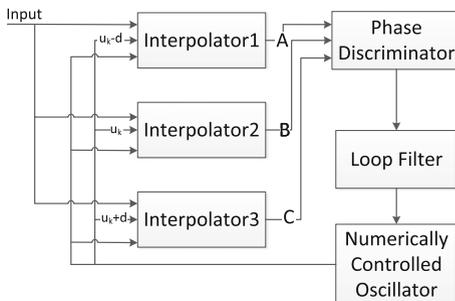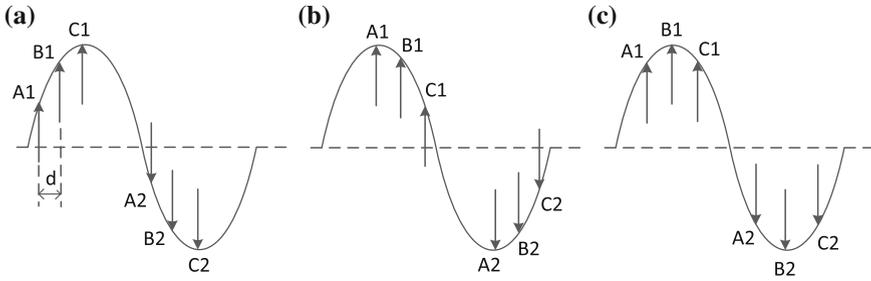
Fig. 6 Principle of MDLL

where $A(n)$ and $C(n)$ are the amplitude of point A and C, respectively; and K is a constant. The interpolator of MDLL adjusts the interpolating point according to the output of the phase discriminator. If $A(n) > C(n)$, shown as Fig. 6a, $\mu_\tau(n) > 0$ the MDLL delay the interpolating point; if $A(n) < C(n)$, shown as Fig. 6b, $\mu_\tau(n) < 0$, MDLL pulls ahead of the interpolating point; if $A(n) = C(n)$, shown as Fig. 6c, $\mu_\tau(n) = 0$, MDLL maintains a steady state. The output of the phase discriminator is determined only by the chip that includes A, B and C, indicating that the continuous change of chip amplitude between chips does not affect MDLL. Therefore, MDLL can be used in CDM links based on Walsh Code.

In practical applications, the amplitude of A, B and C is unstable due to noise interference, which will affect the output of the phase discriminator. To diminish the effect of noise, the mean value of multipoint A, B and C is used as the input of the phase discriminator. The value of K should change with time to balance the lock-in time and precision. The output of phase discriminator is as follows:

$$u_\tau(n) = K(n) \times \frac{\overline{A} - \overline{C}}{\overline{A} + \overline{C}} \tag{6}$$

where

$$K(n) = \begin{cases} K_2 \exp[K_1(1-n)], & n < a \\ K_1 \exp[K_1(1-a)], & n \geq a \end{cases} \tag{7}$$

$\overline{A}$ and $\overline{C}$ are the mean values of multiple point A and C; a is the value of n at the lock moment; K_1 and K_2 are constants (K_2 decides the initial value of $K(n)$, which will affect the lock-in time of MDLL, while K_1 decides the rate of convergence of $K(n)$, which will affect both the lock-in time and precision of MDLL). In practical applications, the value of K_2 should represent the maximum value, while the value of K_1 depends on the application.

4 Simulation

The down link of the IS-95 system also divided channel by Walsh Code; the synchronization scheme of the IS-95 system is based on a pilot which is described in previous research [9]. The simulation conditions are assumed to be as follows:

- The pilot used for IS-95 synchronization is an alternation of +1 and −1 sequence, and the power of pilot is 10 dB higher than the power of the data signal;
- The value of d in MDLL is 0.2 chip; $\overline{A}$ and $\overline{C}$ are the mean values of fifty A and C; $K_1 = 0.2$; $K_2 = 1.2$
- The spreading gain is equal to eight; all other conditions are identical.

As shown in Fig. 7, the synchronization error of MDLL is smaller than that of the IS-95 system when SNR is 10 dB or above; this is primarily because the data signals disturb synchronization in the IS-95 system, and the synchronization scheme used in IS-95 system cannot function when SNR is −10 dB or below, while MDLL can because MDLL accumulate the power of fifty chips, which sacrifices synchronization time, but improves the SNR of the received signals.

The comparison of synchronization time between MDLL and the IS-95 system at the same SNR are shown in Fig. 8. The synchronization time of the IS-95 system is approximately 300 chips, while that of MDLL is approximately 800 chips. It is observed that although MDLL accumulated the power of fifty chips when synchronizing, the increase in synchronization time is acceptable, indicating that the synchronization time of MDLL is eligible for GBBF.
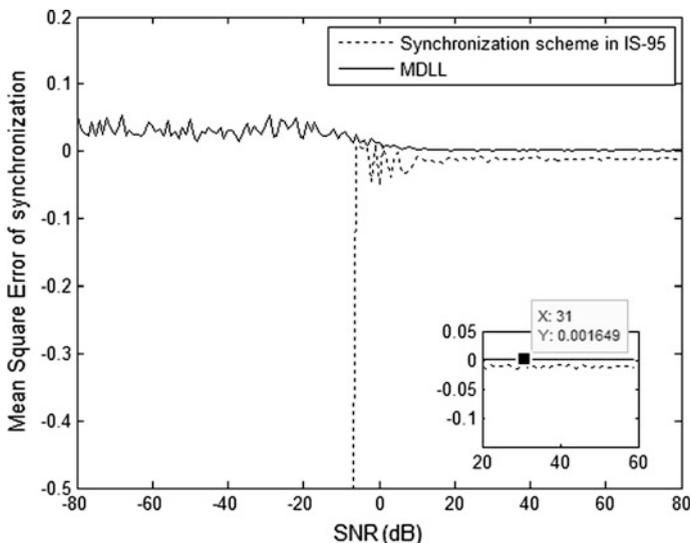
Fig. 7 Synchronization precision comparison
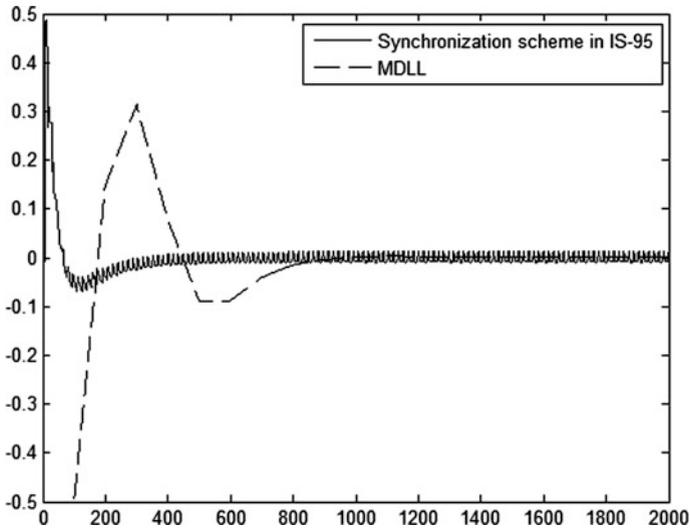
Fig. 8 Synchronization time comparison

5   Conclusion

This paper analyzed the synchronization algorithms of CDM links based on Walsh Code, and determined that synchronization schemes based on pilots will decrease transport efficiency and increase demodulation complexity. To settle this problem, MDLL was proposed and achieved blind synchronization without a pilot, and improved the SNR of received signals by accumulation. The simulation results indicate that precision of MDLL is high and that the synchronization time of MDLL is acceptable.

References

1. Gao, Z., Zhao, M., Pan, W., Wang, J.: Distributed ground-based beamforming. In: 31st AIAA International Communications Satellite Systems Conference, pp. 200–208. American Institute of Aeronautics and Astronautics, Reston (2013)
2. Shao, W., Yin, T.H., Meng, F.Q., Qian, Z.P.: An efficient spectrum utilization technique for feeder link in GBBF satellite-terrestrial system. In: The 8th Satellite Communications Academic Essays, pp. 224–231. The National Radio Application and Management of Academic Conference Proceedings, Tianjing (2013)
3. Yarkosky, M., Narayanabhatla, S., Cheong-Wai, N,J., Oh, D.: Method and system of reusing Walsh Codes to increase forward channel capacity. United States Patent, US8929195-B1, 06 Jan 2015

4. Kuzhaloli, S., Shaji, K.S.: Performance analysis of MC-CDMA in Rayleigh channel using Walsh Code with BPSK modulation. In: The 2nd International Conference on Information Systems Design and Intelligent Applications, pp. 565–572. Springer, Kalyani (2015)
5. Ho, C.K.: High bandwidth efficiency and low power consumption Walsh Code implementation methods for body channel communication. IEEE Trans. Microw. Theor. Tech. **62**(9), 1867–1878 (2014)
6. Von Der, E.U.A., Huang, K.Y., Chang, D.C.D., Mayfield, W.W.: Ground based beam forming utilizing synchronized code division multiplexing. United States Patent, US005903549A, 11 May 1999
7. Li, Y., Zhao, Y.L., Zhao, W.L., Wang, Y., Liu, Q.L.: Improved timing error detection algorithm in satellite communication. Syst. Eng. Electron. **36**(1), 149–154 (2014)
8. Nair, J.P., Bynam, K., Hong, Y.J., Kang, J.: Timing synchronization in super-regenerative receivers with a single quench cycle per symbol. In: 2014 IEEE International Symposium on Circuits and Systems, pp. 738–741. Institute of Electrical and Electronics Engineers Inc, Melbourne (2015)
9. Li, S.W., Chen, H.P.: The spread spectrum communication use Walsh Code based on broadband signal. Commun. Technol. **2**, 75–79 (2000)

The Application of FPGA in Optical Fiber Sensing and Communication Technology

Yan-Bing Li, Sheng-Peng Wan, Hao-Liang Lu and Qiang Gong

Abstract To obtain pulsed light signal used as pulsed pump light for optical fiber sensing and communication systems, a design scheme of generating pulsed light based on continuous laser and Field Programmable Gate Array (FPGA) is proposed in this paper. The pulsed light signals with minimum pulse width of 10 ns were obtained in the experiment, the pulse widths of the pulsed light signals could be regulated with the smallest increment of 5 ns, while the deviations of the pulse widths were less than 0.8 ns between the measurements and the expected values. The repetition frequency could also be adjusted from 0.05 Hz to 100 MHz. Not only is the pulsed light signal stable and reliable, but the complexity of the system is also reduced by using the design method in this paper.

Keywords Pulsed light · Repetition frequency · Electro-optic intensity modulator (EOM) · Pulse-width modulation (PWM) · FPGA

1 Introduction

For converting useful signals into electrical ones which are easily processed and conveniently observed in optical fiber sensing and optical fiber communication technology, it often needs to modulate the optical carrier wave transmitting in optical fiber by adopting PWM technique [1–4]. In view of the booming FPGA devices

Y.-B. Li · S.-P. Wan (✉) · H.-L. Lu · Q. Gong
Jiangxi Engineering Laboratory for Optoelectronics Testing Technology,
Nanchang Hangkong University, Nanchang, China
e-mail: sp_wan@163.com

Y.-B. Li
e-mail: bstir–myself@163.com

H.-L. Lu
e-mail: luhaolianglhl@163.com

Q. Gong
e-mail: moyi_gq@163.com

which are characterized by low cost, high integration and miniaturization, high arithmetic speed, flexible interface mode and control mode, using FPGA device as the hardware platform to design the pulse signal generator needed by producing pulsed light with the aid of the development software named QUARTUS II will be more convenient and flexible [5, 6]. Currently, some researchers have obtained electrical pulses with pulse width of 1 ns and repetition rates of 1–10 kHz in electrical field by using high-end FPGA development board and ultra high-speed transmission line [7].

The stable and reliable light pulses with different features could be obtained by drawing support from EOM and by means of electric control mode to modulate the light intensity of the continuous laser. Furthermore, the purpose of regulating the type of the pulsed light, the pulse width and the repetition frequency of the pulsed light signal could be realized perfectly by changing the states of the control keys on the basis of FPGA platform.

2 The Constitution of the Testing System

The testing system for generating pulsed light consists of narrow linewidth and continuous laser, Optical Isolator (ISO), Polarization Controller (PC), attenuator, EOM, FPGA, electrical pulse signal filtering and amplifying circuit, Direct Current (DC) signal filtering and amplifying circuit, Erbium-doped Optical Fiber Amplifier (EDFA), optical coupler with coupling ratio of 1:99, photodetector (PD) and oscilloscope. The structure diagram of the pulsed light generating system is shown in Fig. 1.
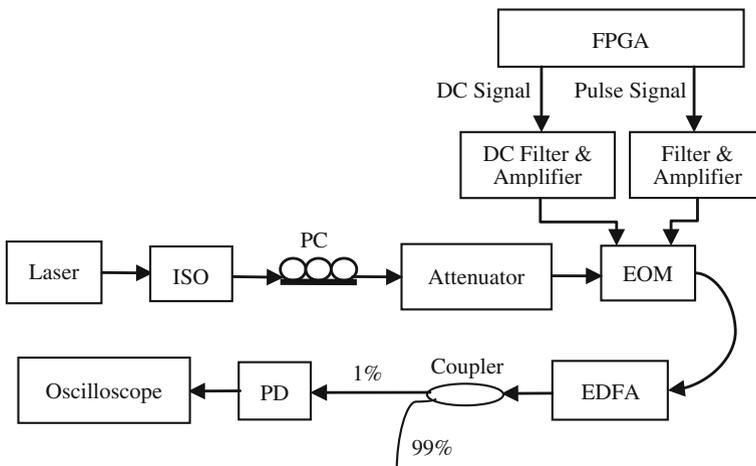
Fig. 1 The structure diagram of the pulsed light generating system

As shown in Fig. 1, the narrow linewidth laser provides the testing system with the initial and continuous laser light. The elliptically polarized light transmitting in the single-mode fiber will be changed into linearly polarized one by regulating the angles of the three rings of PC, the optical power of the linearly polarized light will be reduced by an attenuator to ensure that the light power supplied to EOM is smaller than the specified upper limit value of EOM, the continuous laser light showing linear polarization state will be turned into pulsed light by EOM which is under the combined action of the electrical pulse signal and DC bias voltage. The power of pulsed light will be amplified properly by EDFA for transmitting and using in the subsequent optical path. For observing the result of PWM intuitively and distinctly, the light pulses from the output port of the coupler whose splitting coefficient accounts for 1 % of the total input power will be converted into electrical pulses by a PD with bandwidth of 20 GHz, and a series of electrical pulses from the PD can be observed and recorded clearly on the screen of the oscilloscope with bandwidth of 1 GHz for 3 dB.

