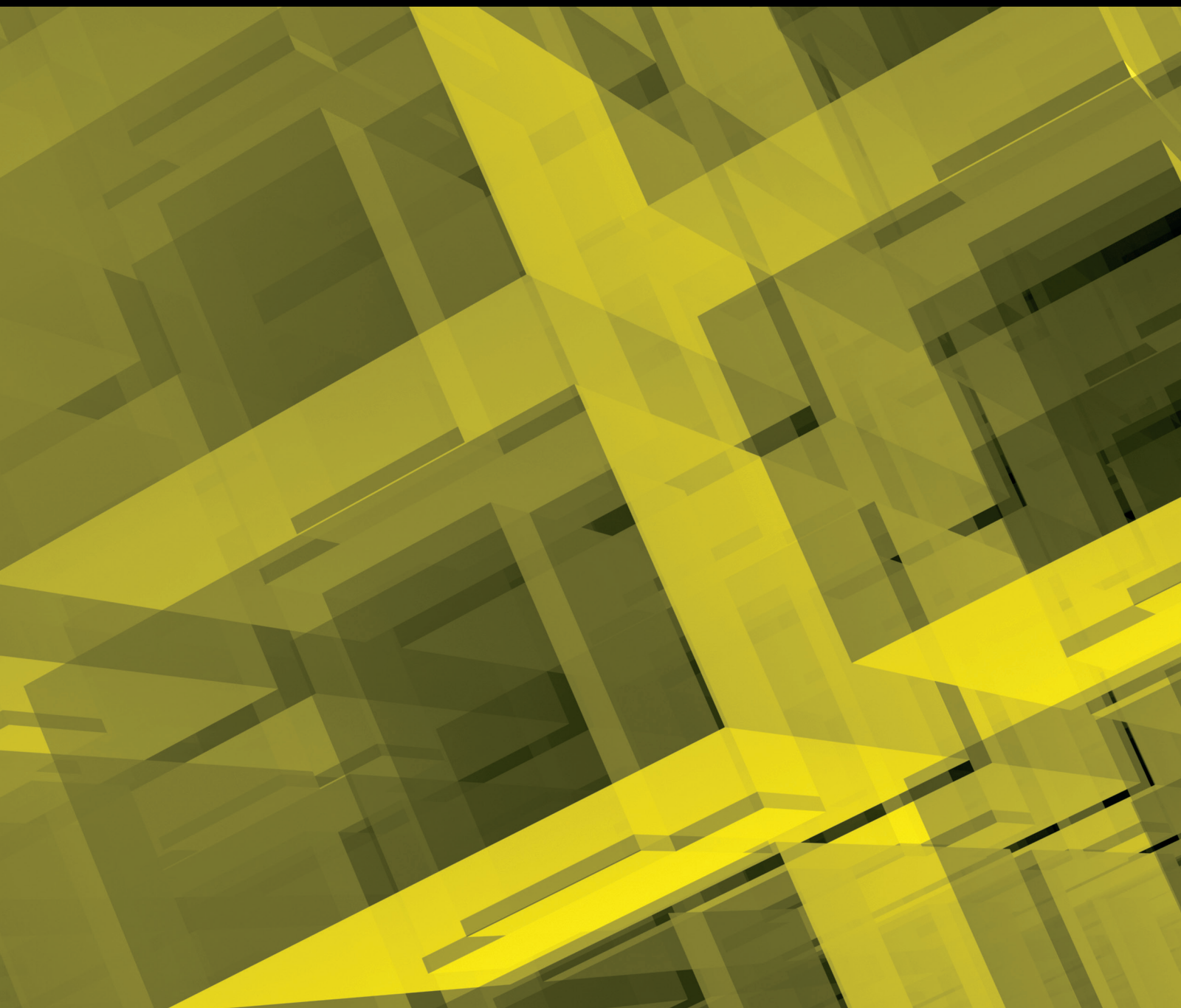


Advances in Mathematical Methods for Image and Signal Processing

Guest Editors: Feng Gao, Gongzhu Hu, Wenying Feng, Tokuro Matsuo,
and Vangalur Alagar





Advances in Mathematical Methods for Image and Signal Processing

Journal of Applied Mathematics

Advances in Mathematical Methods for Image and Signal Processing

Guest Editors: Feng Gao, Gongzhu Hu, Wenying Feng,
Tokuro Matsuo, and Vangalur Alagar



Copyright © 2014 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in “Journal of Applied Mathematics.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Saeid Abbasbandy, Iran
Mina B. Abd-El-Malek, Egypt
Mohamed A. Abdou, Egypt
Subhas Abel, India
Jnos Abonyi, Hungary
M. Montaz Ali, South Africa
Mohammad R. Aliha, Iran
Carlos J. S. Alves, Portugal
Mohamad Alwash, USA
Igor Andrianov, Germany
Boris Andrievsky, Russia
Whye-Teong Ang, Singapore
Abul-Fazal M. Arif, Saudi Arabia
Sabri Arik, Turkey
Ali R. Ashrafi, Iran
Allaberen Ashyralyev, Turkey
Francis T. K. Au, Hong Kong
Francesco Aymerich, Italy
Seungik Baek, USA
Olivier Bahn, Canada
Antonio Bandera, Spain
Jean-Pierre Barbot, France
Mostafa Barigou, UK
Roberto Barrio, Spain
Alfredo Bellen, Italy
Jafar Biazar, Iran
Anjan Biswas, Saudi Arabia
Abdellah Bnouhachem, Morocco
Gabriele Bonanno, Italy
Stephane P.A. Bordas, USA
James Robert Buchanan, USA
Humberto Bustince, Spain
Alberto Cabada, Spain
Xiao Chuan Cai, USA
Piermarco Cannarsa, Italy
Jinde Cao, China
Yijia Cao, China
Zhenfu Cao, China
Alexandre Carvalho, Brazil
Song Cen, China
Shih-sen Chang, China
Tai-Ping Chang, Taiwan
Wei-Der Chang, Taiwan
Shuenn-Yih Chang, Taiwan
Kripasindhu Chaudhuri, India

Ru-Dong Chen, China
Zhang Chen, China
Zhi-Zhong Chen, Japan
Xinkai Chen, Japan
Rushan Chen, China
Ke Chen, UK
Eric Cheng, Hong Kong
Ching-Hsue Cheng, Taiwan
Qi Cheng, USA
Chin-Hsiang Cheng, Taiwan
Jin Cheng, China
Hui Cheng, China
Francisco Chiclana, UK
Jen-Tzung Chien, Taiwan
Han H. Choi, Republic of Korea
Sazzad H. Chowdhury, Malaysia
Hung-Yuan Chung, Taiwan
Angelo Ciaramella, Italy
Pedro J. Coelho, Portugal
Carlos Conca, Chile
Vitor Costa, Portugal
Livija Cveticanin, Serbia
Binxiang Dai, China
Youjun Deng, China
Orazio Descalzi, Chile
Raffaele Di Gregorio, Italy
Kai Diethelm, Germany
Daniele Dini, UK
Urmila Diwekar, USA
Vit Dolejsi, Czech Republic
Bo-Qing Dong, China
Rafael Escarela-Perez, Mexico
Magdy A. Ezzat, Egypt
Meng Fan, China
Ya Ping Fang, China
Didier Felbacq, France
Antonio J. M. Ferreira, Portugal
Michel Fliess, France
Marco A. Fontelos, Spain
Dimitris Fotakis, Greece
Tomonari Furukawa, USA
Maria Gandarias, Spain
Xiao-wei Gao, China
Huijun Gao, China
Xin-Lin Gao, USA

Laura Gardini, Italy
Bernard J. Geurts, The Netherlands
Sandip Ghosal, USA
Pablo Gonzalez-Vera, Spain
Alexander N. Gorban, UK
Laurent Gosse, Italy
Keshlan S. Govinder, South Africa
Said R. Grace, Egypt
Jose L. Gracia, Spain
Maurizio Grasselli, Italy
Zhi-Hong Guan, China
Nicola Guglielmi, Italy
Frédéric Guichard, Canada
Kerim Guney, Turkey
Shu-Xiang Guo, China
Vijay Gupta, India
Saman K. Halgamuge, Australia
Abdelmagid S. Hamouda, Qatar
Bo Han, China
Maoan Han, China
Pierre Hansen, Canada
Ferenc Hartung, Hungary
Jesper H. Hattel, Denmark
Xiao-Qiao He, China
Yuqing He, China
Onesimo Hernandez-Lerma, Mexico
Luis J. Herrera, Spain
J. C. D. Hoenderkamp, The Netherlands
Thomas H'ne, Germany
Wei-Chiang Hong, Taiwan
Sun-Yuan Hsieh, Taiwan
Ying Hu, France
Ning Hu, Japan
Jianguo Huang, China
Dan Huang, China
Zhilong L. Huang, China
Ting-Zhu Huang, China
Zhenkun Huang, China
Mustafa Inc, Turkey
Gerardo Iovane, Italy
Anuar Ishak, Malaysia
Takeshi Iwamoto, Japan
Lucas J. Sánchez, Spain
George Jaiani, Georgia
GunHee Jang, Republic of Korea

| | | |
|-----------------------------------|-----------------------------------|------------------------------------|
| Zhongxiao Jia, China | Dongfang Liang, UK | Roberto Natalini, Italy |
| Daqing Jiang, China | Ching-Jong Liao, Taiwan | Srinivasan Natesan, India |
| Haijun Jiang, China | Yong-Cheng Lin, China | Tatsushi Nishi, Japan |
| Jianjun Jiao, China | Chong Lin, China | Andreas Öchsner, Australia |
| Xing Jin, China | Chein-Shan Liu, Taiwan | Wlodzimierz Ogryczak, Poland |
| Zhen Jin, China | Kang Liu, USA | Roger Ohayon, France |
| Zlatko Jovanoski, Australia | Zhijun Liu, China | Javier Oliver, Spain |
| Tadeusz Kaczorek, Poland | Yansheng Liu, China | Soontorn Oraintara, USA |
| Ido Kanter, Israel | Peter Liu, Taiwan | Donal O'Regan, Ireland |
| Abdul Hamid Kara, South Africa | Weiying Liu, China | Martin Ostoj-Starzewski, USA |
| Hamid Reza Karimi, Norway | Tao Liu, China | Turgut Özi, Turkey |
| Dimitrios A. Karras, Greece | Shutian Liu, China | Claudio Padra, Argentina |
| Ihsan Kaya, Turkey | Fei Liu, China | Vincent Pagneux, France |
| Dogan Kaya, Turkey | Chongxin Liu, China | Reinaldo Martinez Palhares, Brazil |
| Chaudry M. Khalique, South Africa | Zhengrong Liu, China | Quan K. Pan, China |
| Khalil Khanafer, USA | Zhuangjian Liu, Singapore | Endre Pap, Serbia |
| Adem Kılıçman, Malaysia | Jose L. López, Spain | Manuel Pastor, Spain |
| Hyunsung Kim, Republic of Korea | Shiping Lu, China | Giuseppe Pellicane, South Africa |
| Younjea Kim, Republic of Korea | Hongbing Lu, China | Francesco Pellicano, Italy |
| Jong Hae Kim, Republic of Korea | Benzhuo Lu, China | Juan Manuel Peña, Spain |
| Kazutake Komori, Japan | Henry Horng-Shing Lu, Taiwan | Jian-Wen Peng, China |
| Vassilis Kostopoulos, Greece | Yuan Lu, China | Ricardo Perera, Spain |
| Jisheng Kou, China | Changfeng Ma, China | Fernando Perez-Cruz, Spain |
| Roberto A. Kraenkel, Brazil | Ruyun Ma, China | Malgorzata Peszynska, USA |
| Vadim A. Krysko, Russia | Li Ma, China | Allan C. Peterson, USA |
| Jin L. Kuang, Singapore | Lifeng Ma, China | Vu Ngoc Phat, Vietnam |
| V. Kumaran, India | Nazim I. Mahmudov, Turkey | Andrew Pickering, Spain |
| Mirosław Lachowicz, Poland | Oluwole D. Makinde, South Africa | Hector Pomares, Spain |
| Hak-Keung Lam, UK | Francisco J. Marcelln, Spain | Mario Primicerio, Italy |
| Tak-Wah Lam, Hong Kong | Giuseppe Marino, Italy | Morteza Rafei, The Netherlands |
| Heung-Fai Lam, Hong Kong | Guiomar Martín-Herrn, Spain | Laura Rebollo-Neira, UK |
| Luciano Lamberti, Italy | Carlos Martín-Vide, Spain | Roberto Renò, Italy |
| Peter G L Leach, Cyprus | Alessandro Marzani, Italy | Juan A. Rodriguez-Velázquez, Spain |
| Jaehong Lee, Republic of Korea | Nikos E. Mastorakis, Bulgaria | Ignacio Rojas, Spain |
| Usik Lee, Republic of Korea | Nicola Mastronardi, Italy | Carla Roque, Portugal |
| Myung-Gyu Lee, Republic of Korea | Panayotis T. Mathiopoulos, Greece | Debasish Roy, India |
| Wen-Chuan Lee, Taiwan | Gianluca Mazzini, Italy | Imre J. Rudas, Hungary |
| Jinsong Leng, China | Marta Mazzocco, UK | Abbas Saadatmandi, Iran |
| Xiang Li, China | Michael McAleer, The Netherlands | Kunihiko Sadakane, Japan |
| Yongkun Li, China | Stephane Metens, France | Samir Saker, Egypt |
| Wan-Tong Li, China | Michael Meylan, Australia | Juan José Salazar González, Spain |
| Qingdu Li, China | Fan Min, China | Miguel A. F. Sanjuan, Spain |
| Hua Li, Singapore | Alain Miranville, France | Bogdan Sasu, Romania |
| Lixiang Li, China | Ram N. Mohapatra, USA | Richard Saurel, France |
| Wenlong Li, Hong Kong | Gisele Mophou, France | Wolfgang Schmidt, Germany |
| Shuai Li, Hong Kong | Cristinel Mortici, Romania | Jose R. Serrano, Spain |
| Yan Liang, China | Jaime E. Munoz Rivera, Brazil | Mehmet Sezer, Turkey |
| Jin Liang, China | Javier Murillo, Spain | Naseer Shahzad, Saudi Arabia |

Pengjian Shang, China
Hui-Shen Shen, China
Jian Hua Shen, China
Yong Shi, China
Yasuhide Shindo, Japan
Patrick Siarry, France
Fernando Simões, Portugal
Theodore E. Simos, Greece
Francesco Soldovieri, Italy
Abdel-Maksoud A. Soliman, Egypt
Qiankun Song, China
Xinyu Song, China
Yuri N. Sotskov, Belarus
Niclas Strömberg, Sweden
Ray K.L. Su, Hong Kong
Housheng Su, China
Jitao Sun, China
Wenyu Sun, China
Chengjun Sun, Hong Kong
Toshio Tagawa, Japan
Ying Tan, China
San-Yi Tang, China
XianHua Tang, China
Zhidong Teng, China
Engang Tian, China
Alexander Timokha, Norway
Yiying Tong, USA
Hossein Torkaman, Iran
Mariano Torrisi, Italy
Jung-Fa Tsai, Taiwan
Ch. Tsitouras, Greece
Antonia Tulino, USA
Sergey Utyuzhnikov, UK
Kuppalapalle Vajravelu, USA
Alvaro Valencia, Chile
Erik Van Vleck, USA
Ezio Venturino, Italy
Jesus Vigo-Aguiar, Spain
Michael N. Vrahatis, Greece

Heng Wang, Singapore
Yaonan Wang, China
Qing-Wen Wang, China
Guangchen Wang, China
Baolin Wang, China
Mingxin Wang, China
Dongqing Wang, China
Youqing Wang, China
Xiang Wang, China
Shuming Wang, Singapore
Yuh-Rau Wang, Taiwan
Pei-Guang Wang, China
Junzo Watada, Japan
Jinjia Wei, China
Guoliang Wei, China
Junjie Wei, China
Li Weili, China
Martin Weiser, Germany
Frank Werner, Germany
Man-Leung Wong, Hong Kong
Min-Hsien Wu, Taiwan
Wei Wu, China
Shi-Liang Wu, China
Cheng Wu, China
Shanhe Wu, China
Yonghui Xia, China
Tiecheng Xia, China
Gongnan Xie, China
Xuejun Xie, China
Wei Xu, China
Daoyi Xu, China
Zhiqiang Xu, China
Yuesheng Xu, USA
Fuzhen Xuan, China
Gregory S. Yablonsky, USA
Chao Yan, USA
Chao Yang, China
Suh-Yuh Yang, Taiwan
Chao-Tung Yang, Taiwan

Bin Yang, China
Guowei Yang, China
Chunyu Yang, China
Her-Terng Yau, Taiwan
Wei-Chang Yeh, Taiwan
Guan H. Yeoh, Australia
Chih-Wei Yi, Taiwan
Simin Yu, China
Bo Yu, China
Xiaohui Yuan, China
Jinyun Yuan, Brazil
Rafal Zdunek, Poland
Ashraf Zenkour, Egypt
Guisheng Zhai, Japan
Jianming Zhan, China
Meng Zhan, China
Long Zhang, China
Ke Zhang, China
Sheng Zhang, China
Jifeng Zhang, China
Heping Zhang, China
Henggui Zhang, UK
Liang Zhang, China
Jian-gang Zhang, China
Jingxin Zhang, Australia
Zhihua Zhang, China
Shan Zhao, USA
Chongbin Zhao, Australia
Renat Zhdanov, USA
Dong Zheng, USA
Huaichun Zhou, China
Bin Zhou, China
William Zhu, China
Xinqun Zhu, Australia
Quanxin Zhu, China
Goangseup Zi, Republic of Korea
Zhiqiang Zuo, China

Contents

Advances in Mathematical Methods for Image and Signal Processing, Feng Gao, Gongzhu Hu, Wenying Feng, Tokuro Matsuo, and Vangalur Alagar
Volume 2014, Article ID 245067, 2 pages

Automatic Segmentation of High Speed Video Images of Vocal Folds, Turgay Koç and Tolga Çiloğlu
Volume 2014, Article ID 818415, 16 pages

A Novel Two-Stage Spectrum-Based Approach for Dimensionality Reduction: A Case Study on the Recognition of Handwritten Numerals, Mohammad Amin Shayegan, Saeed Aghabozorgi, and Ram Gopal Raj
Volume 2014, Article ID 654787, 14 pages

Performance Comparison of Wavelet and Multiwavelet Denoising Methods for an Electrocardiogram Signal, Balambigai Subramanian, Asokan Ramasamy, and Kamalakannan Rangasamy
Volume 2014, Article ID 241540, 8 pages

Capacity of 60 GHz Wireless Communications Based on QAM, Jingjing Wang, Na Li, Wei Shi, Yangyang Ma, Xiulong Liang, and Xinli Dong
Volume 2014, Article ID 815617, 5 pages

The Determinants, Inverses, Norm, and Spread of Skew Circulant Type Matrices Involving Any Continuous Lucas Numbers, Jin-jiang Yao and Zhao-lin Jiang
Volume 2014, Article ID 239693, 10 pages

Application of EMD-Based SVD and SVM to Coal-Gangue Interface Detection, Wei Liu, Kai He, Qun Gao, and Cheng-yin Liu
Volume 2014, Article ID 283606, 6 pages

Palm Print Edge Extraction Using Fractional Differential Algorithm, Chunmei Chi and Feng Gao
Volume 2014, Article ID 896938, 7 pages

A Joint Encryption and Reversible Data Hiding Scheme Based on Integer-DWT and Arnold Map Permutation, Shun Zhang, Tiegang Gao, and Guorui Sheng
Volume 2014, Article ID 861782, 12 pages

A Novel Algorithm of Quantum Random Walk in Server Traffic Control and Task Scheduling, Dong Yumin and Xiao Shufen
Volume 2014, Article ID 818479, 8 pages

A New Method of Image Denoising for Underground Coal Mine Based on the Visual Characteristics, Gang Hua and Daihong Jiang
Volume 2014, Article ID 362716, 7 pages



A Multiresolution Image Completion Algorithm for Compressing Digital Color Images, R. Gomathi
and A. Vincent Antony Kumar
Volume 2014, Article ID 757318, 13 pages

Image Sequence Fusion and Denoising Based on 3D Shearlet Transform, Liang Xu, Junping Du,
and Zhenhong Zhang
Volume 2014, Article ID 652128, 10 pages

Exact CS Reconstruction Condition of Undersampled Spectrum-Sparse Signals, Ying Luo, Qun Zhang,
Guozheng Wang, and Youqing Bai
Volume 2013, Article ID 715848, 6 pages

A Novel Optimization-Based Approach for Content-Based Image Retrieval, Manyu Xiao, Jianghu Lu,
and Gongnan Xie
Volume 2013, Article ID 785824, 6 pages

Adaptive Self-Occlusion Behavior Recognition Based on pLSA, Hong-bin Tu, Li-min Xia,
and Lun-zheng Tan
Volume 2013, Article ID 506752, 9 pages

Image Matching Using Dimensionally Reduced Embedded Earth Mover's Distance, Fereshteh Nayyeri
and Mohammad Faizul Nasrudin
Volume 2013, Article ID 749429, 11 pages

Editorial

Advances in Mathematical Methods for Image and Signal Processing

Feng Gao,¹ Gongzhu Hu,² Wenying Feng,³ Tokuro Matsuo,⁴ and Vangalur Alagar⁵

¹ Science School, Qingdao Technological University, Qingdao, Shandong 266033, China

² Department of Computer Science, Central Michigan University, Mount Pleasant, MI 48859, USA

³ Department of Computing & Information Systems, Department of Mathematics, Trent University, Peterborough, ON, Canada K9J 7B8

⁴ Graduate School of Industrial Technology, Advanced Institute of Industrial Technology, Higashi-Oi, Shinagawa 140-0011, Japan

⁵ Department of Computer Science and Software Engineering, Concordia University, Montreal, QC, Canada H4B 1R6

Correspondence should be addressed to Feng Gao; gaofeng99@sina.com

Received 28 May 2014; Accepted 28 May 2014; Published 16 June 2014

Copyright © 2014 Feng Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mathematical methods for signal and image processing present the mathematical methodology for generic image and signal analysis tasks. The existing literature on image methodology is rather scattered. This special issue brings together some seemingly different points of view to stress their conceptual relations and analogies. In this issue, some papers focus on specific applications, while some other papers discuss the methodological frameworks on which specific applications are built. Covering many topics at the forefront of current research, this issue will be of particular interest to researchers in the fields of computer vision, image processing, signal processing, medical imaging, visual perception, pattern recognition, and so on.

Among the 16 exciting papers compiled in this special issue, two papers discuss fundamental mathematical concepts. J.-J. Yao and Z.-L. Jiang consider the skew circulant and skew left circulant matrices with continuous Lucas numbers. They discuss the invertibility of the skew circulant matrices and present the determinant and the inverse matrices by constructing the transformation matrices. H.-B. Tu et al. focus on the problem of self-occlusion in the field of human action recognition. In their paper, a new adaptive occlusion state behavior recognition approach is presented based on Markov random field and probabilistic latent semantic analysis.

Two papers present the specific applications of mathematical methods in image and signal processing. W. Liu et al. propose a new vibration signal analysis approach to detect the

coal-gangue interface based on singular value decomposition (SVD) techniques and support vector machines (SVMs). The empirical mode decomposition (EMD) is also used in their work. G. Hua and D. Jiang propose a new self-adaptive method for image denoising based on visual characteristics.

One paper deals with high speed video images. T. Koç and T. Çiloglu propose an automatic method for segmenting glottis in high speed endoscopic video (HSV) images of vocal folds. Their method is based on image histogram modeling.

In the area of image encryption and image coding, S. Zhang et al. propose a joint encryption and reversible data hiding scheme. R. Gomathi and A. Vincent Antony Kumar introduce a new framework for image coding that uses image inpainting method. In the proposed algorithm, the input image is subjected to image analysis to remove some of the portions purposefully.

There are two papers in the area of communication signal processing. J. Wang et al. discuss 60 GHz wireless communications over the additive white Gaussian noise channel. Channel capacity with quadrature amplitude modulation (QAM) is investigated for the unlicensed 59–64 GHz radio spectrum set aside by FCC. Their results prove that QAM is an attractive scheme for 60 GHz wireless communications. D. Yumin and X. Shufen propose quantum random walk optimization model and algorithm in network cluster server traffic control and task scheduling.

Two papers deal with image retrieval and image matching. M. Xiao et al. propose new optimization strategies on vocabulary tree building, retrieval, and matching methods. F. Nayyeri and M. F. Nasrudin study image matching by using dimensionally reduced embedded earth mover's distance.

In the area of image recognition and image reconstruction, M. A. Shayegan et al. propose a new two-stage approach for dimensionality reduction. Their method is based on one-dimensional and two-dimensional spectrum diagrams of standard deviation and minimum to maximum distributions for initial feature vector elements. Y. Luo et al. study exact CS reconstruction condition of undersampled spectrum-sparse signals. They mathematically prove that, in certain cases, the exact CS reconstruction of a spectrum-sparse signal from undersampled data is impossible.

Three papers focus on classical mathematical methods of image processing. B. Subramanian et al. discuss the properties of wavelets and multiwavelets with regard to the denoising capability compared to conventional filtering techniques. L. Xu et al. propose a novel algorithm for image sequence fusion and denoising simultaneously in 3D shearlet transform domain. C. Chi and F. Gao propose an algorithm based on fractional difference for the edge extraction of thenar palm print image.

By compiling these papers, we hope to enrich our readers and researchers with respect to the advances in mathematical methods for image and signal processing.

*Feng Gao
Gongzhu Hu
Wenyi Feng
Tokuro Matsuo
Vangalur Alagar*

Research Article

Automatic Segmentation of High Speed Video Images of Vocal Folds

Turgay Koç^{1,2} and Tolga Çiloğlu²

¹ Department of Electronic Communication Engineering, Süleyman Demirel University, 03200 Isparta, Turkey

² Department of Electrical and Electronics Engineering, Middle East Technical University, 06800 Ankara, Turkey

Correspondence should be addressed to Turgay Koç; turgaykoc@sdu.edu.tr

Received 24 January 2014; Revised 18 April 2014; Accepted 20 April 2014; Published 5 June 2014

Academic Editor: Feng Gao

Copyright © 2014 T. Koç and T. Çiloğlu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An automatic method for segmenting glottis in high speed endoscopic video (HSV) images of vocal folds is proposed. The method is based on image histogram modeling. Three fundamental problems in automatic histogram based processing of HSV images, which are automatic localization of vocal folds, deformation of the intensity distribution by nonuniform illumination, and ambiguous segmentation when glottal gap is small, are addressed. The problems are solved by using novel masking, illumination, and reflectance modeling methods. The overall algorithm has three stages: masking, illumination modeling, and segmentation. Firstly, a mask is determined based on total variation norm for the region of interest in HSV images. Secondly, a planar illumination model is estimated from consecutive HSV images and reflectance image is obtained. Reflectance images of the masked HSV are used to form a vertical slice image whose reflectance distribution is modeled by a Gaussian mixture model (GMM). Finally, estimated GMM is used to isolate the glottis from the background. Results show that proposed method provides about 94% improvements with respect to manually segmented data in contrast to conventional method which uses Rayleigh intensity distribution in extracting the glottal areas.