3 The Driving Circuit Module of EOM

Both DC bias voltage and electrical pulse modulating signals loaded to the input terminals of EOM are generated by FPGA device. The circuit units that we need to design by means of FPGA device include PLL frequency-multiplier circuit, key control module, DC signal and electrical pulse signal generating circuit. In addition, an external crystal circuit with main vibration frequency of 50 MHz is needed in circuit design which provides the driving circuit module with the initial reference clock. The chip named EP4CE15F17C8 of the series of Cyclone IVE made by Altera Corporation was used as the core platform of the circuit design in experiment, and each modularized circuit could be designed by drawing support from hardware description language for programming under the development environment called QUARTUS II 13.0. The structure diagram of the driving circuit module for EOM is shown in Fig. 2.
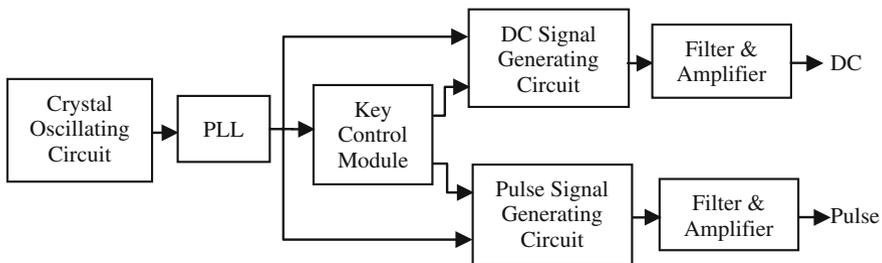
Fig. 2 The structure diagram of the driving circuit module for EOM

3.1 The DC Signal's Generation and Control

The clock signal with main frequency of 50 MHz from the crystal resonance circuit would be made frequency multiplication of 4 times for achieving the reference clock signal of 200 MHz whose clock period was 5 ns by the PLL frequency-multiplier circuit, the clock signal after frequency multiplication and timing constraint would be used as the reference input clock of the key control module and the DC signal generating circuit. The key control module would control FPGA whether or not to output the DC signal, and the output from the designated pin of FPGA remained high continuously under the action of the frequency multiplication clock signal if the DC signal control key was in the state of allowing DC voltage to be outputted. So, the device would offer the external circuit a stable DC signal which stemmed from the LVTTL electrical signal of 3.3 V coming from the internal integrated circuits of FPGA. Then, the DC voltage could be adjusted continuously for meeting the operating requirements of EOM by an external anti-jamming and amplifying circuit for DC signal.

3.2 The Pulse Signal's Generation and Control

The frequency multiplication clock signal of 200 MHz from the Phase Locked Loop (PLL) frequency-multiplier circuit was used as the reference input clock signal of the pulse signal generating circuit, and the clock frequency of 200 MHz could be splitted by adopting frequency division technique for obtaining the single pulse and double pulse signals whose repetition frequencies and pulse widths were adjustable. The pulse width of the single pulse signal can be described by

$$\tau = \frac{K+2}{F_{clk} \times N} \times 10^9 \tag{1}$$

where K is an adjustable parameter for pulse width and it belongs to natural number, F_{clk} is the main clock frequency of 50 MHz offered by the crystal unit circuit, N denotes the coefficient of frequency multiplication, the measure unit of pulse width τ is ns. For insuring high fidelity of pulse signal as much as possible [8], the minimum pulse width was limited to 10 ns for single pulse signal and double pulse signal, and the frequency multiplication coefficient was set to 4, furthermore, the variation of 5 ns was set as the smallest increment of pulse width in the experiment. Therefore, the goal of changing the pulse width can be achieved through regulating the parameter K.

The repetition frequency f of the single pulse signal is given by

$$f = \frac{F_{clk} \times N}{C \pm J} \tag{2}$$

as shown in Eq. (2), both F_{clk} and N are the same as in Eq. (1), besides, C denotes the count value of being aimed at the clock cycles of the frequency multiplication signal within the frequency divider, and J is the change of the number of clock cycles of the frequency multiplication clock, where C is a positive integer, and $C \geq 2$, and J is a natural number, as well as $J + 1 < C$, MHz is used as the unit of f. The value of f will decrease in steps when taking $C + J$ and it will increase gradually when taking $C - J$.

For double pulse signal, the minimum time interval between the two close neighbor pulses of the double-pulse would be no less than 10 ns in order to reduce their wave distortion as much as possible [9], and it could be changed with the smallest increment of 5 ns. The pulse widths of the two near neighbor pulses of the double-pulse can be regulated at the same time, and their pulse widths also can be determined by Eq. (1). In addition, the time span d can be described by

$$d = \frac{M + 2}{F_{clk} \times N} \times 10^9 \tag{3}$$

as shown in Eq. (3), both F_{clk} and N are the same as in Eq. (1), M is the regulated number of times of time span between the two close neighbor pulses of the double pulse signal, where M is a natural number. The unit of d is ns, and its minimum increment was limited to 5 ns in the design. Besides, the repetition frequency between the two distant neighbor pulses of the double-pulse can be similarly determined by Eq. (2).

All parameters of electrical pulse signals from FPGA could be changed and controlled by changing the states of keys of the key control module. Besides, The narrow pulse signal coming from FPGA is often accompanied by overshoot noise, undershoot and ringing, so, the filtering and amplifying circuit which is characterized by wide frequency bandwidth, high precision and high rate of voltage conversion is needed to process the pulse signals for obtaining quality pulses whose peak voltage and pulse waveform both can meet the usage requirements for EOM [10].

The electrical pulse signal and the DC signal after filtering and amplifying will be separately loaded to the Radio Frequency input end and the Bias input end of EOM for driving it to convert the continuous laser into the pulsed light who can change its features following the changing rule of the electrical pulse modulation signal [11].

4   The Experimental Result

It's necessary to transform the pulsed light into electrical signal which is more conveniently displayed and observed on the screen of the oscilloscope for observing the result of the modulated light signal clearly. However, it should be noted that only the pulsed light from the coupler's output port with coupling ratio of 1 % would be transmitted to the PD for conversion in order to avoid damaging the light-sensitive cell.

The results show that the deviations of the pulse widths for different types of pulsed light signals are less than 0.8 ns between the measurements and the expected values through comparing the experimental results to the expected pulse widths, so, they can meet the application requirements in distributed optical fiber sensing and optical fiber communication technology. A part of the measured waveforms of different pulse signals after photoelectric conversion are shown in Figs. 3 and 4.
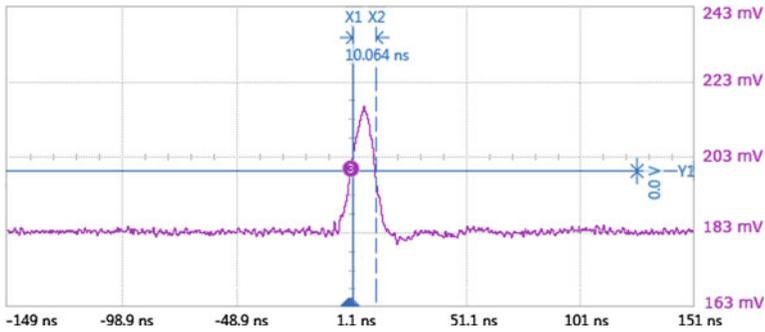
Fig. 3 The measured waveform of the single pulse signal with expected pulse width of 10 ns
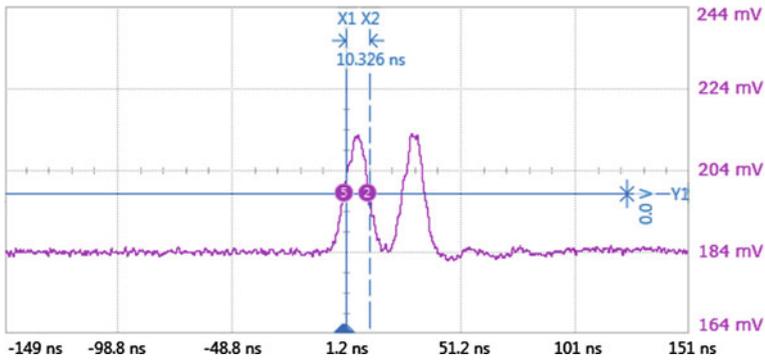
Fig. 4 The measured waveform of the double pulse signal with expected single pulse width of 10 ns and expectant repetition frequency of 100 MHz between the two close neighbor pulses

The light pulses with pulse width of 10 ns and different repetition frequencies were obtained in the experiment, and these pulses could meet the requirement for achieving the spatial resolution of 1 m when they were used as the pulsed pump light in the distributed optical fiber sensing system.

5 Conclusion

The experimental results show that the design method proposed in this paper for obtaining pulsed light signals is practicable, it not only can generate different types of optical pulse signals, but also the goals can be realized easily that both the pulse width and repetition frequency of the pulsed light can be regulated through adjusting the corresponding parameter of the electrical pulse from FPGA. Moreover, the output of the system is stable and reliable, and the expected design effect is achieved basically.

References

1. Xu, H.J., Fan, C.C.: Pilot tone modulation and its application in WDM optical fiber communication system. J. High Technol. Lett. **5**, 20–23 (1999) (in Chinese)
2. Song, M.P.: The technique of Brillouin scattering distributed optical fiber sensing based on microwave electro-optical modulation. J. Acta Opt. Sinica. **24**, 1111–1114 (2004) (in Chinese)
3. Fabien, R., Bao, X.Y., Jeff, S.: Characterization of Brillouin fiber generator and amplifier for optimized working condition of distributed sensors. J. Opt. Fiber Technol. **15**, 304–309 (2009)
4. Yang, A.Y., Wu, X.Y., Qiao, Y.J., Sun, Y.N.: Bit-rate adaptive optical performance monitoring method for fiber communication systems. J. Opt. Commun. **284**, 436–440 (2011)
5. Hao, J.W.: A signal generator of pulse width modulation based on FPGA. J. Comput. Eng. **39**, 260–264 (2013) (in Chinese)
6. Wang, L., Xu, L.S., Yao, Y., Song, D.: FPGA-based design and implementation of arterial pulse wave generator using piecewise Gaussian-Cosine fitting. Comput. Biol. Med. **59**, 142–151 (2015)
7. Yin, C.Q., Tian, H., Li, Y.Q., Huang, H.J.: Design of ultrahigh-speed square wave pulser in a Brillouin sensing system. J. Laser Technol. **38**, 679–683 (2014) (in Chinese)
8. Zhang, S.J., Huang, Z.: Signal integrity and its implementation scheme in high speed digital system and implementation scheme. J. App. Electron. Tech. **11**, 31–33 (2002) (in Chinese)
9. Dahlen, P.E.: Enhancing signal integrity design methodologies utilizing discrete frequency domain techniques. Ph.D., University of Minnesota, pp. 23–46 (2014)
10. Sen, S., Mukhopadhyay, S.: Reduction of V_π voltage of an electro-optic modulator by jointly using oblique end cutting and multi-rotation. Opt. Laser Technol. **59**, 19–23 (2014)
11. Wegert, M., Schwochert, D., Lüdge, K.: Integrated quantum-dot laser devices: modulation stability with electro-optic modulator. J. Opt. Quant. Electron. **46**, 1337–1344 (2014)

The Influence Analysis of Pseudorandom Number Generators and Low Discrepancy Sequences for the Family of Compact Genetic Algorithms: Search Behavior Research from Outside Causes to Internal Causes

Hong-Guang Zhang, Bi-Hua Tang and Kai-Ming Liu

Abstract According to probability vectors, compact genetic algorithms (CGAs) generate new individuals by using pseudorandom number generators (PRNGs) or low discrepancy sequences (LDSs). These new generated individuals are the only factors which determine the search directions in CGAs. Therefore, we experimentally study the relationship between probability vectors and PRNGs (or LDSs). Moreover, we primarily investigate the influence analysis of PRNGs and LDSs for the effectiveness and efficiency of the family of CGAs by using analysis of variance (ANOVA), success rate and success performance. According to experimental results, we provide conclusive evidence to suggest using PRNGs (or LDSs) for CGAs. In essence, the frameworks of CGAs and the update method of probability vectors of CGAs are the internal causes that determine the performance of CGAs for different PRNGs and LDSs.

Keywords Influence analysis · Pseudorandom number generators · Low discrepancy sequences · Compact genetic algorithms · Convergence analysis

H.-G. Zhang (✉) · B.-H. Tang · K.-M. Liu
School of Electronic Engineering, Beijing Key Laboratory of Work Safety Intelligent Monitoring, Beijing University of Posts and Telecommunications, Beijing 100876, China
e-mail: hongguang-zhang@bupt.edu.cn

B.-H. Tang
e-mail: bhtang@bupt.edu.cn

K.-M. Liu
e-mail: kmliu@bupt.edu.cn

1   Introduction

Massive practices have proved that evolutionary algorithms (EAs) are fit for solving specific-application optimization problems. EAs use pseudorandom number generators (PRNGs) or low discrepancy sequences (LDSs) to implement the heuristic search of each individual. However, the heuristic search of each individual corresponds to the move of each individual in solution space. More important is the problem that PRNGs and LDSs influence the behavior characteristics of heuristic searches of individuals in EAs.

In fact, the concluding remarks of different papers about the research of EAs for random number generators are not contradictory to each other. Generally, there are different research conditions in the related papers, such as real-coding genetic algorithm or binary-coding genetic algorithm, the difference of the features of test problems, and the implementation difference for the same PRNG or LDS.