1. Introduction

Vocal system monitoring is essential for clinical analysis of voicing and investigation of speech production models. Today, high speed endoscopic video (HSV) of vocal folds is a state-of-the-art method to investigate vocal fold vibration. With the development of high speed video cameras, the vibration of vocal folds can be captured at 4000 fps with an image resolution of 256×256 pixels. Such a system can provide 20 frames in a glottal cycle of 200 Hz vibration. Reduction of the dimensionality of spatiotemporal information and representation of high speed video data in a simple, convenient, and lossless manner is a challenge. By means of HSV, some of the characteristics of vocal fold vibrations, such as asymmetric vibration, glottal area, and glottal width, are measured for clinical and engineering applications. Quantification of the difference between vibration patterns of left and right vocal folds in both spatial and temporal domains is required for objective analysis of voice disorders [1–4]. The time variation of glottal area is of particular interest in HSV based analysis

[2, 5–9]. It is used as a reference signal in estimating the parameters of biomechanical models of vocal folds [5, 6]. These models are used for functional analysis of vocal folds vibrations as well as articulatory speech synthesis.

Currently, glottal area extraction from HSV images is based on histogram thresholding [1, 10], region-growing [7, 11], and active contours [12]. Histogram thresholding method uses the difference between intensity distributions of the object and background. In HSV images, pixels corresponding to vocal folds and tissue around them have large intensity values compared to the pixels corresponding to the opening between the vocal folds, that is, glottis. The aim in the histogram based methods is to find a threshold to discriminate the low intensity pixels corresponding to glottis from the high intensity pixels. The intensity distributions are usually modeled by parametric functions. First, distributions of the glottis and background intensities are estimated and a threshold is determined. Glottal area can be estimated quite accurately provided that the intensity distribution is bimodal.

However, intensity distribution is very sensitive to illumination. Nonuniform illumination can ruin the modality of intensity distributions. Thus, it is a significant problem in the histogram based segmentation methods. The advantages of the method are its suitability for real-time applications due to little computational requirement, and under uniform illumination it is very effective for glottal area extraction.

Region-growing based methods use the histogram thresholding as a first step of the algorithm. After applying thresholding, a binary image is obtained for further processing. In the next step, one of the connected regions is used as an initial seed; then the seed points are propagated up to a state at which the difference in the intensity of the boundary pixels and intensities of their neighbours reaches a certain limit. Region-growing methods require an intelligent algorithm to select multiple seeds in video frames, since vocal folds can have partial contacts during glottal opening and closing phases. Therefore, their performances depend on the combination of histogram thresholding and accurate selection of seed points.

In the active contour based methods, an initial contour for glottis is found by using edge detection operation; then it is shrunk and expanded iteratively to minimize an energy function. One disadvantage is that the convergence requires significant amount of time due to computational burden. Another disadvantage is that the final boundary is affected by the selection of initial contour and noise in the HSV images [8, 12]. It is not suitable for applications requiring processing large number of video frames in a small fraction of time or real-time, such as clinical evaluation. However, segmentation of HSV in an automatic manner is indispensable for applications requiring analysis of large number of frames, such as vocal fold vibration functional analysis and speech production analysis.

Currently, the major challenge is automatic extraction of glottal area [7, 8, 12]. Most of the existing image processing algorithms in the literature are not completely automatic and require user intervention. In [7], the intensity distribution of an HSV image is modeled by Rayleigh distribution and a binary image is obtained by using histogram thresholding according to Bayes's decision rule. After selection of single seed, a region-growing operation is applied to find the final boundaries. However, as it is shown later in this study, modeling the intensity distribution of HSV without considering the region of interest is unreliable and finding an accurate threshold may involve ambiguities. In addition, the results may get worse due to the nonuniform illumination. The automatic algorithms used in [8, 12] are based on active contour method. However, analysis of a single image can take about several minutes with these methods. To extract glottal area in an automatic and efficient way, a histogram based method is suitable due to its computational efficiency. In this paper, three fundamental problems in automatic histogram based HSV processing, which are automatic localization of vocal folds, deformation of the bimodality of intensity distribution by nonuniform illumination, and ambiguous segmentation when glottal gap is small, are addressed. The problems are solved by a novel approach which involves

automatic TV-Norm based masking, illumination, and reflectance modeling.

The intensity histogram depends on the region of interest in the image. To obtain a bimodal intensity distribution for segmentation, region of interest must contain glottis and tissue in its proximity. Automatic calculation of region of interest is the first step of automatic vocal fold segmentation system. It is required not only in histogram based algorithms but also in other automatic HSV segmentation algorithms to reduce the amount of processed data to minimize computational complexity and to avoid the possibility of false region determination. In the literature, to the best of our knowledge, automatic vocal fold localization is performed in [9, 12]. In [12], first, the darkest image is determined as a representative HSV image; then an edge detection operation followed by connected component analysis is applied on it. At the final stage, a rectangular mask corresponding to the largest vertical connected region is used as a region of interest in the image. However, it is quite sensitive to noise due to the fact that even little noise in the image can degrade the edges and connected components in the image. In [9], an image sequence related to a glottal cycle is used to determine the glottis. It is assumed that the lowest intensity value is in the glottis. The row-wise and column-wise intensity minima are calculated from each frame in the image sequence. Then, their averages form vertical profile, V_n , and horizontal profile, H_n , vectors. They expected that V_n and H_n have a minimum value between the locations of vocal folds margins. The two neighboring maxima on each side of the minimum are used to locate the margins. It is our experience that V_n and H_n are highly sensitive to illumination and tissue structures in the image. Usually it is not possible to locate glottis clearly with the use of these profiles. The proposed solution is based on the intensity variation caused by the vibratory motion of vocal folds. The intensities of vocal fold edges and glottis change almost periodically during the vibration of vocal folds. It is realized that the largest intensity variation in HSV is observed at the glottis. In this study, a novel masking algorithm is proposed which uses the total variation norm (TV-norm) of HSV image at the automatic mask determination stage. It can be used as a preprocessing step in many vocal fold segmentation systems.

The second problem which is addressed here is the deformation of the modality of the intensity distribution due to the nonuniform illumination. Nonuniform illumination over a scene can be reduced by modeling reflectance and illumination. Intensity can be considered as the product of reflectance and illumination [13]. One solution is to model the illumination and then recover the reflectance of the object by removing the effect of illumination. The estimated reflectance image can be used in segmentation. In this paper, a planar illumination model is proposed for vocal fold monitoring systems. By means of the proposed model, bimodal HSV reflectance image histograms can be acquired from complex multimodal HSV intensity histograms.

In HSV image segmentation systems, images are processed consecutively; each frame is first analyzed and then segmented. However, the vibration of vocal folds brings

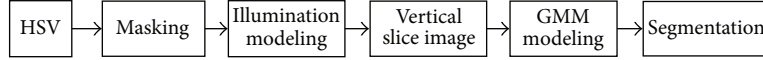


FIGURE 1: HSV segmentation algorithm.

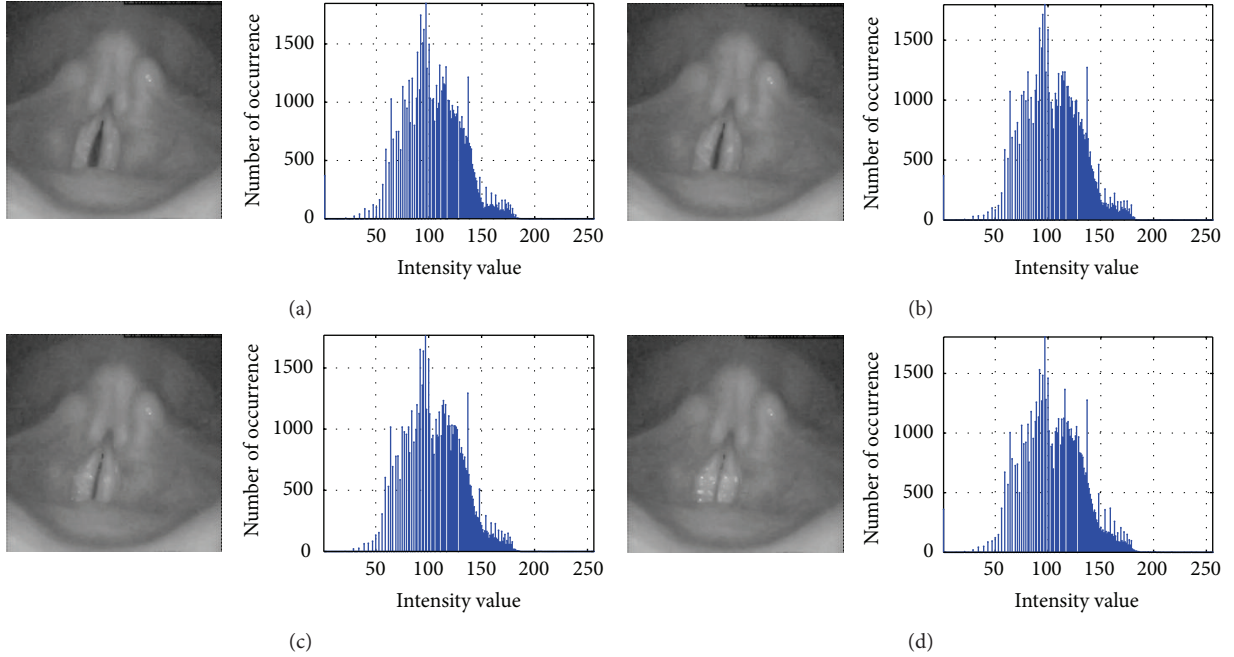


FIGURE 2: Four snapshots from HSV and corresponding histograms.

some difficulties in histogram based segmentation. The vocal folds' edges approach each other and make a partial or full contact at each glottal period. At the beginning of the glottal opening phase and at the end of the glottal closing phase, a small portion of the image pixels belongs to glottal opening or gap. It causes the intensity histogram of the image to be unimodal which makes finding an accurate and reliable threshold for segmentation difficult. In this study, to remove the uncertainty in threshold determination, a threshold determination algorithm which uses the intensity variation provided by the vibratory motion of vocal folds from an image frame sequence is proposed.

The main steps of the developed algorithm are shown in Figure 1. First, using the TV-norm of an image sequence, a mask is determined to crop a portion of an original frame that contains vocal folds. Then, an illumination model is estimated from the mean temporal intensity variation of the HSV. A reflectance image is estimated after the illumination modeling phase. In the next step, reflectance histogram along central longitudinal cross sections of glottis over a sequence of masked frames is used to determine the threshold for glottal boundary detection. The reflectance distribution is modeled by using Gaussian mixture model (GMM). Finally, in the segmentation step, a reflectance threshold is determined by a Bayesian approach.

This paper is organized as follows. Section 2 describes the automatic vocal fold localization problem and then presents automatic masking algorithm. The degradation of

intensity distribution due to the nonuniform illumination and proposed planar illumination model is presented in Section 3. Constructing vertical slice image from masked frames and GMM based reflectance modeling for segmentation are presented in Section 4. Segmentation results and comparison of proposed algorithm with Yan's method [7] on manually marked HSV images are presented in Section 5. Finally, conclusion is given in Section 6.

2. Masking

The accurate determination of region of interest in a HSV image is necessary for intensity histogram processing. Some HSV image frames and corresponding intensity histograms are shown in Figure 2. The vocal folds are located in the middle of the consecutive image frames. The dark region between the folds is called glottis. As the folds come together, the area of the glottis decreases as seen in the images.

Despite the considerable change in the glottal area in each image shown in Figure 2, the intensity histograms are identical and almost unimodal. The intensity distribution is insensitive to the change in the glottal area if the whole HSV image is chosen as a region of interest. By visual inspection of the histograms, it is not easy to decide whether the vocal folds are open or close. Furthermore, due to the almost unimodal intensity distribution, it is hard to determine a reliable threshold value to discriminate the intensity values of pixels of the glottis from the intensity values of its neighbouring

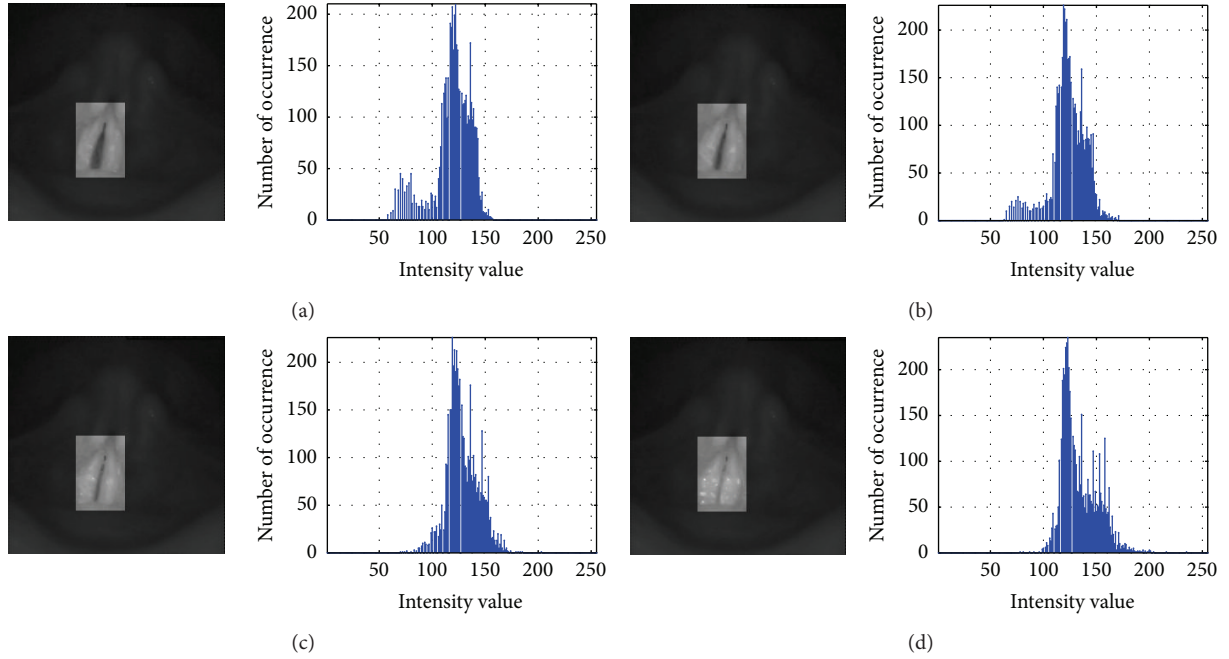


FIGURE 3: Four snapshots from masked HSV and corresponding histograms.

pixels in the intensity histograms. However, by reducing the region of interest as shown in Figure 3, the intensity histograms can be made bimodal and sensitive to the changes in the glottal area.

The intensity histograms of the masked image frames shown in Figure 3 are either unimodal or bimodal depending on the state of the vocal folds. When the folds are closed or almost closed as shown in Figures 3(c) and 3(d), the small intensity values corresponding to the dark region at the glottis are replaced by the high intensity values of vocal folds edges; hence background intensity distribution is dominant in the corresponding histograms. As a result, the intensity distributions turn out to be unimodal. On the other hand, when the vocal folds are open as shown in Figures 3(a) and 3(b), the small intensity values due to the darkness in the glottis yield the distributions have a stronger bimodal character because of the reduced size of data collection region (by masking). A reliable threshold estimation from an intensity histogram is possible when the opening between the vocal folds is sufficiently large in the masked HSV image. An intensity value at the left edge of the background intensity distribution can be chosen as a threshold [13]. For automatic processing of HSV images, an automatic algorithm that determines the location of the glottis is required. A novel automatic HSV masking algorithm is presented in the following subsection.

2.1. Automatic Masking Algorithm. The algorithm yields a rectangular mask from consecutive HSV images in which vocal folds vibrate. It uses the total variation of the intensity values of each pixel. Total variation norm is used in image restoration and noise reduction [14–16]. In this paper, it is used in the opposite direction. In HSV images, the most

active structures are vocal folds. Their vibration produces large intensity changes at pixels corresponding to the glottis (later demonstrated in Section 3). Hence, a large frame to frame intensity variation is a significant indication of a pixel corresponding to glottis.

TV-norm is computed over a sequence of frames as

$$TV(x, y) = \sum_{n=1}^{N-1} |I(x, y, n+1) - I(x, y, n)| \quad \forall x, y \in I, \quad (1)$$

where x and y are the spatial variables and n denotes frame index. $I(x, y, n)$ is the intensity function and N is the number of frames in the sequence. It should be emphasized that TV-norm is used in time, not in space. $TV(x, y)$ produces larger values at the locations of high intensity variation. TV obtained from 300 HSV image frames is shown in Figure 4.

The largest TV values (red color) are accumulated at the glottal region while the smaller values are distributed over the background. This is used for the determination of a mask to locate the glottis in HSV frames automatically.

In placing the mask, horizontal and vertical maxima statistics are used. Let M_x and M_y be two sets whose elements are determined by

$$\begin{aligned} M_x(y) &= \max_x (TV(x, y)), \quad y = 1, 2, \dots, N, \\ M_y(x) &= \max_y (TV(x, y)), \quad x = 1, 2, \dots, N. \end{aligned} \quad (2)$$

M_x and M_y calculated over TV-image in Figure 4 can be seen in Figure 5. The largest M_x values are located between values about 90 and 120. This interval points out the horizontal interval of the most active region, glottis, in the

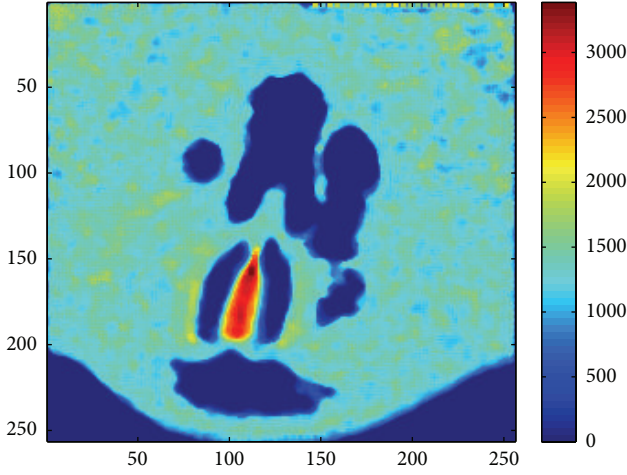
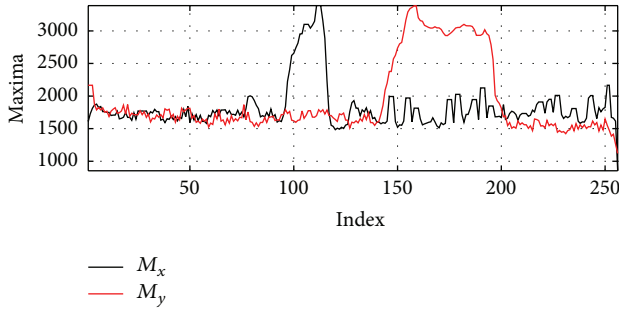


FIGURE 4: TV-image obtained from 300 frames.

FIGURE 5: Plots of $M_x(y)$ and $M_y(x)$ calculated from TV-image shown in Figure 4.

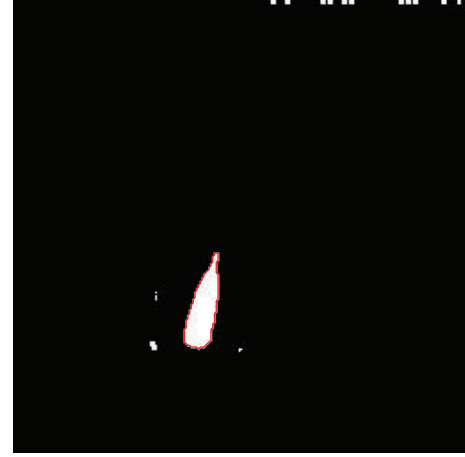
TV-image. The largest M_y values are between about 150 and 200 and they indicate the minimum and maximum vertical axis points of the location of the glottis in TV image.

To extract the glottis region from TV-image, a threshold can be determined using M_x and M_y . Let μ and σ be the mean and standard deviation, respectively, of the union of the values of M_x and M_y . TV-image is converted to a binary image by using $T = \mu + \sigma$ as a threshold. Resulting binary images may have nonzero regions, R_i , other than the glottal region, R_g . Figure 6(a) shows the results of the thresholding for the TV-image shown in Figure 4. One large region corresponding to the glottis and several small regions due to noise are seen in the figure. To get rid of these small regions, average TV values inside the regions can be used. Let E_i be the average of the TV values in R_i ; that is,

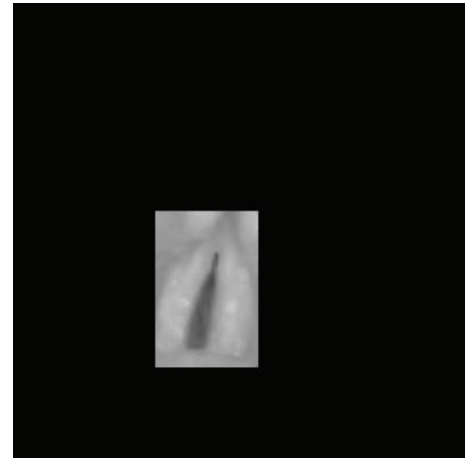
$$E_i = \frac{1}{A_i} \sum_{x,y \in R_i} \text{TV}(x, y), \quad (3)$$

where A_i is the area of the region R_i . By choosing the region having the largest E_i as glottis region the nonzero regions can be eliminated. The index of the glottal region, R_g , is determined as

$$g = \arg \max_i (E_i). \quad (4)$$



(a)



(b)

FIGURE 6: (a) Boundary of the glottal activity region, (b) a masked HSV image.

To locate the mask, the boundary, B_g , of the largest TV-norm region, R_g , is determined by an edge detection algorithm. B_g is a set of (x, y) pairs (coordinate values). To construct the mask, the extrema of the elements of B_g are defined as

$$\begin{aligned} x_{\min} &= \arg \min_x (x, y) \in B_g, & y_{\min} &= \arg \min_y (x, y) \in B_g, \\ x_{\max} &= \arg \max_x (x, y) \in B_g, & y_{\max} &= \arg \max_y (x, y) \in B_g. \end{aligned} \quad (5)$$

Then the mask is defined as follows:

$$\text{mask}(x, y) = \begin{cases} 1, & x_{\min} \leq x \leq x_{\max}, y_{\min} \leq y \leq y_{\max}, \\ 0, & \text{elsewhere.} \end{cases} \quad (6)$$

Masked image is obtained by the product of the original image and the mask $I_{\text{MASKED}}(x, y, n) = I(x, y, n) \cdot \text{mask}(x, y)$. An example of a glottal boundary and the corresponding masked image are shown in Figure 6. The vocal folds are cropped from the HSV image; thus the masked image

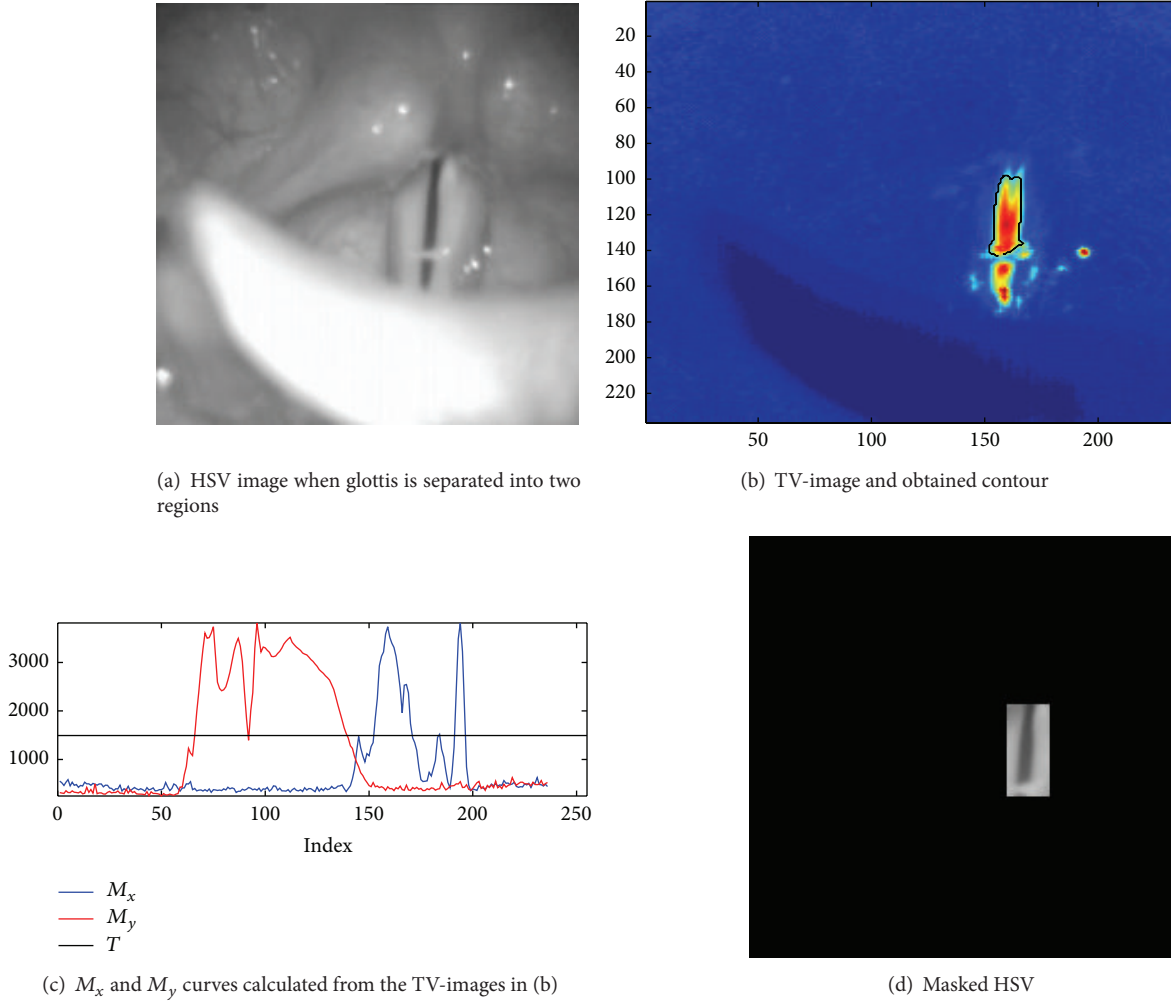


FIGURE 7: Application of the proposed method without 2D median filtering when glottis is separated into two regions.

contains only glottis and its neighbours as desired. This mask is used for processing all 300 frames. The size and location of the mask can be updated for long HSV recordings by processing sliding HSV image blocks containing a number of consecutive frames.

2.1.1. Special Cases

Case 1 (finding ROI when glottis has more than one area). A special case in vocal fold vibration is separation of glottis into more than one vibrating region. An example from the IRCAM database is seen in Figure 7(a). The glottis is separated into two regions. In this special case, the movement of the middle part of vocal folds is limited. The TV-image obtained from Figure 7(a) is shown in Figure 7(b). The total variation is high at both glottal regions but quite small at the connection of the regions. It is also seen in the maxima curves calculated from the TV-image plotted in Figure 7(c). M_y curve has a sharp minimum approximately at $y = 90$ (row = 145, note that y is different than row number and equal to $y = 255$ -row) which corresponds to the location of the connected

edges of vocal folds. This causes the separation of the glottis into two disjoint regions if the TV-image is thresholded by the sum of the mean and the standard deviation of the union of maxima curves. Since in the region selection step one of the regions is selected, calculated mask does not cover the glottis completely as shown in Figure 7(d). This problem can be solved by applying either one or both of the following methods:

- (1) thresholding TV-image by a smaller threshold, for example, $\mu + 0.5\sigma$, or an adaptive method for selection of threshold from a set of candidates, that is, choosing a threshold from the set $T_i = \{\mu + 0.1\sigma, \mu + 0.2\sigma, \dots, \mu + 0.9\sigma, \mu + \sigma\}$;
- (2) 2D median filtering of TV-image.