Meysenburg and Foster [1] examined the effect of the PRNG quality on the performance of genetic programming. According to experimental results, they find no statistically significant evidence to support the notion that higher quality PRNGs can improve the performance of genetic programming [1]. Meysenburg et al. [2] made a detailed study of the phenomena that the poor quality PRNGs can drive genetic algorithms to superior performance for certain problems. These research results show no direct correlation between improving the quality of PRNG and improving the performance of genetic algorithm [2]. In some cases, the better PRNGs can cause the worse performance of genetic algorithm [2]. Tirronen et al. [3] empirically evaluated different PRNGs in the context of differential evolution for the real-valued optimization. This paper presents strong evidence that it is safe to assume near identical performance from most PRNGs for differential evolution [3]. Maucher et al. [4] experimentally investigated the robustness of simulated annealing and genetic algorithm, when using pseudorandom numbers with limited quality. For example, they used Mersenne twister (MT) as a PRNG, and artificially reduced the period length to values ranging from 1009 to 512009. The study shows that the increased diversity of the pseudorandom numbers leads to the improved results for simulated annealing and genetic algorithm [4]. Interestingly, they find that when using the diverse quasi-random sequences, genetic algorithm outperforms its own results using quantum random number generator device [4]. Numbers from the quantum random number generator device are regarded as truly random bits, which are independent and uniformly distributed. Cantú-Paz [5] studied the effect of PRNGs on the simple genetic algorithm for trap problem, and used ablation experiments to isolate the stochastic components of the simple genetic algorithm. Experiment results show that the choice of PRNGs can cause large variations in the performance of the simple genetic algorithm [5]. On the other hand, these ablation experiments suggest that PRNGs, used to initialize the population, are critical to the performance of the simple genetic algorithm for trap problem [5]. Krömer et al. [6] provided an initial empirical comparison of genetic algorithm, differential

evolution, and particle swarm optimization using different PRNGs and deterministic chaos based generators (DCh-based generators). Experimental results confirm that the choice of the source of stochasticity has an effect on the results of the algorithm [6]. However, this effect appears to be algorithm dependent and problem dependent. On the other hand, simulation results show that the re-mapped logistic equation (a DCh-based generator) seems to be a viable computationally inexpensive method [6]. Cárdenas-Montes et al. [7] dealt with an empirical study of the impact of the change of PRNGs [MT and GNU Compiler Collection rand()] in the performance of the four EAs: particle swarm optimization, differential evolution, genetic algorithm, and firefly algorithm. They use Wilcoxon signed-rank test to compare the sensitivity of these four EAs to the choice of PRNGs [7]. Experimental results show that each EA has a very different sensitivity in relation to the choice of PRNGs [7]. Omran et al. [8] investigated the use of LDSs to generate an initial population for population-based optimization algorithms. They use non-parametric statistical tests, two popular LDSs (Sobol LDS and Halton LDS), and two well-known population-based optimization methods (particle swarm optimization and differential evolution) [8]. According to simulation results, there is no evidence that the use of LDSs improves the performance of population-based search methods [8]. Monica et al. [9] proposed a new artificial bee colony (ABC) based on Sobol LDS to guide the movement of scout bee and protect the population diversity. The performance of this proposed ABC is compared to that of basic ABC for several benchmark test problems. Experimental results show that this proposed ABC converges faster than basic ABC [9].

There is sixteen years history since 1999 when the original compact genetic algorithm [10] was proposed. The original intention of compact genetic algorithms (CGAs) [10, 11] is to extend the application scope of EAs from the large computers to the tiny embedded microcomputers. In every year, many very serious and fresh researches are investigated for solving different interdisciplinary problems, such as the works in [12–15] from 2014 to 2015. The family of CGAs grows fast in these sixteen years.

Almost all CGAs in the family of CGAs are based on the architecture of probability vectors. By using PRNGs and LDSs, CGAs use probability vectors to generate the new individuals. The most important is that these new generated individuals are the only factor to determine the search directions in CGAs. Therefore, these probability vectors are the most important factors in the family of CGAs.

In this paper, we experimentally study the relations between probability vectors and PRNGs (or LDSs), and provide the conclusive evidence and suggestion of using PRNGs (or LDSs) for CGAs. The main works include the influence analysis of PRNGs and LDSs for the effectiveness and efficiency of the family of CGAs. The details are given as follows. Firstly, different from most of genetic algorithms, CGAs use probability vectors instead of the population. Therefore, a research work is to find the relationships between random number generators (PRNGs or LDSs) and the effectiveness of the convergence performance of CGAs. Secondly, the

efficiency of EAs indirectly represents the computation time. The computation time of EAs is proportional to the number of function evaluations. Therefore, the number of function evaluations is frequently used to compare the efficiency of EAs. Using the number of function evaluations as the evaluation criterion, a research work is to give the clear conclusion that whether there is a perfect random number generator (a PRNG or LDS) to improve the efficiency of CGAs or not.

2 Experimental Scheme

2.1 Compact Genetic Algorithms

In essence, each CGA is only characterized by its definition style and the update method of probability vector. We select three CGAs in different styles. Three selected CGAs are the original CGA (CGA) [10], the persistent elitist compact genetic algorithm (peCGA) [11], and the nonpersistent elitist compact genetic algorithm (neCGA) [11]. We summarize the features of three selected CGAs, as shown in Table 1.

2.2 Test Problems

Generally, CGAs run on embedded microcomputers with limited memory, and so on. In essence, CGAs are fit for the binary optimization problems with low complexity. Therefore, we select one max problem, trap problem [11 copies of the 9-bit basic function of trap problem, which is defined by (1)], deceptive problem (33 copies of the 3-bit basic function of deceptive problem), and MAX SAT problem (file_rms_cnf_L2_V100_C100_0.cnf from SAT 2006 [16]) as test problems. These test problems have different features, such as unimodal, multimodal, linear,

Table 1 Summary of selected compact genetic algorithms

Algorithm	Feature
CGA	CGA uses the random walk model. Generate two new individuals according to probability vector in one generation, and let them compete
peCGA	peCGA persistently uses elitist mechanism. Generate a new individual according to probability vector in one generation, and update the elite individual by using a winner persistently
neCGA	neCGA nonpersistently uses elitist mechanism. Generate a new individual according to probability vector in one generation, and update the elite individual by using a winner nonpersistently

Table 2 Summary the features of the selected test problems

Problem	Feature	Mapping from search space to solution space
One max	Unimodal	Linear, one-to-one
Trap	Trap	Nonlinear, many-to-one
Deceptive	Deceptive	Nonlinear, many-to-one
MAX SAT	Multimodal	Nonlinear, many-to-one

nonlinear, one-to-one, many-to-one, and so on. The features of the selected test problems are summarized, as shown in Table 2.

In this paper, we use the normalized fitness of each test problem.

$$f(X) = \begin{cases} 9 & u = 9 \\ 8 - u & u \neq 9 \end{cases} \tag{1}$$

where X is the 9-bit binary substring, and u is the number of ones in X.

2.3 Pseudorandom Number Generators and Low Discrepancy Sequences

We select Mersenne twister (MT) [17], Multiplicative congruential generator (MCG) [18], Sobol LDS (Sobol) [19] and Halton LDS (Halton) [20] as random number generators. Mersenne twister and Multiplicative congruential generator are the representative algorithms of PRNGs. We have used the built-in functions provided by Matlab R2012a (7.14.0.739) for selected PRNGs and LDSs. LDSs skip 1000 initial points to guarantee the quality of LDSs. Note that LDSs do not have a seed value. Therefore, we randomly discard zero or one point for selected LDSs after generating a random number matrix in one generation. These random number matrices are used to generate new individuals.

The sequences, generated by PRNGs, are not truly random. In fact, PRNGs are deterministic algorithms for generating pseudorandom number sequences by using the PRNGs' seed (initial values of PRNGs). On the other hand, LDSs are designed specifically to generate number sequences as uniformly as possible. In mathematics, lower discrepancy is the unique feature of LDSs. To demonstrate the actual distributions of selected PRNGs and LDSs, the actual distributions of selected PRNGs and LDSs for two new generated individuals in CGA are given as an example in Fig. 1. As shown in Fig. 1, the distributions of LDSs are almost uniform, and own lower discrepancy than PRNGs. On the other hand, there are 5000 points in Fig. 1a–d, which correspond to 50 times of generating two new generated individuals with 100-bit coding in the simulations.
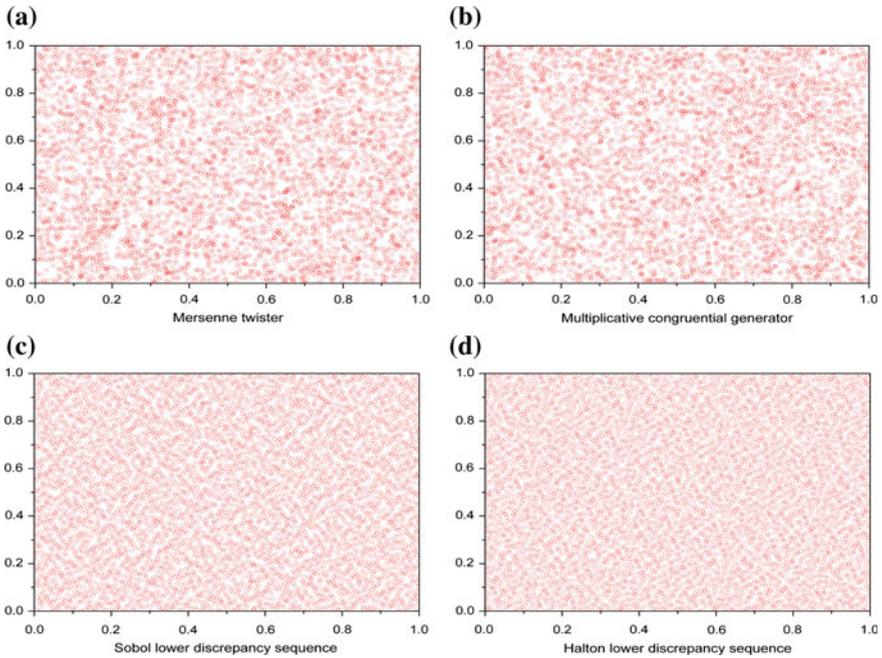
Fig. 1 The actual distributions of selected pseudorandom number generators and low discrepancy sequences for two new generated individuals in CGA

2.4 Objectives and Methods

Experimental objectives and methods were given in Table 3. We executed 1000 independent experiments for each kind of experiments. In each kind of experiments, each selected PRNG or LDS was regarded as an independent variable, and the

Table 3 Experimental objectives and methods

Section	Objective	Method
Section 3.1	To evaluate the influence of selected PRNGs and LDSs for the effectiveness of CGAs	Stop when the number of function evaluations is equal to 2000. We use ANOVA as the experimental method
Section 3.2	To evaluate the influence of selected PRNGs and LDSs for the efficiency of CGAs	Stop when the best fitness is greater than fitness threshold, or the number of function evaluations is equal to 2000. The fitness threshold values are 0.75, 0.80, 0.85, and 0.90 for Trap, Deceptive, One Max, and MAX SAT problems. These fitness threshold values are obtained by experiments. Evaluation criteria are success rate and success performance [see (2) and (3)]

performance of each CGA was regarded as the dependent variable. We used ANOVA, success rate, and success performance to evaluate the influence of selected PRNGs and LDSs for the effectiveness and efficiency of the family of CGAs.

Parameter settings of CGA, peCGA, and neCGA were given as follows. The population sizes in CGA, peCGA, and neCGA were all equal to 50. These settings of population sizes referred to Figs. 3 and 4 in [10] and Fig. 4 in [11]. The allowable length of inheritance η in neCGA was equal to 5 (0.1 × population size). This setting referred to [11].

3 Experimental Results and Discussion

3.1 Compare Effectiveness

To evaluate the influence of selected PRNGs and LDSs for the effectiveness of CGAs, we used ANOVA to compare the means of the best fitness of each CGA for the selected PRNGs and LDSs. The null hypothesis of ANOVA is that the means of the best fitness of each CGA for selected PRNGs and LDSs (MT, MCG, Sobol and Halton) are equal. The alternative hypothesis is that the means of the best fitness of each CGA for the selected PRNGs and LDSs are not equal.

If p-value is equal to or less than 0.05 (a 0.05 level of significance), it casts doubt on the null hypothesis. As shown in Table 4, p-values were all less than 0.05. These results mean that at least one sample mean is significantly different than the other sample means. These phenomena indicate that different PRNGs or LDSs have a direct effect on the effectiveness of each CGA.

The unit of generating new individuals in the frameworks of CGA, peCGA, and neCGA is the only step of using a PRNG or LDS. The new generated individuals determine the search directions of each CGA. Therefore, PRNGs or LDSs are the only factor that is strongly correlated with the search directions. These strongly correlated relationships are in accord with the above conclusion that different PRNGs or LDSs have a direct effect on the effectiveness of each CGA.

3.2 Compare Efficiency

To evaluate the influence of selected PRNGs and LDSs for the efficiency of CGAs, we used success rate (SR) and success performance (SP) as the evaluation criteria. SR [21] was given by (2) to represent the ability of an algorithm to satisfy the fitness threshold of stop condition. SP [21] was given by (3) to represent the performance of the convergence speed of an algorithm for successful runs.

Table 4 Execute 1000 independent runs, and stop condition is 2000 function evaluations

Problem	PRNG/LDS	CGA		peCGA		neCGA	
		MBF	*p*-value	*MBF*	*p*-value	*MBF*	*p*-value
One max	MT	0.9991	0	0.8441	0	0.9320	5.50e−227
	MCG	0.9995		0.8442		0.9328	
	Sobol	0.9856		0.9006		0.8845	
	Halton	0.9990		0.9280		0.9125	
Trap	MT	0.8882	0	0.7288	0	0.8194	1.04e−275
	MCG	0.8883		0.7278		0.8193	
	Sobol	0.8749		0.7763		0.7515	
	Halton	0.8884		0.8142		0.7941	
Deceptive	MT	0.9372	0	0.8800	5.04e−68	0.9073	0
	MCG	0.9372		0.8798		0.9061	
	Sobol	0.9210		0.8587		0.8047	
	Halton	0.8956		0.8656		0.8150	
MAX SAT	MT	0.9806	2.27e−163	0.9366	0	0.9565	0
	MCG	0.9806		0.9372		0.9574	
	Sobol	0.9822		0.9071		0.8856	
	Halton	0.9698		0.9124		0.8955	

Note To evaluate the influence of selected PRNGs and LDSs for the effectiveness of CGAs, compare the difference of MBF and *p*-value of ANOVA, where MBF represents the mean of the best fitness. To assess the statistical significance of selected PRNGs and LDSs for CGAs, we compute *p*-value by performing ANOVA, and use a 0.05 level of significance

$$SR = \frac{N_s}{N_t} \tag{2}$$

$$SP = mean(the\ number\ of\ function\ evaluations\ for\ successful\ runs)\frac{N_t}{N_s} \tag{3}$$

where N_s was the number of successful runs, and N_t was the number of total runs ($N_t = 1000$ in this paper). A successful run was a run during which the algorithm found a feasible solution, and this feasible solution satisfied the fitness threshold of stop condition. *mean*() returned the mean of values.

We discussed the influence of selected PRNGs for the efficiencies of CGAs and the influence of selected LDSs for the efficiencies of CGAs, respectively. (I) Experimental results of using MT and MCG were in bold font, as shown in Table 5. SR and SP by using MT were very similar to that of MCG for CGA, peCGA, and neCGA. These results indicate that when high qualities of PRNGs, such as the period length, discrepancy, and so on, are guaranteed, there is little influence of selected PRNGs for the efficiencies of CGAs. (II) However, there were obvious difference of the efficiencies of CGAs between Sobol and Halton, as shown

Table 5 Execute 1000 independent runs, and stop condition is to satisfy fitness threshold or 2000 function evaluations

Problem	PRNG/LDS	CGA		peCGA		neCGA	
		SR	*SP*	*SR*	*SP*	*SR*	*SP*
One max	MT	**1**	**523.800**	**0.487**	**678.065**	**1**	**294.200**
	MCG	**1**	**523.436**	**0.454**	**728.415**	**1**	**293.728**
	Sobol	1	1074.322	0.893	306.889	0.785	492.271
	Halton	1	743.952	0.959	262.272	0.953	388.869
Trap	MT	**1**	**572.586**	**0.221**	**1645.667**	**1**	**330.404**
	MCG	**1**	**575.780**	**0.22**	**1632.541**	**1**	**333.079**
	Sobol	1	1156.356	0.679	458.055	0.564	735.297
	Halton	1	757.934	0.922	300.312	0.820	484.325
Deceptive	MT	**1**	**665.914**	**0.993**	**181.844**	**1**	**326.716**
	MCG	**1**	**673.038**	**0.995**	**181.008**	**1**	**326.346**
	Sobol	1	933.356	0.932	327.815	0.583	758.401
	Halton	1	421.468	0.944	307.212	0.598	754.485
MAX SAT	MT	**1**	**319.452**	**0.997**	**132.307**	**1**	**177.907**
	MCG	**1**	**321.200**	**0.992**	**133.835**	**0.999**	**180.419**
	Sobol	1	364.152	0.758	392.470	0.320	1188.086
	Halton	1	546.248	0.850	337.538	0.574	619.757

Note To evaluate the influence of selected PRNGs and LDSs for the efficiency of CGAs, compare success rate (*SR*) and success performance (*SP*), where *SR* and *SP* are defined by using (2) and (3), respectively

in Table 5. These results indicate that the differences between Sobol and Halton have obvious effect on the efficiencies of CGAs.

On the other hand, it cannot simultaneously improve the efficiencies of CGA, peCGA, and neCGA by only using a PRNG or LDS. As shown in Table 5, the consistency of each PRNG (or LDS) for all of CGAs were totally different. These experimental results also demonstrated that there was not a perfect random number generator (such as MT, MCG, Sobol and Halton) that improved the efficiencies of all of CGAs for all test problems.

In essence, there is an obvious difference of the mapping relationship from search space to solution space between binary-coding test problems and real-coding test problems, such as Hamming cliff. Moreover, there are also obvious differences of fitness landscapes between different binary-coding test problems. It is feasible to find an excellent random number generator (a PRNG or LDS) that improves the efficiency of one CGA for one kind of test problems. However, it is not feasible to find a perfect random number generator (a PRNG or LDS) that improves the efficiency of one CGA for all kinds of test problems.

3.3 Discussion

We discussed the relationships among the features of the data sets, the performance of CGAs, and the random number generators (PRNGs and LDSs) in this section.

As shown in Tables 4 and 5, PRNGs were not better than LDSs only for the results of peCGA of One Max and Trap problems. We further synthesized and analyzed the results of peCGA of One Max and Trap problems. In fact, these above results of peCGA were not very satisfactory. We used Trap problem as an example, and compared peCGA with CGA. As shown in (1), the maximum fitness value of local optimal solution of Trap problem is equal to 8/9 (8/9 ≈ 0.8889). As shown in Table 4, CGA obtained the maximum fitness value of local optimal solution. But, peCGA did not obtain the maximum fitness value of local optimal solution. According to the contrast of Trap problem between peCGA and CGA in Table 4, we can also conclude that the results of peCGA in Table 5 were not very satisfactory. We believe that the elitist mechanism for updating winner in peCGA leads to these above results. In essence, this elitist mechanism, which peCGA persistently update the elite individual by using a winner, improves the performance of the efficiency of peCGA. However, this elitist mechanism simultaneously influences the performance of the effectiveness of peCGA. The no-free-lunch-theorem is always there.

On the other hand, PRNGs were better than or similar to LDSs for the other results. This indicates that PRNGs are more suitable for CGAs. Why LDSs, with almost uniform distribution, cannot improve the performance of CGAs. Our analysis is given as follows.

1. There are three important factors for this problem. These three factors are the distribution of random number generators (PRNGs and LDSs), the fitness functions, and the distribution of sample points in the solution space.
2. Generally, when the distribution of sample points in the solution space is uniform, there are much more opportunities to improve the performance of CGAs. Although the distributions of LDSs are almost uniform, the distribution of sample points in the solution space may also be not uniform. The fitness functions are the mapping relationships from the searching space to the solution space. The multimodal, nonlinear, many-to-one fitness functions can totally change the distributions of sample points in the solution space.
3. On the other hand, LDSs own lower discrepancy than PRNGs. This means the generation manner of LDSs is very precise. But, this precise generation manner of LDSs also means that there is a regular law to guide the generation of random numbers when we use LDSs. The regular law is not a good thing for random searching and heuristic searching, especially for the multimodal, nonlinear, many-to-one fitness functions. On the contrary, the generation manners of PRNGs are simple and not precise. This simple generation manners of PRNGs increase the freedom of the search behavior of the sample points in the solution space. This also means PRNGs own the better robustness than LDS, especially for the multimodal, nonlinear, many-to-one fitness functions.

4 Conclusion

In this paper, we regarded each PRNG or LDS as an independent variable, and regarded the performance of each CGA as the dependent variable. Based on experimental results, we conclude that the frameworks of CGAs and the update method of probability vectors of CGAs are the internal causes that determine the performance of CGAs for different PRNGs and LDSs.

The concluding remarks of each kind of the main works are explained as follows.

Firstly, we used ANOVA to evaluate the influence of selected PRNGs and LDSs for the effectiveness of CGAs in Sect. 3.1. All of p-values in Table 4 were less than 0.05 (a 0.05 level of significance). This means that different PRNGs or LDSs have a direct effect on the effectiveness of CGA, peCGA and neCGA. In essence, the effectiveness of each CGA is all strongly correlated with PRNGs or LDSs.

Secondly, according to experimental results in Sect. 3.2, the research problem of the efficiencies of CGAs is the algorithm dependent, the test problem dependent, and the random number generator dependent. In essence, this research problem does not belong to the problem of a single independent variable, such as the problem of a single random number generator independent variable. It is not feasible to find a perfect random number generator (a PRNG or LDS) that improves the efficiency of one CGA for all kinds of problems.

On the other hand, when we use a PRNG or LDS as a random number generator for CGAs, we suggest that the period length, discrepancy, etc. of PRNGs and LDSs should be guaranteed. Adopting the above suggestion can avoid the "lucky" accidents of experimental results, especially for the research of the efficiencies of CGAs.

On the whole, the influence analysis from outside causes (random number generators, such as PRNGs and LDSs) to internal causes (such as the frameworks of CGAs and the update method of probability vectors of CGAs) is a feasible approach to find the essential laws of using random number generators in CGAs.

References

1. Meysenburg, M.M., Foster, J.A.: Random generator quality and GP performance. In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 1121–1126. Morgan Kaufmann, Orlando (1999)
2. Meysenburg, M.M., Hoelting, D., Mcelvain, D., Foster, J.A.: How random generator quality impacts genetic algorithm performance. In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 480–483. Morgan Kaufmann, San Francisco (2002)

3. Tirronen, V., Äyrämö, S., Weber, M.: Study on the effects of pseudorandom generation quality on the performance of differential evolution. In: Dobnikar, A., Lotrič, U., Šter, B. (eds.) ICANNGA 2007. LNCS, vol. 4431, pp. 361–370. Springer, Heidelberg (2011)

4. Maucher, M., Schöning, U., Kestler, H.A.: Search heuristics and the influence of non-perfect randomness: examining genetic algorithms and simulated annealing. Comput. Stat. **26**, 303–319 (2011)

5. Cantú-Paz, E.: On random numbers and the performance of genetic algorithms. In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 311–318. Morgan Kaufmann, San Francisco (2002)

6. Krömer, P., Zelinka, I., Snášel, V.: Behaviour of pseudo-random and chaotic sources of stochasticity in nature-inspired optimization methods. Soft. Comput. **18**, 619–629 (2014)

7. Cárdenas-Montes, M., Vega-Rodríguez, M.A., Gómez-Iglesias, A.: Sensitiveness of evolutionary algorithms to the random number generator. In: Dobnikar, A., Lotrič, U., Šter, B. (eds.) ICANNGA 2007. LNCS, vol. 4431, pp. 371–380. Springer, Heidelberg (2011)

8. Omran, M.G.H., Al-Sharhan, S., Salman, A., Clerc, M.: Studying the effect of using low-discrepancy sequences to initialize population-based optimization algorithms. Comput. Optim. Appl. **56**, 457–480 (2013)

9. Monica, T., Rajasekhar, A., Pant, M., Abraham, A.: Enhancing the local exploration capabilities of artificial bee colony using low discrepancy Sobol sequence. In: Aluru, S., Bandyopadhyay, S., Catalyurek, U.V., Dubhashi, D.P., Jones, P.H., Parashar, M., Schmidt, B. (eds.) IC3 2009. CCIS, vol. 40, pp. 158–168. Springer, Heidelberg (2009)

10. Harik, G.R., Lobo, F.G., Goldberg, D.E.: The compact genetic algorithm. IEEE Trans. Evol. Comput. **3**, 287–297 (1999)

11. Chang, W.A., Ramakrishna, R.S.: Elitism-based compact genetic algorithms. IEEE Trans. Evol. Comput. **7**, 367–385 (2003)

12. Jiau, M.K., Huang, S.C.: Services-oriented computing using the compact genetic algorithm for solving the carpool services problem. IEEE Trans. Intell. Transp. Syst. **16**, 2711–2722 (2015)

13. Satman, M.H., Diyarbakirlioglu, E.: Reducing errors-in-variables bias in linear regression using compact genetic algorithms. J. Stat. Comput. Simul. **85**, 3216–3235 (2015)

14. Soares, A.S., de Lima, T.W., Soares, F.A.A.M.N., Coelho, C.J., Federson, F.M., Delbem, A. C.B., Van Baalen, J.: Mutation-based compact genetic algorithm for spectroscopy variable selection in determining protein concentration in wheat grain. Electron. Lett. **50**, 932–934 (2014)

15. Al-Dabbagh, R.D., Kinsheel, A., Baba, M.S., Mekhilef, S.: A combined compact genetic algorithm and local search method for optimizing the ARMA (1,1) model of a likelihood estimator. Sci. Asia **40**, 78–86 (2014)

16. SAT 2006: Artificial Intelligence Research Institute—Spanish National Research Council. http://www.iiia.csic.es/conferences/maxsat06/ms06-bench.tgz

17. Matsumoto, M., Nishimura, T.: Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul. **8**, 3–30 (1998)

18. Park, S.K., Miller, K.W.: Random number generators: good ones are hard to find. Commun. ACM **31**, 1192–1201 (1988)

19. Bratley, P., Fox, B.L.: Algorithm 659 implementing Sobol's quasirandom sequence generator. ACM Trans. Math. Softw. **14**, 88–100 (1988)

20. Kocis, L., Whiten, W.J.: Computational investigations of low-discrepancy sequences. ACM Trans. Math. Softw. **23**, 266–294 (1997)

21. Liang, J.J., Runarsson, T.P., Mezura-Montes, E., Clerc, M., Suganthan, P.N., Coello, C.C.A., Deb, K.: Problem definitions and evaluation criteria for the CEC 2006 special session on constrained real-parameter optimization. http://www.researchgate.net/publication/216301032_Problem_definitions_and_evaluation_criteria_for_the_cec_2006_special_session_on_constrained_real-parameter_optimization (2006)

TLS Channel Implementation for ONOS's East/West-Bound Communication

Jun Huy Lam, Sang-Gon Lee, Hoon-Jae Lee and Yustus Eko Oktian

Abstract East/West-bound communication is the communication channel which exists only in the distributed software-defined network (SDN) that governs the communication within the control plane of the network. Unlike most SDN projects, that have neglected the security of, and have achieved east/west-bound communication with the assistance of a network application on the management plane through the representational state transfer (REST) application program interface (API), the Open Networking Operating System (ONOS) was designed with distributed functionality as one of its core features. Hence, it supports both the native intra-cluster and the network application's inter-cluster communication. In this paper, the transport layer security (TLS) channel for ONOS's native east/west-bound communication was implemented and the performance impact was evaluated.

Keywords East/West-bound security · Distributed SDN · Intra-cluster security · Inter-cluster security · Intra-domain security · Inter-domain security · Distributed SDN security · Identity-based cryptography

J.H. Lam · S.-G. Lee (✉) · H.-J. Lee · Y.E. Oktian
Dongseo Univesity, Busan, South Korea
e-mail: nok60@dongseo.ac.kr

J.H. Lam
e-mail: timljh@msn.com

H.-J. Lee
e-mail: hjlee@dongseo.ac.kr

Y.E. Oktian
e-mail: yustus.oktian@gmail.com

1 Introduction

The separation of the network control and data planes created the SDN in which a SDN controller centralizes the network control plane and manages the network data plane. Introduction of SDN also provides a platform for network applications of the management plane to make use of the information gathered by the SDN controller. These information allows network applications to have a network-wide view and hence capable of efficient resource management [1]. Some use cases can be found [2].

The distributed SDN introduces a new communication channel: east/west-bound communication. It is used for communication within the network control plane, i.e., communication between the SDN controllers or data stores. The existence of this particular communication channel is what differentiates the conventional centralized SDN and the distributed SDN. Unlike the southbound communication that manages communication between the network control plane and data plane or between the SDN controllers and the network switches, east/west-bound communication have neither an agreed-upon protocol nor the security for it. However, the security for this communication channel is particularly important for the distributed SDN. It ensures that no malicious SDN controllers are snooping for network information or even driving the network.

Most SDN projects have proposed or implemented southbound or northbound protection through Open Flow's optional TLS mode [3], RESTCONF (REST Configuration protocol) and Neutron with TLS-enabled Hypertext Transfer Protocol (HTTPS) [3], REST API with TLS [4], secure enhancement in wireless mobile networks [5] or other southbound or northbound protocols. However, none have attempted to secure the east/west-bound communication in which the core of the distributed SDN resides. The control plane of the network is spread across all available controllers within the distributed SDN. In order to ensure that the network information which will be synchronized between these controllers are secure, the east/west-bound communication has to be protected.

In this paper, the TLS channel was implemented to secure the east/west-bound communication of the ONOS project. To the best of our knowledge, this is the first work that secures the east/west-bound communication with a proven protocol such as TLS. Our contributions are as listed below:

1. We described the east/west-bound communication and why it needs to be protected.
2. We implemented the TLS mode for ONOS's intra-cluster communication and evaluated the performance difference between the unsecured mode and that protected by TLS. We concluded the implementation and result of our findings.
3. The source code of the implementation was accepted into the ONOS master code after the code review process [6].