2D median filtering of TV-images is chosen as a general solution and used in the masking method. The TV-image obtained by 9×9 2D median filtering is shown in Figure 8(a). The TV of the glottal region is distributed over the location of the glottis after median filtering. Furthermore, some of the small regions having large TV resulting from shining of

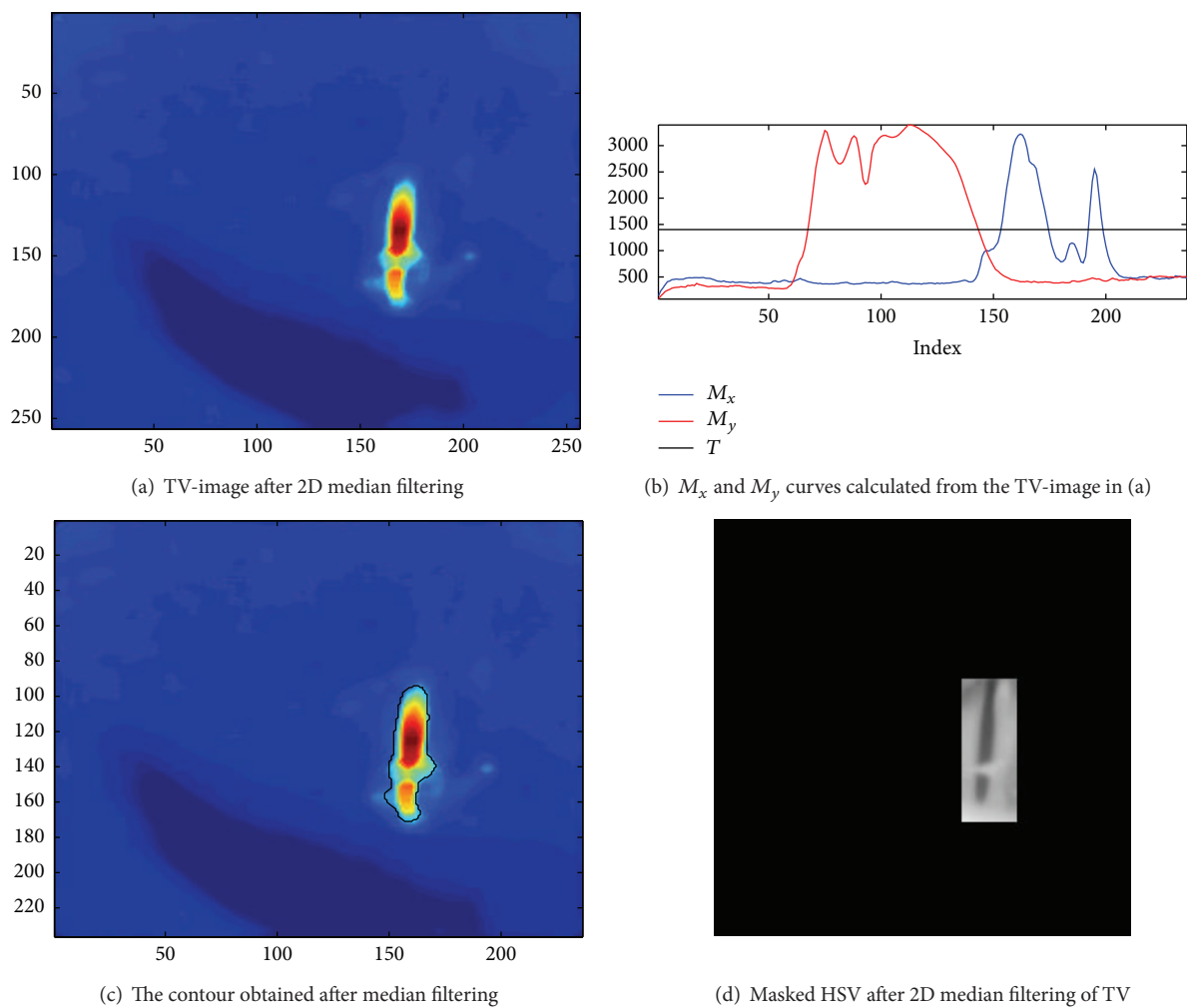


FIGURE 8: Application of the proposed method with 2D median filtering when glottis is separated into two regions.

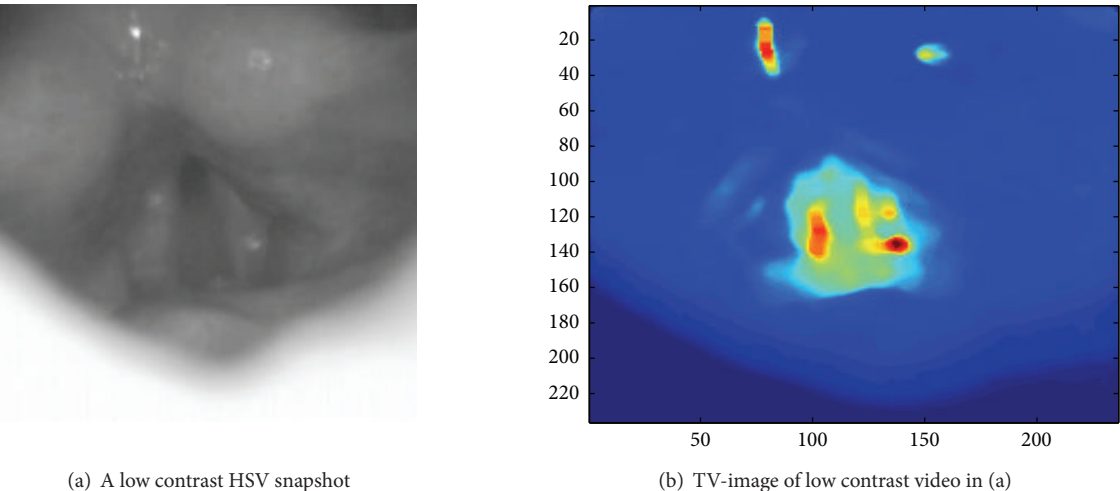


FIGURE 9: A low contrast HSV and its TV-image.

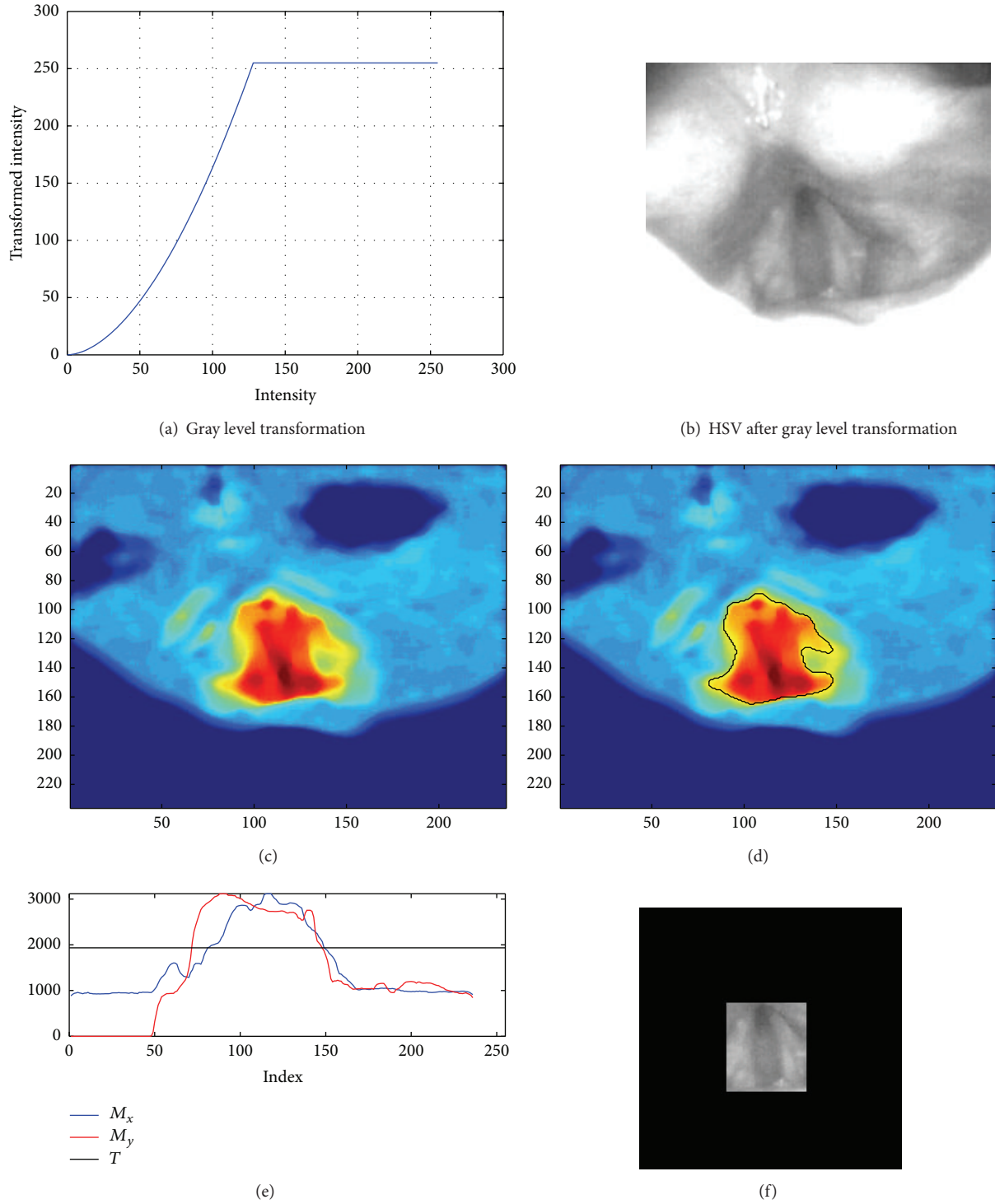


FIGURE 10: Application to a low contrast HSV.

tissues are diminished by the filtering (e.g., region around $x = 200$). Therefore, TV-image is enhanced by 2D median filtering. The maxima curves obtained from median filtered TV-image are plotted in Figure 8(b). Median filtering of the TV-image smooths the minimum in M_y (around $y = 90$) corresponding to the location of the connection of the two

regions at the glottis. The contour of TV-image obtained after thresholding is shown in Figure 8(c). Now, both glottal regions are covered by the contour. The final constructed mask is shown in Figure 8(d). Without using 2D median filter, the masked HSV is erroneous as seen in Figure 7(d). However, by the use of median filtering on TV, the mask

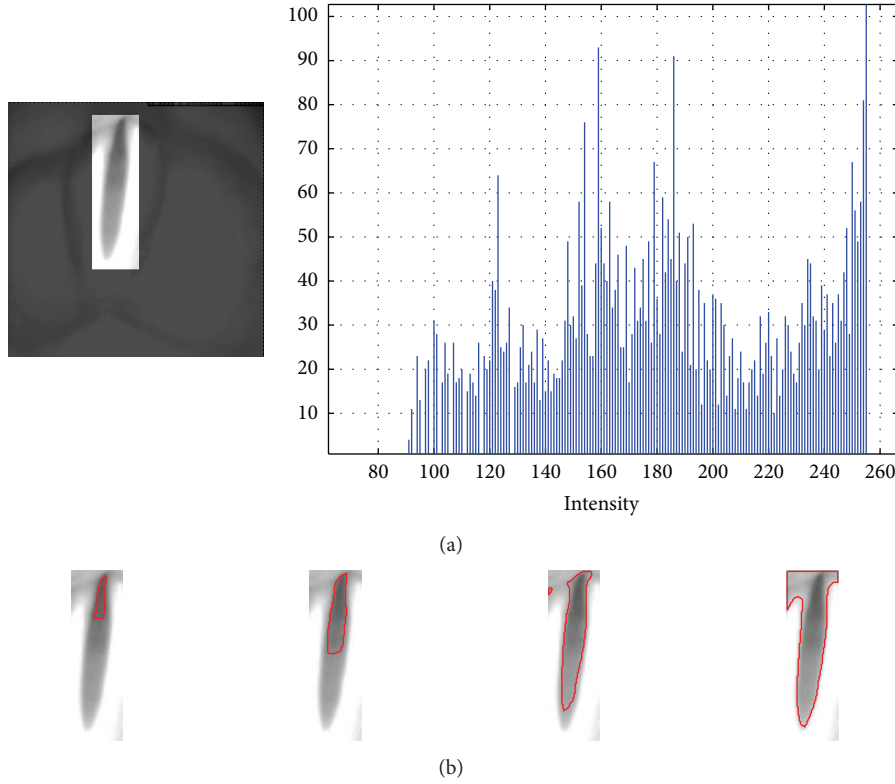


FIGURE 11: (a) Masked HSV image and its histogram. (b) Boundary of the vocal folds in the masked image obtained by thresholding with 110, 140, 170, and 210, respectively.

is corrected and the glottis is completely covered by the mask as plotted in Figure 8(d). In general, using 2D median filter enhances TV-images; hence, it is used in the automatic masking method.

Case 2 (enhancing low contrast images). Low contrast brings some difficulties in automatic masking HSV images. It requires a special approach which is described here. One low contrast HSV image is shown in Figure 9(a). The contrast of the image is relatively less than the previous HSV image examples. When the contrast of consecutive images is reduced, the total variation of intensity of pixels corresponding to moving objects decreases. Therefore, low contrast reduces TV values at the glottal activity region. The TV-image obtained from the image is shown in Figure 9(b). Note that there are two regions having significantly large TV value above the vocal folds. These regions result from the shining points above the vocal folds. The TV values of the glottis are high but distributed in an area about the location of the vocal folds. One solution is to use contrast transformation. The contrast of the HSV is transformed by the following nonlinear transformation:

$$T(I) = \begin{cases} 255 * \left(\frac{I}{I_{\text{High}}} \right)^{1.8}, & I \leq I_{\text{High}}, \\ 255, & I > I_{\text{High}}, \end{cases} \quad (7)$$

where I is the intensity and I_{High} is the top 10% of all pixel values [13].