2 Security of the East/West-Bound Communication

2.1 East/West-Bound Communication

In the distributed SDN, there are two types of east/west-bound communication: intra-cluster communication and inter-cluster communication. As per their names, intra-cluster communication is the communication between controllers that belong within the same cluster/domain, whereas inter-cluster communication is the communication between controllers that belong to different clusters/domains.

OpenDaylight (ODL) implemented inter-cluster communication as a network application (ODL-SDN interface, ODL-SDNi) [7] on the management plane through REST API but does not provide intra-cluster communication, as it was not initially designed with the distributed primitives.

ONOS implemented both intra-cluster and inter-cluster communication as shown in Fig. 1. The inter-cluster communication of ONOS is similar to that of ODL's in the form of network application (Inter-Cluster ONOS Network Application, ICONA) [8]. In addition, ONOS also achieved intra-cluster communication within the control plane itself [9]. Therefore, in theory, for intra-cluster communication, no performance or latency penalty goes through the REST API and the management plane.

Most SDN projects have implemented east/west-bound communication through the management plane and have emphasized on the information that can be retrieved by the network application through REST APIs. They protect the communication channel by restricting the access of the network application to certain network information. By doing so, it also limits the functionality of the east/west-bound communication channel.
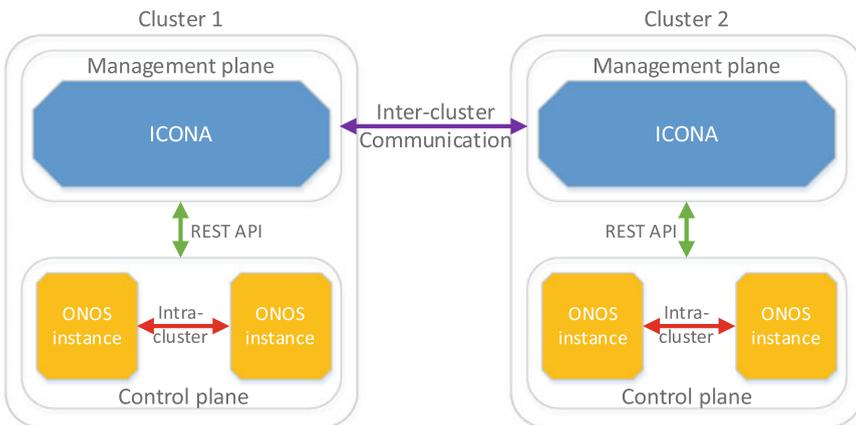

Fig. 1 The two types of east/west-bound communication in ONOS

On the other hand, intra-cluster communication allows full access to the network information and hence, it is even more crucial to protect this particular communication channel in which unrestricted network information will be exchanged. Therefore, in this paper, security for the intra-cluster communication was implemented and its performance impact evaluated.

2.2 *Related Work*

Santos et al. [10] proposed the application of identity-based cryptography (IBC) to secure the communications of Master Controller–Secondary Controller (MC-SC), SC-SC, the client-side and server-side within their framework. In their proposal, the IBC protocol was based on the protocol established by Sakai et al. [11]. However, the IBC protocol was not finalized or adopted by the industry. According to research conducted by Chen et al. [12] and Chatterjee et al. [13], the Type 1 pairing used by the IBC protocol of Santos et al. is suitable for security levels of up to 80 bits; for security levels higher than 80 bits, the performance will degrade significantly. Additionally, the performance impact was not evaluated for acceptance by the industry. Unlike IBC, TLS was well-proven and widely adopted by the industry. Hence, this paper employs the well-proven and widely-accepted TLS protocol so that it can be adopted by the industry.

3 Performance Evaluations

The system setup runs on a desktop computer with an Intel Core i5 760, quad-core processor in which each core runs at 2.8 GHz with multi-threading capability, 10 GB of memory and a 120 GB solid state drive that has its transfer speed limited by the Serial Advanced Technology Attachment 2 (SATA2). The system software runs on Linux Mint 17.2 with three LXC (Linux Container) Ubuntu containers. One ONOS instance is running on each container.

The required java optional parameters were already listed within the script file, onos/tools/package/bin/onos-service, but the default setting for the security of east/west-bound communication is unprotected. In order to enable the TLS mode for the intra-cluster east/west-bound communication in ONOS, the java optional parameters must be set for the script file and the java keystore and truststore will be required. Java keytool can be used to initiate both the keystore and truststore; the location of the java keystore and trustsore and their respective passwords must be set in the onos-service accordingly.

The source code was merged into the ONOS master code [6] and hence can be checked out from the ONOS master code at https://gerrit.onosproject.org/onos. For a detailed guide of ONOS's source code checkout and setup, please refer to the ONOS Wiki guide [14].

The performance of ONOS's east/west-bound communication channel was evaluated with an ONOS application called "messagingperf" which measures the number of messages that it attempted to send at one time through the intra-cluster east/west-bound communication as well as the number of messages that were sent through. The default configuration was used in the evaluations in which the application was assigned two sender threads, two receiver threads and enabled serialization. The performance difference between the TLS secure channel and open channel were evaluated. In order to obtain stable and more consistent results, the results were recorded after all ONOS instances were completely initiated, and an average of ten readings was recorded.

In Table 1, some results indicate that the ONOS instance actually completed sending more messages than it attempted to send. This is due to the ONOS instance actually completing the transfer that it attempted to send during the previous attempt in addition to the current attempt. Hence, the "completed readings" consist of all messages that were successfully sent at that particular time, which may also include messages sent during the previous attempt.

Table 1 Performance comparison between TLS-enabled and open-mode intra-cluster communication

	With TLS		Without TLS		Performance degradation	
	Attempted	Completed	Attempted	Completed	Attempted (%)	Completed (%)
Instance 1	24716	24715	25798	25796	−4.38	−4.37
	24533	24535	25793	25794	−5.14	−5.13
	23714	23713	25646	25647	−8.15	−8.16
	23825	23824	25378	25378	−6.52	−6.52
	23716	23716	25695	25695	−8.34	−8.34
	24188	24188	25970	25969	−7.37	−7.36
	24360	24362	24960	24961	−2.46	−2.46
	23512	23512	26235	26234	−11.58	−11.58
	24547	24545	25582	25582	−4.22	−4.22
	25410	25410	25236	25237	0.68	0.68
Average	**24252.1**	**24252**	**25629.3**	**25629.3**	**−5.68**	**−5.68**
Instance 2	24247	24246	26077	26076	−7.55	−7.55
	24674	24674	25692	25694	−4.13	−4.13
	23577	23577	25421	25419	−7.82	−7.81
	24801	24802	25527	25528	−2.93	−2.93
	24161	24161	26442	26441	−9.44	−9.44
	24223	24222	26198	26200	−8.15	−8.17
	24608	24608	25304	25303	−2.83	−2.82
	23775	23775	25909	25909	−8.98	−8.98
	24438	24438	25612	25611	−4.80	−4.80
	24499	24501	24920	24921	−1.72	−1.71

(continued)

Table 1 (continued)

	With TLS		Without TLS		Performance degradation	
	Attempted	Completed	Attempted	Completed	Attempted (%)	Completed (%)
Average	**24300.3**	**24300.4**	**25710.2**	**25710.2**	**−5.80**	**−5.80**
Instance 3	25050	25050	25163	25165	−0.45	−0.46
	25326	25327	25020	25019	1.21	1.22
	24951	24952	25218	25219	−1.07	−1.07
	24515	24514	24907	24907	−1.60	−1.60
	25758	25757	24980	24978	3.02	3.02
	24811	24813	25190	25192	−1.53	−1.53
	24781	24781	24932	24930	−0.61	−0.60
	25227	25226	25150	25152	0.31	0.29
	23773	23772	25204	25203	−6.02	−6.02
	25440	25440	24996	24996	1.75	1.75
Average	**24963.2**	**24963.2**	**25076**	**25076.1**	**−0.45**	**−0.45**

According to the results listed in Table 1, the performance degraded by approximately 5 % for ONOS instance 1 and instance 2, while the performance degradation of ONOS instance 3 is less than 1 %. This performance degradation is to be expected with the cryptography processes securing the channel, and is also somewhat insignificant with recorded values between 0.45 and 5.8 %. Hence, the security of the communication channel should not be ignored at such a small performance penalty.

4 Conclusions

The aim of the implementation is to secure the intra-cluster east/west-bound communication channel of ONOS and evaluate the subsequent performance degradation. The implementation of the TLS channel was also merged into the ONOS master code after the code review process and can be easily obtained, since ONOS is an open source project. The performance evaluation results indicate that the performance degradation brought about by TLS is minimal and hence, the performance degradation alone should not justify the negligence of the distributed SDN projects to secure the channel.

However, the primary issue of securing the east/west-bound or the southbound communication is not only in regards to the subsequent performance degradation; it may also be due to the tedious system setup that requires the network administrator to pre-install the public and private keys to all devices prior to deployment, and to re-install the key whenever it expires. Therefore, our future works will attempt to minimize the system setup by implementing an identity-based cryptography (IBC) protocol to secure the east/west-bound communication that is able to overcome

the shortcomings of Santos et al. [10]. By doing so, it can ease the system setup process in which the public key can be derived from the identity (ID) information of the device such as the controller ID, Media Access Control (MAC) address, Internet Protocol (IP) address, combinations of the IDs, and etc., while the private key will be generated by the PKG and distributed accordingly.

References

1. Ali, S.T., Sivaraman, V., Radford, A., et al.: A survey of securing networks using software defined networking. IEEE Trans. Reliab. **60**(3), 1086–1097 (2015)
2. Jacek, W., Thorsten, R., Khoa, T.D., et al.: SDN controller mechanisms for flexible and customized networking. Intl. J. Electron. Telecommun. **60**(4), 299–307 (2014)
3. OpenDaylight's Wiki. https://wiki.opendaylight.org/view/CrossProject:OpenDaylight_Security_Analysis
4. ONOS Project's Wiki. https://wiki.onosproject.org/pages/viewpage.action?pageId=4162614
5. Aaron, Y.D., Jon, C., Sasu, T., et al.: Software defined networking for security enhancement in wireless mobile networks. Comput. Netw. **66**, 94–101 (2014)
6. ONOS Project's Gerrit. https://gerrit.onosproject.org/#/c/4168/
7. OpenDaylight's Wiki. https://wiki.opendaylight.org/view/Project_Proposals:ODL-SDNi_App
8. Gerola, M., Santuari, M., Salvadori, E., et al.: ICONA: inter cluster ONOS network application. In: 1st IEEE Conference on Network Softwarization (NetSoft), pp. 1–2. IEEE, London (2015)
9. ONOS Project's Wiki. https://wiki.onosproject.org/display/ONOS/Cluster+Coordination
10. Santos, M.A.S., Nunes, B.A.A., Obraczka, K., et al.: Decentralizing SDN's Control Plane. IEEE Local Computer Networks (LCN), IEEE, Edmonton, Canada (2014)
11. Sakai, R., Kasahara, M.: ID based cryptosystems with pairing on elliptic curve. In: Cryptology ePrint Archive. Report 2003/054. http://eprint.iacr.org/ (2003)
12. Chen, L., Cheng, Z., Smart, N.P.: Identity-based key agreement protocols from pairings. Int. J. Inf. Secur. **6**(4), 213–241 (2007)
13. Chatterjee, S., Hankerson, D., Menezes, A.: On the efficiency and security of pairing-based protocols in the Type 1 and Type 4 settings. Arithmetic of finite fields. In: 3rd International Workshop WAIFI 2010, pp. 114–134. Springer, Berlin (2010)
14. ONOS Project's Wiki. https://wiki.onosproject.org/display/ONOS/ONOS+Wiki+Home

Towards Load Balance and Maintenance in a Structured P2P Network for Locality Sensitive Hashing

Da-Wei Liu and Zhi-Hua Yu

Abstract In this paper, we consider load balancing and maintenance of distributed similarity search systems using locality sensitive hashing (LSH) in a structured peer-to-peer (P2P) network based on Distributed Hashing Table (DHT). LSH has been proven efficient in K-Nearest Neighbor (KNN) search in high dimensions. Recently, a number of schemes have been proposed to implement LSH over DHT-based P2P systems to process distributed similarity searches. We provide an efficient structure using virtual nodes to manage the multi-dimensional LSH bucket space in DHT peers and the maintenance algorithm, which improves load balancing in comparison with state-of-the-art techniques such as the virtual node algorithm. Here, we demonstrate effectiveness of the proposed method by experiments.

Keywords Load balance · DHT · LSH · Virtual nodes · Distributed similarity search

1 Introduction

Performing similarity searches, especially in high dimensions, is a classical problem in many applications. A number of similarity search methods for high dimensional data have been proposed to enable effective similarity searches to keep pace with the rapid growth of information from text documents and multimedia content. Locality Sensitive Hashing (LSH) [1, 2] is an efficient hash-based approximate method, which assigns similar objects that are more likely to collide (hash to the

D.-W. Liu (✉) · Z.-H. Yu
Institue of Network Technology, Institute of Computing Technology, CAS, Yantai,
Shandong, People's Republic of China
e-mail: liudw@int-yt.com

Z.-H. Yu
e-mail: yzh@ict.ac.cn

Z.-H. Yu
Institute of Computing Technology, CAS, Beijing, People's Republic of China

same bucket in a hash table). In high dimensions, LSH has demonstrated better performance to perform similarity searches than other tree-based methods such as KD tree.

Distributed Hash Table (DHT) is a well-known approach to build structured Peer-to-Peer (P2P) networks. Various DHT-based P2P overlay algorithms have been proposed, such as Chord [3], Pastry [4] and Tapestry [5]. Chord uses consistent hashing to assign keys to Chord nodes. Consistent hashing maintains good scalability and lookup efficiency. However, the uniform distribution of keys makes Chord unable to support range queries, which require data locality. Recently, several schemes have focused on similarity search over high-dimensional data in structured P2P overlay networks [6–11]. These schemes design different indexing structures in DHT-based P2P networks to support complicated similarity search.

While Chord assumes that keys are uniformly distributed and the actual loads of the whole overlay network are often skewed, several load balancing schemes [12, 13] are designed to perform in DHT-based P2P networks. For example, Rao [12] uses the concept of virtual nodes, which transfers load between nodes to get perfect load balancing.

In this paper, we consider a DHT-based distributed similarity search system using LSH, which processes distributed similarity search by mapping LSH to DHT-identifier space. We discuss the load balancing problem in existing mapping methods, and present Chord-based structure using virtual nodes and maintenance algorithms to ensure fair load balance among nodes and to efficiently process dynamic nodes changes in overlay network.

2 Overview and Problem Statement

Here we briefly describe the approach of generating similarity search using LSH [6]. We apply locality-sensitive hashing functions based on p-stable distributions [1]. For each data point v, k independent hash functions of the form (1) are considered, where a is a d-dimensional vector whose elements are chosen independently from the Normal distribution. $W \in IR$, and B is chosen uniformly from [0, W].

$$h_{a,B}(v) = \left\lfloor \frac{a \cdot v + B}{W} \right\rfloor \tag{1}$$

Each hash function maps a *d*-dimensional data point onto the set of integers and the final result is a vector of length *k* of the form (2).

$$g(v) = (h_{a_1 B_1}(v), \ldots, h_{a_k B_k}(v)) \tag{2}$$

In order to process a similarity search in a distributed P2P network, these hash tables need to be mapped to the peer identifier space of DHT overlay network. The

linear mapping function ξ can be simply based on sum in the form (3), where b is the k-dimensional vector of integer. Each element in b represents a bucket label.

$$\xi_{sum}(b) = \sum_{i=1}^{k} b_i \tag{3}$$

Note that the mapping function based on the sum treats all bucket labels equally, which assigns buckets likely to hold similar data to the same peer. As described in [6], given all data points $v_1,...,v_M$, after two-level mapping, the result follows the normal distribution.

$$N\left(\frac{k}{2}, \frac{\sqrt{k\sum_i \|vi\|_2^2}}{W}\right) \tag{4}$$

The above distributed similarity search system (4) enables KNN search over LSH indices, which is different from LSH in centralized settings. Meanwhile, the predictable loads following the normal distribution make the allocating and processing problem less likely in DHT-based P2P networks.

As discussed above, the values generated by mapping function (3) follows a normal distribution. However, the peer identifier space of DHT-based overlay network is designed to follow uniform distribution. In the original Chord protocol [3], consistent hashing is used to assign keys to Chord nodes. Each node receives roughly the same number of keys. In our distributed similarity search system, mapping of keys (LSH bucket labels) onto a node does not use consistent hashing.

To provide perfect load balancing in the whole distributed similarity search system, different structures or algorithms should be proposed to assign the LSH indexes, which follow the normal distribution, onto the linear uniform node IDs space. Instead of using some gateway peers with predefined positions, which is efficient only for the initial period, we consider the underlying structure of DHT network, which is treated as a black box by [6].

3 Structure and Algorithms

We introduce a virtual node space on top of the underlying Chord node space. Each virtual node is like a single peer in the original DHT. In our structure, each Chord node (physical node) can be responsible for more than one virtual node. For example, each virtual node is responsible for a contiguous region of the key identifier space in original DHT, which in our system are the LSH buckets labels. While each Chord node can now own noncontiguous portions in the key identifier ring space. Generally speaking, virtual nodes layer is like an additional "mapping" step. In the virtual nodes space, we use consistent hashing to get node IDs and

allocate the keys generated by hash function based on sum to each node. The keys are the LSH buckets labels following normal distribution, which are generated by mapping function present in Sect. 2.

Let l_i denote the load of virtual node v_i, where l_i represents the sum of elements of all LSH bucket labels stored in virtual node v_i. For each physical Chord node n_i, the load L_i can be represented as the sum function of the load of all virtual node of node n_i, where l_j is the load of virtual node v_j:

$$L_i = \sum_{j=1}^{m_i} l_j$$

Node n_i is responsible for virtual nodes set $\{v_1,..., v_{mi}\}$, where m_i represents the number of virtual nodes.

As discussed in Sect. 2, by using mapping function based on sum (3), we can predict the distribution of the output of ξ_{sum} as (4), having an estimated mean $k/2$. The normal distribution has a nice property known as the 68-95.99.7 rule, which we can benefit from. Therefore, the overall load of all indexes is predictable. Considering the number of virtual nodes, we can assign an average load threshold T_{load} to every virtual node. T_{load} is determined by k, W, and $\{v_i\}$. If we assume that all the physical nodes are the same (heterogeneity of the physical nodes is not considered in this paper), the load threshold T_{load} can be a global value in the whole DHT.

Using the load l_i of virtual node v_i and the global load threshold T_{load}, we can divide the physical nodes, which are the Chord nodes in our system, into several different levels, according to the load L_i of each Chord node i. Here, similar to [1], we give two categories, heavy and light. A node is considered heavy if $L_i > T_{load}$, and is light otherwise.

Our goal is to perform load balancing by transferring the load (virtual nodes) of heavy nodes to light ones. Since our system is for distributed similarity search, the load related to virtual nodes are the LSH bucket labels, we assume that in our load balancing algorithms the load of all virtual nodes will not split.

4 Experiments

We now describe the experiments used to demonstrate the effectiveness of our load balancing algorithms and to better understand their performance. We consider the load of the heaviest node (l_{max}) and the lightest node (l_{min}). The number of traffic packages, which reveals the additional load to the overlay network by our algorithms, is also recorded. Table 1 shows the results of our experiment. We deployed 10 physical nodes in original Chord and assigned 50 virtual nodes in our algorithms (the size of successor list in virtual nodes $r = 1$ in this experiment).

Table 1 Load balancing results

DHT	Experiment results			
	Total load	l_{max}	l_{min}	Traffic number
Original chord	745	255	17	29878
Our algorithms	745	134	56	33902

When compared against the work in [6], which uses the original Chord without load balancing schemes in DHT overlay network, our algorithm achieves better load balancing for allocating the total load (combined of elements and LSH bucket labels) to the same number of physical nodes. While we obtained a smaller heaviest load and a bigger lightest load, which indicates that the load distribution is more uniform, we also increased the bandwidth burden by about 13 % relative to the original Chord protocol. The additional traffic packages are used to update the load information in our data structure. When the system is in a fair load balancing statement, we can increase the time period for running our algorithms to decrease the traffic load.

5 Conclusion

Based on the DHT-based distributed similarity search system using LSH and load balancing schemes using virtual nodes in DHT, we have discussed the particular load balancing problem with mapping the multi-dimensional LSH bucket space following a predictable distribution to the DHT naming space. We presented a chord-based structure using virtual nodes to perform our load balancing algorithms. Experiments have demonstrated that our work is effective with less traffic loads and achieves better load balancing.

References

1. Datar, M., Immorlica, N., Indyk, P., Mirrokni, V.S.: Locality-sensitive hashing scheme based on p-stable distributions. In: The Twentieth Annual Symposium on Computational Geometry, pp. 253–262. ACM, New York (2004)
2. Andoni, A., Indyk, P.: Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In: The 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 06), pp. 459–468. IEEE Press, New York (2006)

3. Stoica, I., Morris R., Karger, D., Kaashoek, M.F., Balakrishnan, H.: Chord: a scalable peer-to-peer lookup service for internet applications. In: 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM 01), pp. 149–160. ACM, New York (2001)
4. Ugo, M., Matteo, S.: Network-conscious π-calculus—a model of pastry. Electron. Notes Theor. Comput. Sci. **312**, 3–17 (2015)
5. Li, J., Li, Q., Liu, C.: Community-based collaborative information system for emergency management. Comput. Oper. Res. **42**(2), 116–124 (2014)
6. Haghani, P., Michel, S., Aberer, K.: Distributed similarity search in high dimensions using locality sensitive hashing. In: The 12th International Conference on Extending Database Technology: Advances in Database Technology (EDBT 09), pp. 744–755. ACM, New York (2009)
7. Lv, Q., Josephson, W., Wang, Z., Charikar, M., Li, K.: Multi-probe LSH: efficient indexing for high-dimensional similarity search. In: The 33rd International Conference on Very Large Data Bases (VLDB 07), pp. 950–961. ACM, New York (2007)
8. Alexandr, A., Piotr, I., Huy, L., Ilya, R.: Beyond locality sensitive hashing. In: The 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 14), pp. 1018–1028. ACM, New York (2014)
9. Bawa, M., Condie, T., Ganesan, P.: LSH forest: self-tuning indexes for similarity search. In: The 14th International Conference on World Wide Web (WWW 05), pp. 651–660. ACM, New York (2005)
10. Tang, Y., Zhou, S., Xu, J., Lee, W.: A lightweight multi-dimensional index for range queries over DHTs. IEEE Trans. Parallel Distrib. Syst. **22**(12), 2046–2054 (2011)
11. Fotiou, N., Katsaros, K.V., Xylomenos, G., Polyzos, G.C.: An inter-domain topology aware overlay for the support of name-resolution services in the future Internet. Comput. Commun. **62**, 13–22 (2015)
12. Sidhu, A., Kinger, S.: Analysis of load balancing techniques in cloud computing. Int. J. Comput. Technol. **4**(2), 60–66 (2013)
13. Dabek, F., Kaashoek, M.F., Karger, D., Morris, R., Stoica, I.: Wide-area cooperative storage with CFS. In: The Eighteenth ACM Symposium on Operating Systems Principles (SOSP 01), pp. 202–215. ACM, New York (2001)

Water Monitoring System Based on Recognition of Fish Behavior

Cong-Ren Lin, Ye Chen, Xi-Zhou Lin, Fei Yuan and Yi Zhu

Abstract Quality assurance of drinking water is vital in the current world. This research used fish to monitor and forewarn water quality in real time. This research adopted a Gaussian mixture model (GMM) to conduct background modeling, extract moving foregrounds, and automatically judge, screen, and dispose of various contours in the binary images through image preprocessing and morphological processing, and outlined the changing curves of indicators using a computer. The experimental results demonstrated that the above method was successful. This system was able to monitor the active state of a fish in real-time and accurately raise the alarm in the case of an abnormal condition. Based on this research, it would be easy to measure other indicators and develop extended functions that would be beneficial for follow up studies.

Keywords Biological monitoring · Gaussian mixture model · Motion detection

C.-R. Lin · Y. Chen · X.-Z. Lin · F. Yuan · Y. Zhu (✉)
Key Laboratory of Underwater Acoustic Communication
and Marine Information Technology, Ministry of Education,
Xiamen University, Xiamen, Fujian, People's Republic of China
e-mail: zhuyi@xmu.edu.cn

C.-R. Lin
e-mail: lincongren@xmu.edu.cn

Y. Chen
e-mail: chenye@stu.xmu.edu.cn

X.-Z. Lin
e-mail: xizhou.lin@outlook.com

F. Yuan
e-mail: yuanfei@xmu.edu.cn

1   Introduction

By using biological evaluation techniques, biological monitoring technology has been widely used in water safety warning and water quality monitoring systems. Different from experimental analyses, computer monitoring systems save manpower, material, and financial resources. With the discovery and cultivation of a number of indicator organisms, biological monitoring of water quality via computer systems will undoubtedly play an increasingly important role.

Taking fish as the biological indicator, this paper mainly studied fish behavior with the use of computer vision technology via online video monitoring. The research goal was to get information on fish behavior characteristics which are related to the biological monitoring of water quality. The specific research steps included a moving object detection method based on a GMM and the design and implementation of an online video monitoring system based on recognition of fish behavior. The remainder of this paper proceeds as follows: Sect. 2 describes the design of the monitoring system and research of key technologies. The specific content includes the introduction of the functions of each part and a brief description of target detection methods and the image information processing method which this article focused on. Section 3 introduces the hardware and software of the system and a brief analysis of system testing. Section 4 concludes the paper.

2   System Design and Technologies

2.1   System Framework

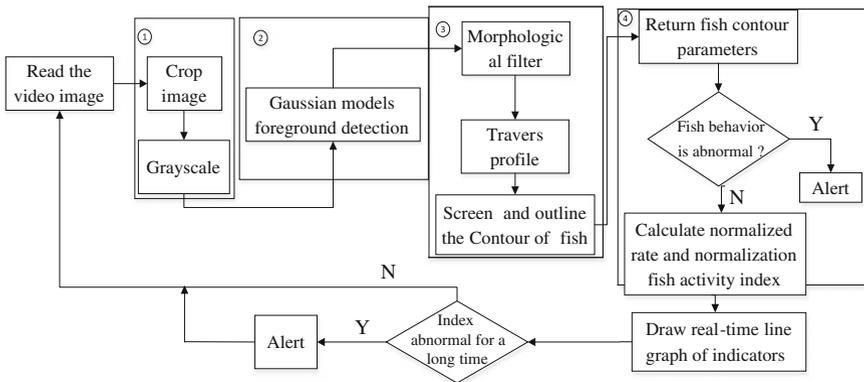Figure 1 shows the overall system framework, where each module is marked with a number.

Fig. 1   The framework of the system

The first module is the image preprocessing module. It included image cropping and gray-scaling. Module 2 is the foreground extraction module. It was the basis of subsequent processing. A GMM was adopted as the foreground extraction algorithm. Module 3 is the image processing and object extraction module. The main task was to process the GMM processed frames and filter out fine noise. Module 4 is the information processing and index calculation module. Through analyzing the value returned by the object extraction module, it calculated and normalized the rate and active index of the fish. Then, these data were displayed in the form of line charts on the software interface in real time. In addition, this module was responsible for checking whether the status of the fish was normal and whether the indicators were in an unreasonable state for a long period of time. When an unusual circumstance appeared, an alert was generated.

2.2 Foreground Extraction Algorithm

This system adopted a background subtraction method based on a GMM to extract the outline of fish. Assuming that the pixel value measured at different times was independently subject to distributed random process, the mixed Gaussian probability density was expressed as (1):

$$P(X_t) = \sum_{k=1}^{M} \omega_{k,t} \cdot \eta\left(X_t, \mu_{k,t}, \sum_{k,t}\right) \tag{1}$$

$$\eta\left(X_t, \mu_{k,t}, \sum_{k,t}\right) = \frac{\exp\left\{-\frac{1}{2}\left(X_t - \mu_{k,t}\right)^T \sum_{k,t}^{-1}\left(X_t - \mu_{k,t}\right)\right\}}{(2\pi)^{n/2}\left|\sum_{k,t}\right|^{1/2}} \tag{2}$$

where M is the number of mixed Gaussian distribution, and usually an integer between 3 and 7 [1]. $\eta\left(X_t, \mu_{k,t}, \sum_{k,t}\right)$ is the probability density function of X_t when governed by the kth Gaussian distribution [2]. Comparing the observed value of the pixel on the frame at time t with the corresponding M Gaussian distributions, if any one of the Gaussian distribution satisfied Eq. (3), that pixel was classified as a background point. Otherwise, the pixel was classified as a foreground point. T usually took the empirical value of 2.5 [3].

$$\left|X_t - \mu_{k,t-1}\right| < T \cdot \sigma_{k,t-1} \tag{3}$$

It was necessary for the GMM to update the background. Online update algorithm is a commonly used Gaussian background update algorithm [4]. First, it judged pixels according to Eq. (3), and if the value of the pixel matched with one of the Gaussian distributions in the model, the weight of that Gaussian distribution

increased, and the variance of the Gaussian distribution decreased. Otherwise, the system created a new Gaussian distribution to replace the original Gaussian distribution whose weight was minimal. For a GMM, it complied with the following expressions:

$$\hat{\mu}_{k,t} = \left(1 - \rho_{k,t}\right)\mu_{k,t} + \rho_{k,t}X_t \tag{4}$$

$$\hat{\omega}_{k,t} = (1 - \alpha)\omega_{k,t} + \alpha M_{k,t} \tag{5}$$

$$\hat{\sigma}_{k,t}^2 = \left(1 - \rho_{k,t}\right)\sigma_{k,t} + \rho_{k,t}(X_t - \mu_{k,t})^2 \tag{6}$$

$$\rho_{k,t} = \alpha \cdot \eta\left(X_t | \mu_{k,t}, \sigma_{k,t}\right) \tag{7}$$

$$M_{k,t} = \begin{cases} 1, & if \left|X_t - \mu_{k,t}\right| < T \cdot \sigma_{k,t} \\ 0, & else \end{cases} \tag{8}$$

where α is the update rate and ρ is the learning rate. The value of ρ affected the foreground extraction result. Taking full account of the speed of change in the foreground and background, the appropriate ρ value was selected as 0.1.