The transformation and transformed HSV are plotted in Figures 10(a) and 10(b). The transformation not only increases the contrast but also reduces the temporal intensity variations at the bright regions. The TV-image obtained after contrast transformation is shown in Figure 10(c). The transformation enhances the TV-image by reducing the effect of shining tissues due to illumination and increases the TV values of the active glottal region. The estimated contour and maxima curves obtained from the TV-image and resulting mask are shown in Figures 10(d) and 10(f), respectively. The method successfully locates the vocal folds.

3. Modeling Nonuniform Illumination

Histogram based image segmentation relies on the lighting conditions on the scene [13]. Under uniform lighting, or illumination, glottis and background in an HSV image can be separated using a single threshold determined by intensity histogram. However, the intensity distribution of the glottis and background is deformed under nonuniform illumination. It results in such a complex intensity distribution that separation of the glottis and background may not be achieved via simple thresholding. For instance, an HSV image distorted by nonuniform illumination and its histogram are shown in Figure 11(a). Note that, despite the successful

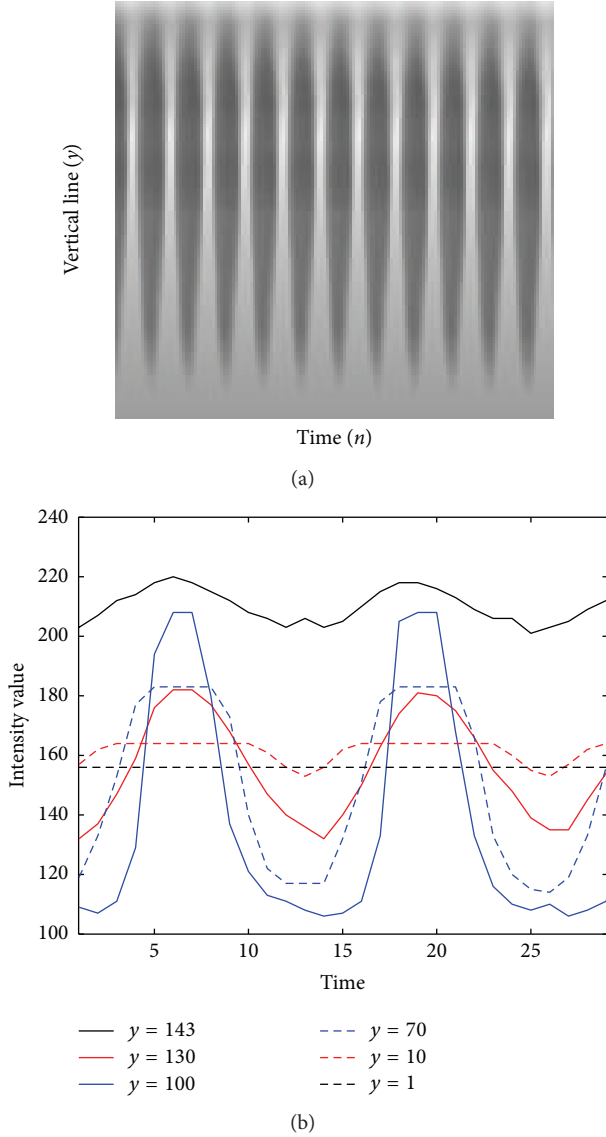


FIGURE 12: (a) Vertical slice image (VSI), intensity variation along the vertical slice of masked image. (b) Intensity variations of VSI image for different vertical lines.

masking over the image, the intensity distribution is far from having a bimodal character as indicated earlier in Section 2. A reliable and accurate threshold cannot be determined from the intensity histogram. For instance, the boundary of the vocal folds obtained by 4 different thresholds corresponding to the local minima of the histogram is shown in Figure 11(b). It is seen that the estimated boundaries are erroneous due to the deformation of the glottis and background intensity distributions.

In this section, we present a novel illumination model for HSV images and a method to estimate the parameters of the model. The method is based on the mean intensity distribution along the longitudinal cross section at the center of the glottis. Vertical slices of masked video frames are used to form, $VSI(n, y) = I_{\text{MASKED}}(x_{\text{center}}, y, n)$, where x_{center} is the horizontal center of the glottis when it is aligned

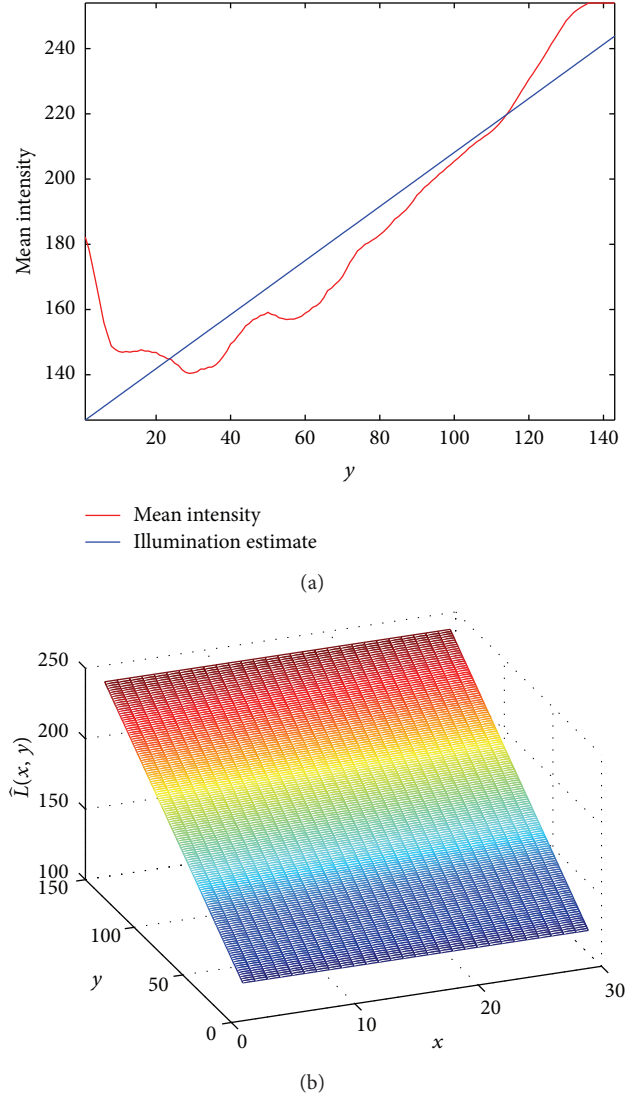


FIGURE 13: (a) Mean intensity along the center vertical line in masked HSV and estimated illumination line. (b) Estimated illumination function, $\hat{L}(x, y)$.

vertically. The vertical slice image, $VSI(n, y)$, obtained from HSV frames plotted in Figure 11(a), is shown in Figure 12(a). It shows temporal intensity variation on the central vertical line of the glottis. The intensity on the vertical slice is low during the glottal opening and high when the folds are closed. Intensity variations for different y values of VSI are shown in Figure 12(b). For $y = 143$ (top line), the intensity is high and has a small variation in time with a small amplitude, and it has high periodicity. The largest variation is observed on $y = 100$ ($VSI(n, 100)$). It is the best representation of the periodic movement of the vocal fold edges. The variation is almost constant and exhibits nonperiodical movement on $y = 1$. This is due to the fact that the variation on $y = 1$ line does not contain information from the glottis region. The variation on the lines, $y = 130$ and $y = 70$, is also periodic with considerably large amplitudes. For $y = 10$, the amplitude variation is small. Note that the mean intensity

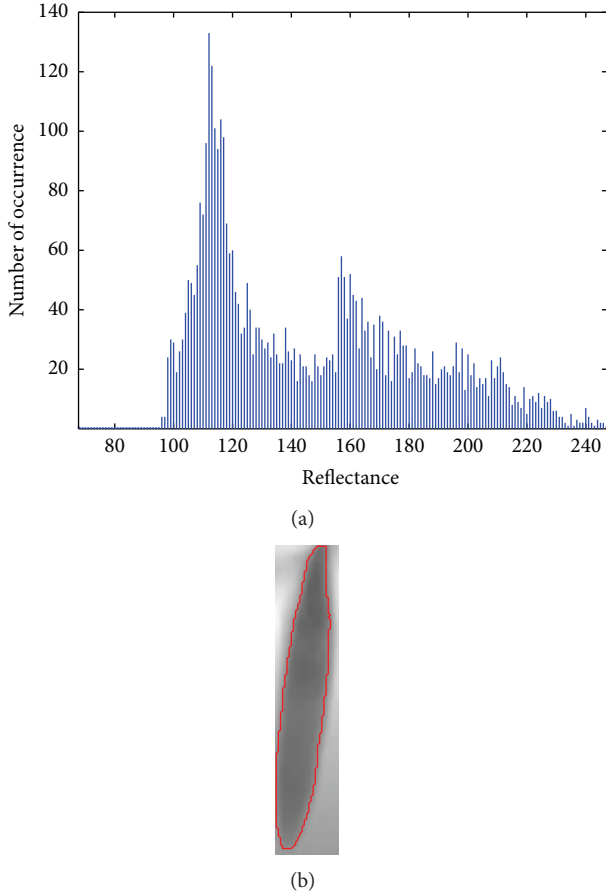


FIGURE 14: (a) Histogram of the reflectance image. (b) Estimated reflectance image and the boundary of the vocal folds obtained by thresholding with $T = 150$ (reflectance is scaled by 255 for grayscale representation).

values on the lines presented in the figure are different and depend on y . The temporal mean of the intensity variation on each horizontal line in the VSI image can be a measure of the illumination on that particular line. The temporal mean of the VSI as a function of y is plotted in Figure 13(a). Note that the mean intensity increases almost linearly from bottom ($y = 20$) to the top ($y = 143$) in the VSI.

The intensity of an image, $I(x, y)$, can be considered as a product of the illumination, $L(x, y)$, and reflectance, $R(x, y)$, the amount of illumination from the objects in the scene ($I(x, y) = R(x, y) \cdot L(x, y)$) [13]. The illumination depends on the energy provided by the light source. The reflectance, $R(x, y)$, is the characteristic of the imaged objects and it is between 0 (total absorption) and 1 (total reflection). To extract the reflectance component, an estimate of the illumination function, $\hat{L}(x, y)$, can be used.

Under uniform illumination ($L(x, y) = L$), intensity is expressed by reflectance times a constant ($I(x, y) = L \cdot R(x, y)$). Furthermore, if a uniformly reflective object ($R(x, y)$ is a constant, R) is imaged under uniform illumination, the intensity, $I(x, y)$, will be a constant. Similarly, if an image sequence is formed by periodic placement and removal of the object at the same location, the mean intensity

variation along the image sequence at each point of the region covered by the object will be the same. As illustrated in this section, the intensities on vertical center line in the glottis are periodic. It is assumed that the vertical center line in the glottis is uniformly reflective ($R(y) = R$). It implies that, under uniform illumination, the mean intensity is the same at each point on the vertical center line. However, under nonuniform illumination, for the vertical center line, the mean intensity is the product of a constant reflectance, R , and the illumination function, $L(y)$. If R is chosen as 1 (for the sake of simplicity), the nonuniform illumination, $L(y)$, will be the same as the mean intensity on the vertical center line. In the same way, the illumination function, $L(x, y)$, can be estimated by temporally averaging the intensity for each pixel location in the glottis. To reduce the computational complexity, we assume that $L(x, y)$ changes only vertically and is the same on each horizontal line ($L(x, y) = L(y)$). The mean intensity is modeled by using a line as shown in Figure 13(a). The nonuniform illumination estimate, $\hat{L}(x, y)$, is obtained by extending the line horizontally as shown in Figure 13(b). The reflectance component of the masked HSV image, $R(x, y)$, can be obtained by dividing the intensity of the masked image, $I(x, y)$, into the illumination estimate, $\hat{L}(x, y)$. To represent the reflectance as a grayscaled image, its values are scaled to be confined to the interval $[0-255]$ by multiplying 255. The estimated reflectance and its histogram are plotted in Figure 14.

The distinction between the glottis and background reflectance distribution can be observed from the clear bimodal structure in Figure 14(a). For discriminating the glottis from the background a reflectance threshold can be selected by visually inspecting the histogram. The reflectance value corresponding to the middle point between the peaks is found to be 150. It is used to convert the reflectance image into binary image and edge detection operation is applied on the binary image to extract the boundary of the vocal folds. Estimated boundary is illustrated in Figure 14(b). The vocal folds' edges are accurately determined by reflectance thresholding. Another nonuniform illumination modeling and reflectance based thresholding example is shown in Figure 15. The intensities of the image are distributed between values 120 and 255. Note that, due to high illumination, a large amount of intensity values is saturated around the upper intensity limit, 255. In this case, the classification of the glottis pixels based on the intensity distribution is problematic. A reliable threshold cannot be found even by visual inspection. However, the reflectance histogram obtained by using the proposed illumination model exhibits bimodal distribution. The result of the segmentation by using the reflectance value corresponding to the minimum between the two highest peaks in the reflectance histogram, $R = 130$, is shown in the upper right panel of the figure. The vocal folds' edges are accurately determined by using reflectance based thresholding.