2.3 Image Processing

After image pre-processing and foreground extraction, the foreground of each frame in the video was extracted by the system. However, there was too much noise in the image. We chose median filtering to eliminate isolated noise. Setting the shape of the structure element a 5×5 square shapes. Besides, morphology can be used to simplify the image data, get rid of noise in the case of keeping the basic shape characteristics of an image [5]. The system mainly used the opening operator to smooth boundary and eliminated the tiny noise.

2.4 Image Bio-Information Processing

Estimation and Extraction of Fish. After the processing steps described above, the contour of the fish was extracted. Through measuring the maximum length and width of each contour, the contour was recognized as a fish and marked by a box when its maximum length and maximum width was greater than the experience value set previously.

   Normalization of Biological Fish. Because different fish have different swimming speeds, it was meaningless to compare the absolute speed of them. In this study, unit cm/s for absolute swimming speed was taken. However, in order to

assess the swimming speed, the absolute speed was normalized to body length/second, or BL/s.

After identifying and labeling the outline of the fish, the system returned a coordinate value for any corner of the contour to calculate the fish's rate. Suppose that coordinate of the upper left corner of preview frame is (x_1, y_1), the upper left corner coordinate of the current frame is (x_2, y_2), the frame rate is N, fish body length is l, the instantaneous rate of fish is v, and normalized velocity is $\dot{v}$.

$$v = N \cdot \sqrt{(x_2 - x_1)^2 + (y_2 - y)^2} \tag{9}$$

$$\dot{v} = \frac{N}{l} \cdot \sqrt{(x_2 - x_1)^2 + (y_2 - y)^2} \tag{10}$$

The above is about the motion of one fish. If there is a group of fish in the pool, the center of mass and center-of-mass velocity of the fish group can be calculate using a similar principle. According to the center-of-mass velocity of the fish group, the water quality can be indicated and the system stability can be improved.

Normalized Active Index of Fish. Generally, the fish randomly distribute in the lower tank when the monitored water was not contaminated. When the water was lightly polluted, the fish moved to the upper water, and appeared as random distribution. When the water was moderately polluted, the fish floated to the surface collectively and kept swimming. However, when the water was heavily polluted, some of the fish began to die. Then the number of active biological fish decreased. If the pollution was further intensified, all the fish died.

The dead fish were relatively static in the water, and were identified as background by Gaussian model. The identified fish's outlines were the contours of living fish. Then all pixels within the fish's contour were summed up to obtain the fish's area A. The active index K of fish was calculated using Eq. (11).

$$K = A/F \tag{11}$$

where F is the size of the video frame. Once the active index suddenly dropped and remained at a low value for a while, or was even reduced to 0 %, it meant the fish had begun to die, and water pollution was serious.

3 System Design and Application

3.1 System Design

The software system was composed of three parts, namely Visual Studio 2010 SP1, Measurement Studio 2010 SP1, and Emgu CV 2.4.0.1717. Measurement Studio 2010 SP1 provides special tools for signal processing, numerical analysis, and visualization. Emgu CV has a package for OpenCV image processing library under

a. NET platform. System framework modules shown in Fig. 1 were encapsulated by the software. All parameters were previously tested, adjusted, and optimized. Besides, the hardware system consisted of computer, water tank, and camera.

3.2 System Test

The biological fish species used in this test was red carp fingerlings. Before the test, the fish were normally bred in the aquarium for more than seven days. A number of responsive healthy juveniles were randomly selected to start the test.

Fish in Normal State. According to the fact that red carp tend to inhabit in the bottom of a water body, when the red carp were placed in unpolluted water, they swam with a durable speed [6] (or cruising speed) in the lower part of the tank. Besides, their swim speed was within the normal range. The quantity index of active fish was always maintained at normal levels. If the fish's state was judged normal by the system, the system did not alert. It can be seen from Fig. 2 that the system extracted the contour of the fish and plot images of normalized swimming rate of fish (in BL/s) and normalized the active index of fish in real time.

However, sometimes the fish floated to the surface to breathe or forage, and sometimes they rested at the bottom statically. Through observation of both states, the fish were not completely stationary but moved at a very slow speed. Through selecting corresponding videos to test, results were obtained. As shown in Fig. 3, the fish swimming speed curve indicated that the fish swam at a low velocity but not at 0.
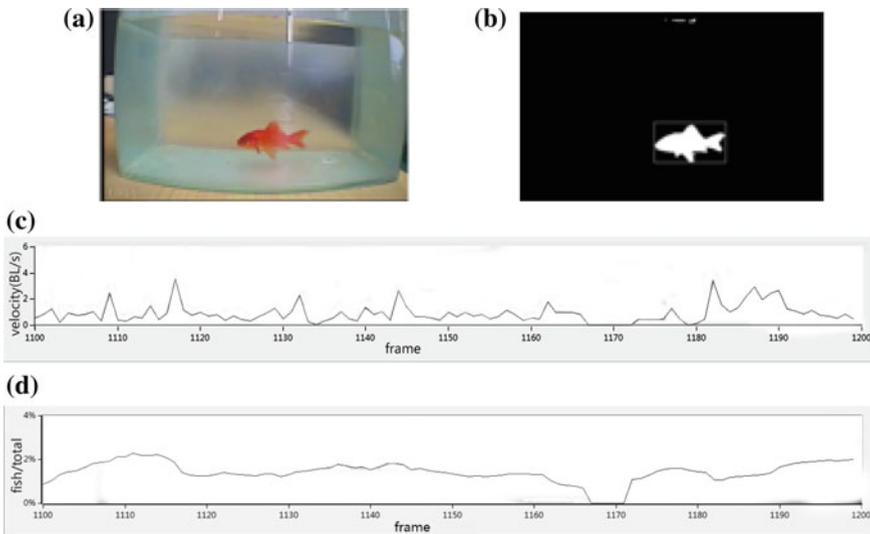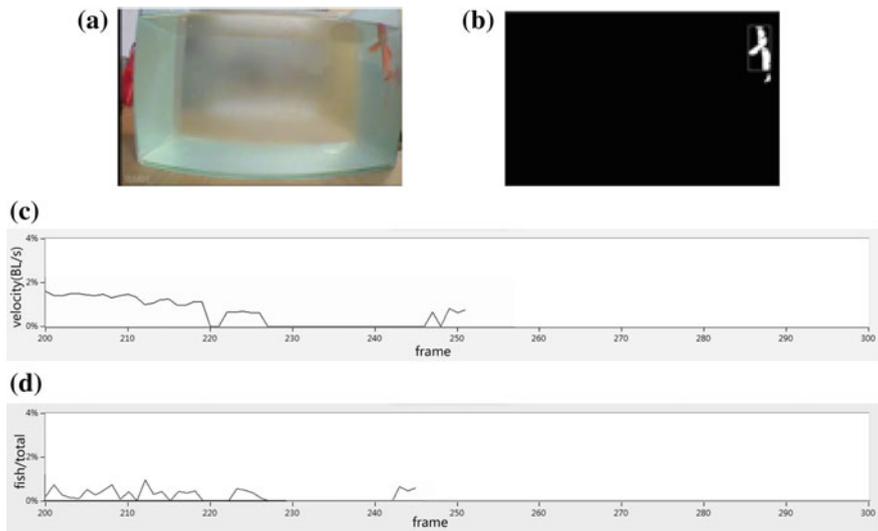

Fig. 2 The test results of fish in normal state, **a** normal state, **b** processing result, **c** normalized swimming speed, **d** normalized active index
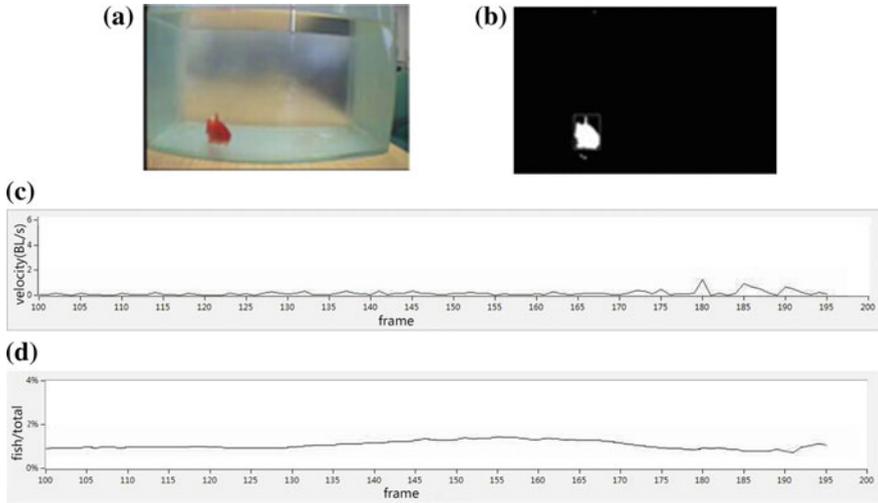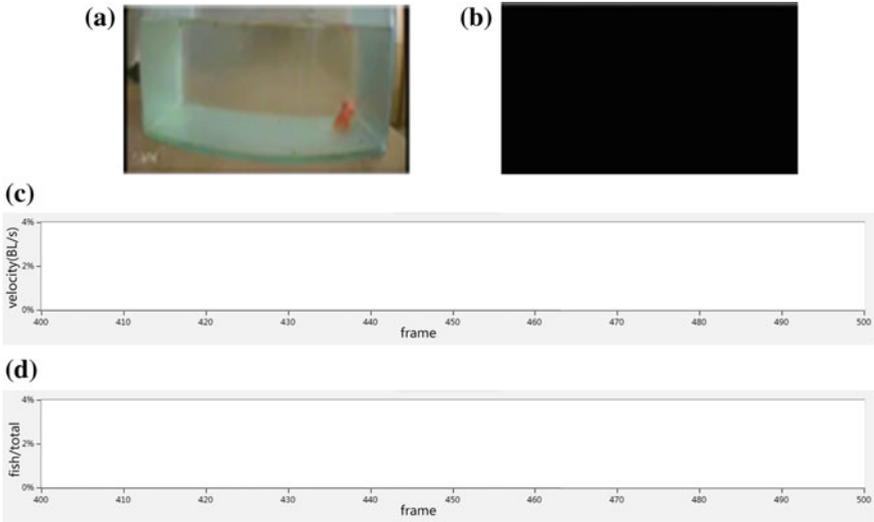
Fig. 3 The test results of fish rest underwater, **a** fish resting, **b** processing result, **c** normalized swimming speed, **d** normalized active index

The normalized active index stayed at a certain low value. The test results showed that the system did not respond with a false detection on these two occasions.

Fish in Abnormal State. Fish have very sensitive perceptions of changes in their water environment. Once the concentration of toxic substances in the water rose, the red carp developed abnormal behaviors. When the fish were stimulated by

Fig. 4 The test results of fish in early abnormal state, **a** initial abnormality, b processing result, **c** normalized swimming speed, **d** normalized active index

Fig. 5 The test results of fish in dead state, **a** death state, **b** processing result, **c** normalized swimming speed, **d** normalized active index

toxic substances, they became excited. The red carp exhibited some disordered behavior such as enhanced alertness, were close to the upper body of water, migrated, and displayed increased activity. In this study, vinegar, which red carp are sensitive to, was used to simulate toxic substances in the water. One red carp's initial abnormal condition is shown in Fig. 4.

If the water's toxic pollution wasn't resolved, the red carp showed the next symptoms. At that time the fish's swimming ability declined, was unable to maintain balance, and finally died. The dead fish was relatively static in the water. After a while, it was identified as background by the GMM, and then the active index decreased to 0 for a long time. System recognition results are shown in Fig. 5.

4 Conclusion

In this paper, a system framework that learned to monitor water quality through the recognition of fish behavior was presented. Thorough experiments demonstrated that monitoring results of the system were highly correlated with the water quality. For further research and more practical applications, the system could be extended and completed by introducing more organisms, such as daphnia magna, algae, plankton and other fish. In addition, the system was designed to biologically monitor only in the case of a single fish. Therefore the system would be improved by calculating the movements of centroid of the group, the dispersion of population,

etc. This would play a better role in quality monitoring of pools those have different kinds of fish.

References

1. Evangelio, R.H., Ptzold, M., Keller, I., Sikora, T.: Adaptively splitted GMM with feedback improvement for the task of background subtraction. IEEE Trans. Inf. Forensics Secur. **9**(5), 863–874 (2014)
2. Lin, L., Xu, Y., Liang, X.D., Lai, J.: Complex background subtraction by pursuing dynamic spatio-temporal models. IEEE Trans. Image Process. **23**, 670–683 (2014)
3. Noh, S.J., Shim, D., Jeon, M.: Background subtraction method using codebook-GMM model. In: 2014 International Conference in Control, Automation and Information Sciences (ICCAIS), pp. 117–120. IEEE, Gwangju, Korea (2014)
4. Priyadharshini, S., Dhanalakshmi, S.: Foreground object motion detection by background subtraction and signaling using GSM. In: 2014 International Conference in Information Communication and Embedded Systems (ICICES), pp. 1–6. IEEE, Chennai, India (2014)
5. Stell, J.G.: Formal concept analysis over graphs and hypergraphs. In: Graph Structures for Knowledge Representation and Reasoning. pp. 165–179. Springer International Publishing, Beijing, China (2014)
6. Kim, D.H., Lee, D.H., Lee, W.D.: Classifier using extended data expression. In: IEEE, Mountain Workshop on Adaptive and Learning Systems. pp. 154–159. IEEE Press, Piscataway (2006)

Weighted Modularity on a *k*-Path Graph

Ying-Hong Ma and Wen-Qian Wang

Abstract Community detection is one of the most interesting problems in the study of social networks. Most recent studies have focused on the design of algorithms to determine communities without knowing the number of communities in advance. This paper defines the *k*-path graph and generalizes Newman's modularity as a weighted modularity, and provides a detailed discussion on the relationship between eigenvalues and the maximum modularity of the network. It is proved that the maximum modularity of the weighted graphs depend on both the maximum eigenvalue and a relative parameter.

Keywords Social network · *k*-path graph · Modularity · Community detection

1 Introduction

The community structure in social networks has been studied for nearly 150 years since the "six-degrees of separation" phenomenon was first articulated [1]. Communities in networks are typically defined as a sub-graph in which links are more dense, and the rest parts of the networks are comparatively sparse [2, 3]. Methods for detecting community are similar to the graph partition problem in graph theory [4]. Newman presented a fast algorithm, which uses maximal modularity Q to determine communities [5, 6]. However the computational complexity of maximum modularity Q has been proved to be NP-complete [7]. Meo [8] exploited a novel measure of edge centrality based on k-paths and discovered the community structure by efficiently maximizing the network modularity. Chen et al. [9] presented two novel finely-tuned community detection algorithms provided by splitting and

Y.-H. Ma (✉) · W.-Q. Wang
School of Management Science and Engineering,
Shandong Normal University, Jinan, China
e-mail: yhma@sdnu.edu.cn

W.-Q. Wang
e-mail: wenqiansuk@163.com

merging the given network community structure. Waltman and Van Eck [10] presented a smart local moving algorithm to identify community structure with higher modularity by analyzing a diverse set of networks. Other researches on the modularity in recent years also presented different algorithms [11–13]. All of the above algorithms detect communities without knowing the number of communities in advance, which makes the solutions uncertain if the real number of communities in networks is unknown.

In this article, the number of communities is determined based on the eigenvalues of the probability matrix of social networks, and an estimate of this value is obtained. This paper is organized as follows: Sect. 2 defines a k-path graph of a given network, and presents a definition to determine a k-path matrix. In Sect. 3, the relationship between the eigenvalues and modularity is explored, and then the modularity of the social networks with eigenvalues is calculated. Finally, experimental analysis and conclusions are presented in Sect. 4.

2   *k*-Path Weight Graphs

A matrix consisting of all paths between any two nodes is defined, and then an approximation algorithm is outlined to determine the number of communities in a social network. The agents are represented by nodes in a network, and the influence between two nodes is represented by a weight on the link. In the following description, a network is denoted by G with n-node set V, and an m-link set E and G represents an undirected connected graph without loops or multi-edges. The adjacency matrix A of G is a $n \times n$ zero-one matrix denoted by $A = (a_{ij})_{n \times n}$, where $a_{ij} = 1$ if there is a link between nodes i and j; otherwise, $a_{ij} = 0$. The adjacency matrix of an undirected graph is symmetric. If it is weighted network, the weight of each link is denoted by w_{ij} and the weight matrix of G by $W = (w_{ij})_{n \times n}$. W represents the weight matrix regardless of whether G is weighted or not.

For a given positive integer $k,$ denote a path from nodes i to j by a k-path if it is a walk with $k + 1$ nodes and without a cycle. The matrix of k-path graph $S^k = (s_{ij}^k)$ is determined as follows:

If $k = 0$, $s_{ii}^0 = 1$ and $s_{ij} = 0$ for all $i \neq j$; that is, S^0 is the identity matrix.
If $k = 1$, $S^1 = W$; that is, S^1 is the weight matrix of G.

For all $k \geq 2$, $s_{ij}^k = \frac{1}{k} \sum_s \sum_{l=1}^{k} w_{i_{l-1}^s i_l^s}$ where s is the number of edge-disjoint k-paths from nodes i to j. The value s_{ij}^k can be viewed as the weight of an edge connecting i and j. Hence, a *k-path weight graph* G_k on G can be defined.

Definition of k-path weight graph: For a fixed k, let $w_{ij}(k) = \sum_{l=1}^{k} s_{ij}^l$ for all l-paths join nodes i and $j(l \leq l \leq k)$. $G_k = (V, E_k)$ represents a *k-path weight graph on G*, where $(i, j) \in E_k$. If there are paths with lengths not greater than k from i to j in G, then denote the weight matrix of G_k by $W(k) = (w_{ij}(k))_{n \times n}$. If $k = n - 1$, then G_{n-1} is a complete graph.

3 Modularity on G_k

The modularity is first defined on binary graphs, and is much more generalized [2, 5]. However, the values of all the generalized modularities are obtained by calculating the direct relations of the nodes, that is, the edges are considered to be weights rather than indirect weights. In fact, there is much useful information about the structure of the networks stored in indirect relations. The modularity is used to measure the direct and indirect information in a path weight graph and to detect the number of communities in the network G.

The modularity Q^* of a k-path weight graph is similar to Newman's. The matrix $W(k)$ of the k-path weight graph is used to replace the modular matrix in Newman's definition of modularity, because the expected Newman's matrix is not a fixed matrix for calculating modularity. Here, the objective function Q^* which maximize the weight of inter communities of k-path weight graph is defined.

The k-path weight graph $G_k = (V, E_k)$ with matrix $W(k)$ and $(i, j) \in E_k$ is studied. Assume that there are $q(q \le n/2)$ non-overlapping communities in G, denoted by $C = \{C_1, C_2, \ldots, C_q\}$, where $\cup_i C_i = V$.

3.1 Case 1. $q = 2$

There are exactly two communities in G: V_1 and V_2. If G is a binary graph and $k = 1$, the k-path weight graph G_k is the graph G. Therefore, Q^* is Newman's modularity.

If G is a weighted graph or $k \ge 2$, define the objective function Q^* as the maximum of the actual weights of the inner communities, that is, $Q^* = \max_{V_1 \cup V_2 = V, V_1 \cap V_2 = \emptyset} \sum w_{ij}$ where $(i, j \in V_1)$ or $(i, j \in V_2)$. Maximizing the weights of the inner communities minimizes the inter weights of communities, since the total actual edge weights of networks $\sum_{(i,j) \in E} W_{ij}$ is constant. This is why the expected matrix of Newman's modularity is not incorporated into the objective function.

To calculate Q^*, an indicator vector $\mathbf{r}$ is defined on V, $\mathbf{r} = (r_1, r_2, \ldots, r_n)$, where $r_i = 1$ if node $i \in V_1$; $r_i = -1$ if $i \in V_2$. Then, $r_i r_j = 1$ if i, j are in the same group; $r_i r_j = -1$ if they are in different groups. That is, $r_i r_j + 1 = 2$ if i, j are in the same group; otherwise, $r_i r_j + 1 = 0$. Hence, Q^* is described as follows:

$$
\begin{aligned}
Q^* &= \frac{1}{2} \max \sum_{(i,j \in V_1) \text{or} (i,j \in V_2)} w_{ij}(r_i r_j + 1) \\
&= \max \sum_{(i,j \in V_1) \text{or} (i,j \in V_2)} w_{ij} r_i r_j n \\
&= \max \mathbf{r}^T W(k) \mathbf{r}.
\end{aligned}
\tag{1}
$$

The eigenvector of $W(k)$ corresponding to the eigenvalue β_i by $\mathbf{u}_i$ is denoted, and $\mathbf{r} = \sum_i b_i \mathbf{u}_i$ is used as the linear normalization of all eigenvectors of $W(k)$. Therefore, $b_i = \mathbf{u}_i^T \mathbf{r}$. Then, Eq. 1 is equivalent to:

$$Q^* = \max \sum_i b_i^2 \beta_i. \tag{2}$$

The optimal value of Q^* in Eq. 2 relies not only on the positive eigenvalues of β_i but also on the values of all b_i. Assume that the eigenvalues are in decreasing order $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$; then, the largest eigenvalue β_1 of $W(k)$ is the possible solution such that Eq. 2 achieves the optimal solution. Let $\mathbf{u_1} = (u_1^{(1)}, u_2^{(1)}, \ldots, u_n^{(1)})^T$, be the eigenvector corresponding to the largest eigenvalue β_1. Then, the indictor vector $\mathbf{r}$ is obtained by $r_i = 1$ if $u_i^{(1)} \geq 0$; $r_i = -1$ if $u_i^{(1)} < 0$. Hence, the value of the objective function is $(\sum_{i=1}^n |u_i^{(1)}|^2)\beta_1$, corresponding to the eigenvalue β_1. Additionally, the two communities have $|\{i|u_i^{(l)} \geq 0\}|$ nodes and n-$|\{i|u_i^{(l)} \geq 0\}|$ nodes, respectively.

3.2 Case 2. q > 2

The objective function Q^* for the two communities in Case 1 can be naturally generalized to fit a case with more communities. Assume that there are q groups in the network. An indicator matrix $\mathbf{R} = (\mathbf{r_1}, \ldots, \mathbf{r_q})$ is defined with $\mathbf{r_i} = (r_{i1}, r_{i2}, \ldots, r_{in})^T$ and $r_{ij} = 1$, if node i is in the community j; otherwise, $r_{ij} = 0$. Since all communities are non-overlapping, each pair of columns are mutually orthogonal and the total number of nodes is n; hence, $Tr(\mathbf{R}^T\mathbf{R}) = $ n. Applying the analysis of Eq. 1 to Eq. 2, we obtain the following:

$$Q^* = \max \sum_{j=1}^n \sum_{s=1}^q (\mathbf{u_j}^T \mathbf{r_s})^2 \beta_j, \tag{3}$$

where $\mathbf{u_j}$ are eigenvectors of matrix $W(k)$ corresponding to eigenvalues β_j ($j = 1, 2\ldots$, q). Without a loss of generality, assume that all positive eigenvalues are in decreasing order $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_c$. Clearly, $q \leq c + 1$, since the objective is to maximize Q^* in Eq. 3 with c positive eigenvalues corresponding to c parts from V, and the remaining is the $(c + 1)$th portion. In order to make Q^* as large as possible, the first q largest eigenvalues are chosen from all positive eigenvalues. Then, $\sum_{j=1}^n \sum_{s=1}^q (\mathbf{u_j}^T \mathbf{r_s})^2 \beta_j$ is the maximum value in Eq. 3. However, there are q indictor vectors $\mathbf{r_i}$ ($q > 2$); therefore, it is not simple to choose the components of the indictor vector in Case 1, and we may not be able to choose as many as in the indictor vectors corresponding to the first q largest eigenvalues and eigenvectors. According to Eqs. 2 and 3, the maximum value of the objective function is closely related to all

positive eigenvalues and the eigenvectors of matrix $W(k)$ of G_k. Because the exact numbers of nodes in each community are unknown, and a method doesn't exist by which to choose the indictor vectors in **R**, we must rely on the known information (matrix $W(k)$) to determine the exact number of communities. Therefore, other means are necessary to estimate the number of communities in G_k. In the worst case, the computational complexity of the weight modularity is $O(n^2q)$, where n and q represent the size of the nodes and the number of communities, respectively.

4   Experimental Analysis and Summary

The primary challenge is to estimate the number of communities from the k-paths weighted graph. The classical sparse networks such as the Zachary club were studied. Figure 1 shows the gaps between eigenvalues which increase with k; the gaps between eigenvalues coincide with the number of communities. The number of communities is the index of the maximum eigenvalue gap subtracted from the number of positive eigenvectors.

This paper does not consider that the network structure may change over time; the number of communities is based on the steady state of the network, which would bias the real number of communities without information regarding the
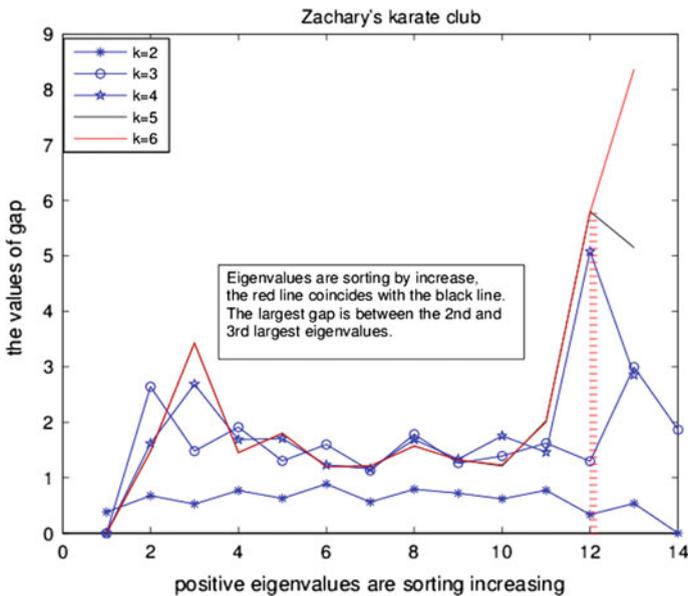
Fig. 1 Estimation of the number of communities from the *k*-paths weighted graph of the Zachary club

evolution of the networks. It will be interesting to characterize the relationship between eigenvalues and the structure of evolving communities.

References

1. Milgram, S.: The small world problem. Psychol. Today. **2**, 60–67 (1967)
2. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. **103**(23), 8577–8581 (2006)
3. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**, 7821–7826 (2002)
4. Scott, J.: Social Network Analysis: A Handbook, 2nd edn. Sage, London (2000)
5. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. Phys. Rev. E **69**, 066–133 (2004)
6. Radicchi, F., Castellano, C., Cecconi, F.: Defining and identifying communities in networks. Proc. Natl. Acad. Sci. **101**, 2658–2663 (2004)
7. Latapy, M., Pons, P.: Computing communities in large networks using random walks. In: Proceedings of the 20th International Symposium on Computer and Information Sciences, 2005. LNCS, vol. 3733, pp. 284–293. Springer, Heidelberg (2005)
8. Meo, P.D., Ferrara, E., Fiumara, G., et al.: Generalized Louvain method for community detection in large networks. In: Proceedings of the 11th International Conference on Intelligent Systems Design and Applications, pp. 88–93. Eprints, Spain (2011)
9. Chen, M.M., Kuzmin, K., Szymanski, B.K.: Community detection via maximization of modularity and its variants. IEEE Trans. Comput. Soc. Syst. **1**(1), 46–65 (2014)
10. Waltman, L., Van Eck, N.J.: A smart local moving algorithm for large-scale modularity-based community detection. Eur. Phys. J. B. **86**(11), 471 (2013)
11. Li, Y.D., Liu, J., Liu, C.L.: A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks. Soft. Comput. **18**(2), 329–348 (2014)
12. Chopade, P., Zhan, J.: Structural and functional analytics for community detection in large-scale complex networks. J. Big Data. **2**(1), 1–28 (2015)
13. Buchmann, T., Pyka, A.: The evolution of innovation networks: the case of a publicly funded German automotive network. Econ. Innov. New Technol. **24**(1–2), 114–139 (2015)