4. GMM Based VSI Reflectance Modeling

One of the problems in the histogram based HSV processing is segmentation of small glottis regions appearing at the

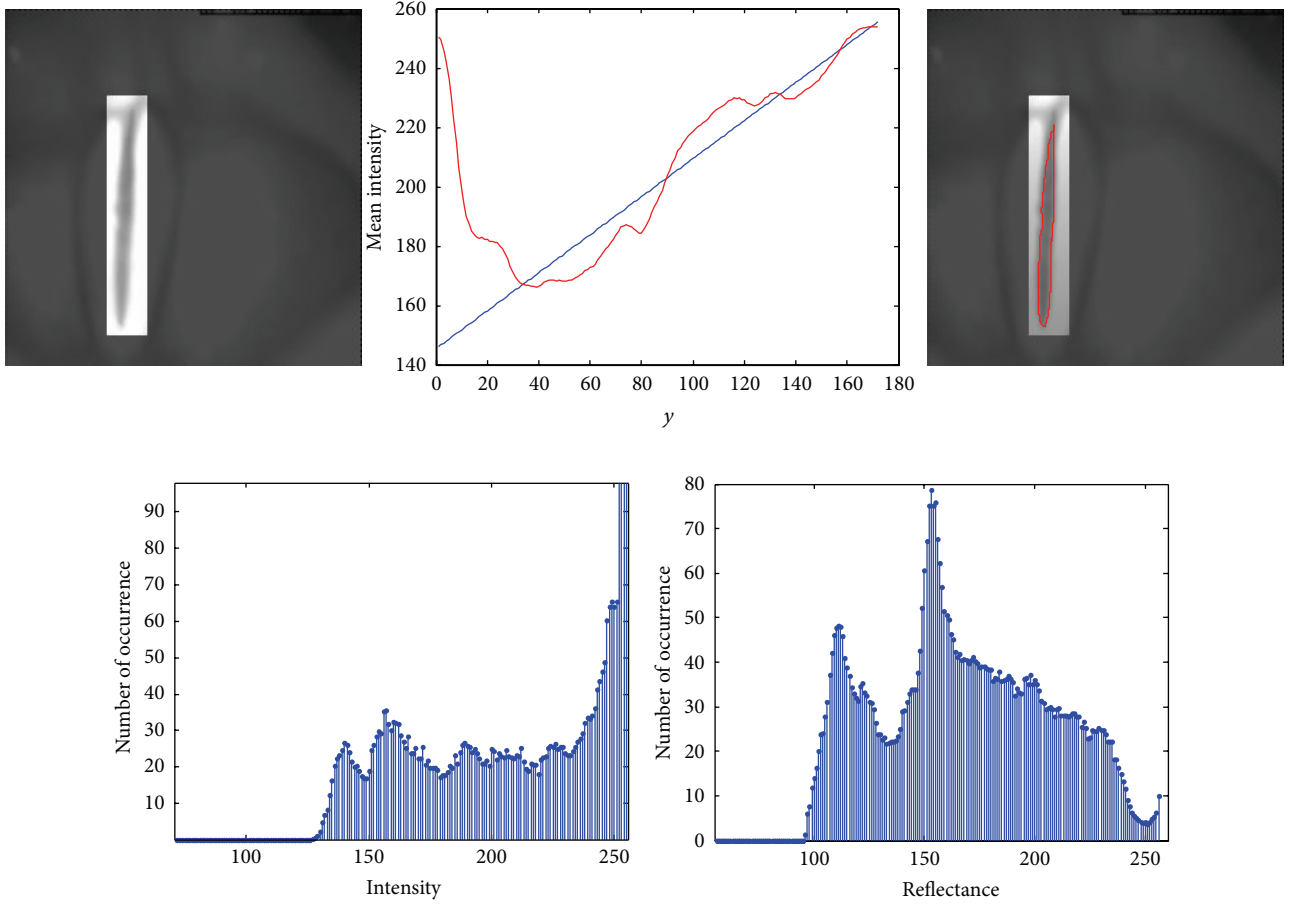


FIGURE 15: Illumination modeling system (reflectance is scaled by 255 for grayscale representation).

beginning of the glottal opening and at the end of the closing phases of vocal folds vibration. As illustrated in the introduction section, the intensity histograms of the images captured during the glottal closing or opening can be unimodal. It is not possible to estimate a reliable threshold even by manual intervention in those cases.

In this section, a novel automatic threshold determination method from consecutive HSV frames is presented. It is shown in the previous section that the reflectance histograms have better modality than the intensity histograms, due to the reduction of the effect of the nonuniform illumination. Therefore, reflectance based image segmentation is preferred in HSV segmentation. The VSI image based on reflectance represents reflectance distribution of the pixels belonging to the glottis and vocal fold edges. A reflection based VSI image is shown in Figure 16(a). The reflectance distribution provided by the vertical slice image is more reliable than that of the original unmasked image in distinguishing the boundary of glottal opening. The reflectance histograms of the unmasked, masked, and VSI images are shown in Figure 16(b). The unmasked image histogram is almost unimodal. Similarly, there is no clear boundary in the masked image histogram and it is not possible to separate the background and glottis easily. However, the histogram of the VSI image has clearly separated two peaks that lead to a robust threshold determination to separate glottis and background.

The VSI reflectance can be represented by a mixture of two Gaussian densities (GMM) [17]. The reflectance distributions at the glottal opening, $P_{R|GO}(r | GO)$, and at other locations, $P_{R|O}(r | O)$, form the two components. $P_{R|GO}(r | GO)$, $P_{R|O}(r | O)$, and $P(R)$ can be written as follows:

$$P_{R|GO}(r | GO) = N(r | \mu_1, \sigma_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{(-1/2)((r-\mu_1)/\sigma_1)^2},$$

$$P_{R|O}(r | O) = N(r | \mu_2, \sigma_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{(-1/2)((r-\mu_2)/\sigma_2)^2},$$

$$P(R) = \sum_{i=1}^2 w_i N(r | \mu_i, \sigma_i), \quad (8)$$

where w_i 's are weights of the Gaussian components. The parameters of the GMM, μ_i , σ_i , and w_i , are estimated by EM (expectation and maximization) algorithm [18]. The posterior probabilities, $P_{GO|R}(GO | r)$ and $P_{O|R}(O | r)$, are written as follows:

$$P(i | R) = \frac{w_i N(R | \mu_i, \sigma_i)}{\sum_{k=1}^2 w_k N(R | \mu_k, \sigma_k)}, \quad (9)$$

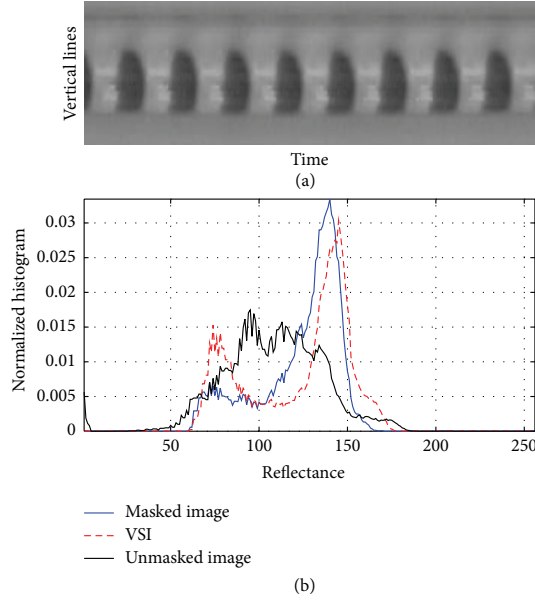


FIGURE 16: (a) Vertical slice image (VSI). (b) Reflectance histograms of unmasked, masked, and VSI images (reflectance is scaled by 255 for grayscale representation).

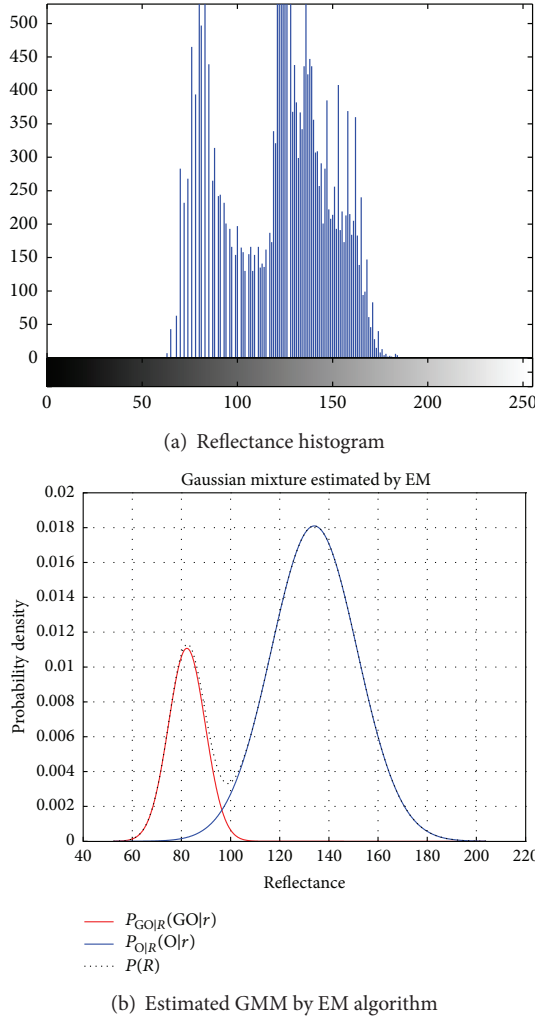


FIGURE 17: Estimation of GMM.

where $i = 1, 2$ denotes glottal opening (GO) and others (O), respectively. A VSI histogram and corresponding GMM are shown in Figure 17. Estimated GMM fits the reflectance distribution well.

The estimated densities are used to segment each HSV frame in the image sequence as

$$g(x, y, n) = \begin{cases} 1, & P(\text{GO} | R(x, y, n)) > P(\text{O} | R(x, y, n)), \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

In the last step of the algorithm, the glottal area variation in HSV is determined. The glottal area in terms of the number of pixels is calculated as

$$a_g[n] = \sum_{x, y \in \text{Mask}} g(x, y, n). \quad (11)$$

The steps of the algorithm are given as follows.

- (1) Get N images from HSV.
- (2) *Start automatic mask determination.*
- (3) Calculate TV-image (see (1)).
- (4) Apply $K \times K$ 2D median filter to TV-image (the smaller K , the sharper TV-image).
- (5) Calculate M_x and M_y (see (2)).
- (6) Threshold for segmentation of TV-image is determined by the union of M_x and M_y statistics, $+\sigma$.
- (7) Segment TV-image and find the regions inside it using connected component method.
- (8) Select the region having largest E (see (3) and (4)) as glottis and determine the boundary of the region by edge detection.
- (9) Construct a mask by using the extrema of the boundary (see (6)).
- (10) *End masking.*
- (11) *Start illumination modeling.*
- (12) Apply masking to image sequence and form VSI image $(\text{VSI}(n, y) = I_{\text{MASKED}}(x_{\text{center}}, y, n))$.
- (13) Calculate the mean intensity at each row in VSI and estimate illumination function $\hat{L}(x, y)$.
- (14) Estimate reflectance images from masked images by using $R(x, y) = I(x, y) / \hat{L}(x, y)$.
- (15) *End illumination modeling.*
- (16) *Start density estimation.*
- (17) Form VSI image of the reflectance image sequence.
- (18) Estimate GMM parameters by EM method and calculate the posterior probabilities, $P_{\text{GO|R}}(\text{GO} | r)$ and $P_{\text{O|R}}(\text{O} | r)$ (see (8)–(9)).
- (19) *End density estimation.*
- (20) *Start segmentation.*
- (21) Segment each image by using the posterior probabilities (see (10)).
- (22) Calculate glottal area (see (11)).
- (23) *End segmentation.*

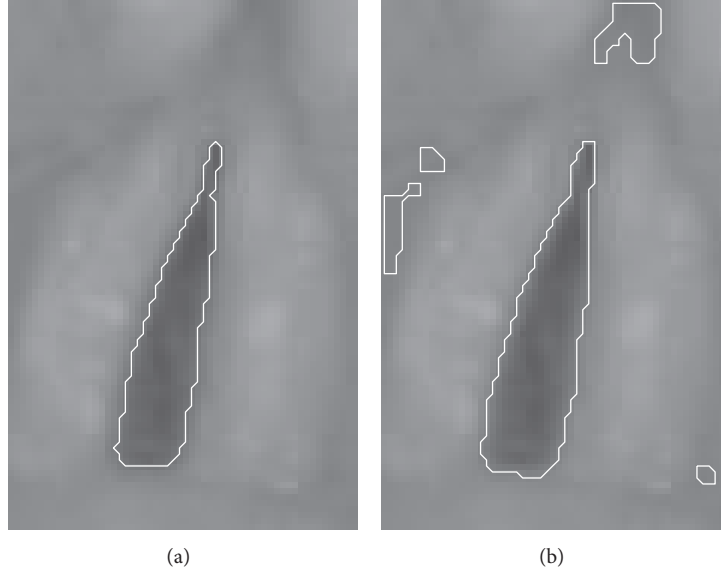


FIGURE 18: Segmentation results. (a) Proposed method. (b) The method of [7].

5. Results

Proposed algorithm is compared with the method of [7] on IRCAM HSV database [19]. In [7], intensities in HSV frames are modeled by Rayleigh distribution; then a threshold is determined by a Bayesian approach. After thresholding, region-growing operation is applied. In this study, region growing is omitted, because it is optional and its success depends on the accurate determination of the glottis region. The core function of [7] is Rayleigh based intensity modeling and thresholding is used in the comparison. However, it is our experience that the application of the method of [7] on unmasked frames frequently produces undesirable segmentation results. It is expected since unmasked images usually have unimodal intensity distribution. Therefore, the method of [7] is applied on the masked images obtained from the first stage of our algorithm. Examples of segmentation results by the proposed method and the method of [7] are shown in Figure 18. Segmented regions other than glottis are observed in the result obtained by the method of [7]. On the other hand, the use of masking and VSI image improves the success of classification significantly. For comparison, 3000 images are segmented from 10 different HSV videos (300 for each) manually to produce a reference data set. Segmentation results for three different cases are shown in Figures 19–21. The glottal area waveforms obtained by manual segmentation, by the method of [7] and the proposed method are shown in Figure 22. Mean square error (MSE) values of the glottal area waveforms obtained by applying the two methods are given in Table 1.

According to the results presented in Table 1, the MSE of the proposed algorithm is smaller than the MSE of the method of [7] by about 94%. The accuracy in the glottal area and vocal fold boundary estimates is increased significantly.

TABLE 1: MSEs over 3000 frames of the estimated glottal area waveforms.

| | Proposed | The method of [7] | Decrease in MSE (%) |
|-----|----------|-------------------|---------------------|
| MSE | 5969.9 | 93573 | 93.60 |

6. Conclusion

Automatic segmentation of HSV images is necessary for investigation of vocal folds' functions in clinical and speech production studies. In this study, an automatic method is introduced for extracting glottal area waveform from HSV images. The first novelty in the method is that the region of interest is determined automatically by a robust masking algorithm using TV-norm of consecutive images which produces a strong indication for moving vocal folds. Using a sliding HSV block, it can be updated easily for processing long sequence of frames. A planar illumination model for HSV images is the second unique feature of the method. By means of the proposed planar illumination model, reflectance images can be obtained and used in the vocal fold segmentation. The other novelty is, by introducing vertical slice image, the use of reflectance variation information in the glottis from multiple images instead of using a single image to model the reflectance distribution. Modeling the reflectance distribution of vertical slice images by GMM is easy, computationally efficient and produces more accurate glottal area waveforms. For long image sequences, the mean and variance of GMM can be recursively updated by forming a sliding frame sequence framework in constructing the VSI image. The proposed approach can be used in any histogram based automatic image segmentation systems.

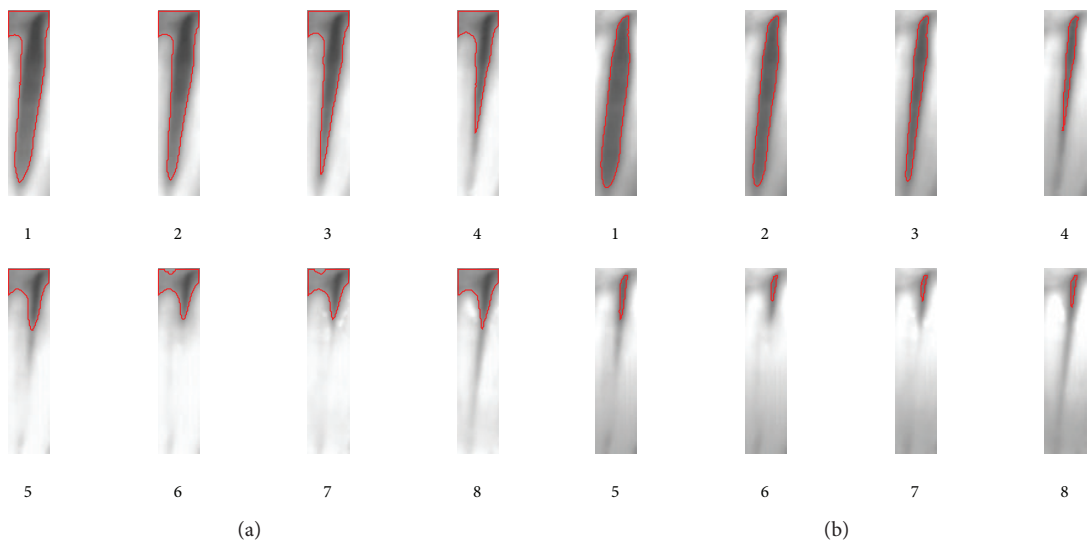


FIGURE 19: Case 1: Segmented HSV image sequence by the method of [7] (a), the proposed method (b).

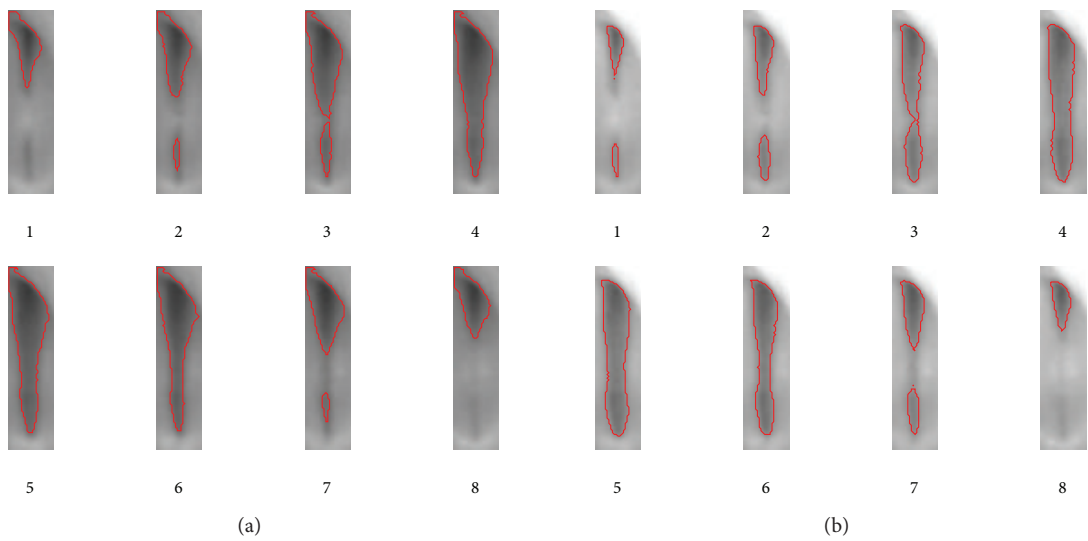


FIGURE 20: Case 2: Segmented HSV image sequence by the method of [7] (a), the proposed method (b).

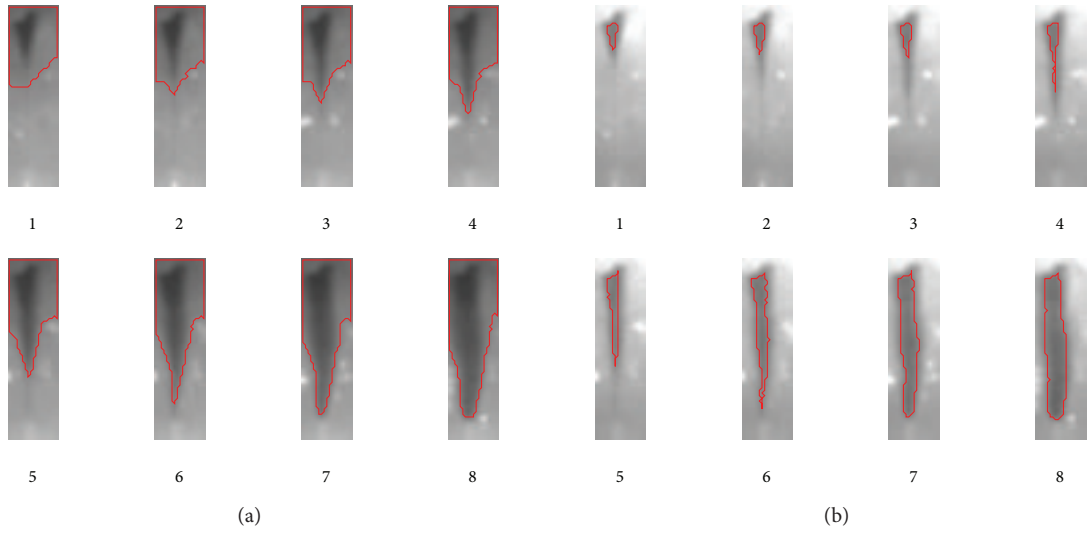


FIGURE 21: Case 3: Segmented HSV image sequence by the method of [7] (a), the proposed method (b).

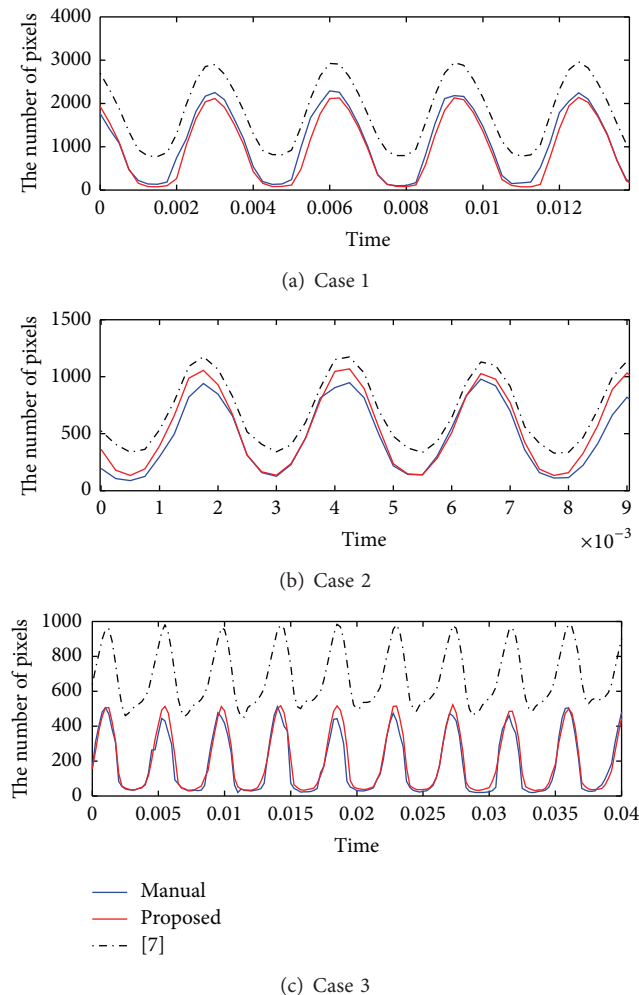


FIGURE 22: Manually extracted and automatically estimated glottal area signals Cases 1–3.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

Special thanks are due to Gilles Degottex and Erkki Bianco for sharing their valuable IRCAM HSV database.

References

- [1] D. D. Mehta, D. D. Deliyski, S. M. Zeitels, T. F. Quatieri, and R. E. Hillman, "Voice production mechanisms following phonosurgical treatment of early glottic cancer," *Annals of Otolaryngology, Rhinology and Laryngology*, vol. 119, no. 1, pp. 1–9, 2010.
- [2] Y. Yan, K. Ahmad, M. Kunduk, and D. Bless, "Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology," *Journal of Voice*, vol. 19, no. 2, pp. 161–175, 2005.
- [3] A. Verikas, A. Gelzinis, D. Valincius, M. Bacauskiene, and V. Uloza, "Multiple feature sets based categorization of laryngeal images," *Computer Methods and Programs in Biomedicine*, vol. 85, no. 3, pp. 257–266, 2007.
- [4] J. Unger, M. Schuster, D. J. Hecker, B. Schick, and J. Lohscheller, "A multiscale product approach for an automatic classification of voice disorders from endoscopic high-speed videos," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '13)*, IEEE, 2013.
- [5] C. Tao, Y. Zhang, and J. J. Jiang, "Extracting physiologically relevant parameters of vocal folds from high-speed video image series," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 794–801, 2007.
- [6] X. Qin, S. Wang, and M. Wan, "Improving reliability and accuracy of vibration parameters of vocal folds based on high-speed video and electroglottography," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 6, pp. 1744–1754, 2009.
- [7] Y. Yan, X. Chen, and D. Bless, "Automatic tracing of vocal-fold motion from high-speed digital images," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1394–1400, 2006.
- [8] Y. Yan, G. Du, C. Zhu, and G. Marriott, "Snake based automatic tracing of vocal-fold motion from high-speed digital images," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '12)*, IEEE, 2012.
- [9] T. Ikuma, M. Kunduk, and A. J. McWhorter, "Preprocessing techniques for high-speed videoendoscopy analysis," *Journal of Voice*, vol. 27, no. 4, pp. 500–505, 2013.
- [10] D. D. Mehta, D. D. Deliyski, T. F. Quatieri, and R. E. Hillman, "Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 1, pp. 47–54, 2011.
- [11] J. Lohscheller, H. Toy, F. Rosanowski, U. Eysholdt, and M. Döllinger, "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos," *Medical Image Analysis*, vol. 11, no. 4, pp. 400–413, 2007.
- [12] S.-Z. Karakozoglou, N. Henrich, C. D'Alessandro, and Y. Stylianou, "Automatic glottal segmentation using local-based active contours and application to glottovibrography," *Speech Communication*, vol. 54, no. 5, pp. 641–654, 2012.
- [13] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2002.
- [14] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [15] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.
- [16] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.
- [17] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2nd edition, 2001.
- [18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B. Methodological*, vol. 39, no. 1, pp. 1–38, 1977.
- [19] G. Degottex and E. Bianco, "IRCAM Databases of High Speed Videoendoscopy," UPMC-Ircam, France, 2010.

Research Article

A Novel Two-Stage Spectrum-Based Approach for Dimensionality Reduction: A Case Study on the Recognition of Handwritten Numerals

Mohammad Amin Shayegan,¹ Saeed Aghabozorgi,² and Ram Gopal Raj³

¹ Image Processing and Pattern Recognition Research Lab, R&D Center, Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

² Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

³ Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

Correspondence should be addressed to Mohammad Amin Shayegan; mashaygan@siswa.um.edu.my

Received 29 January 2014; Accepted 11 April 2014; Published 12 May 2014

Academic Editor: Feng Gao

Copyright © 2014 Mohammad Amin Shayegan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dimensionality reduction (feature selection) is an important step in pattern recognition systems. Although there are different conventional approaches for feature selection, such as Principal Component Analysis, Random Projection, and Linear Discriminant Analysis, selecting optimal, effective, and robust features is usually a difficult task. In this paper, a new two-stage approach for dimensionality reduction is proposed. This method is based on one-dimensional and two-dimensional spectrum diagrams of standard deviation and minimum to maximum distributions for initial feature vector elements. The proposed algorithm is validated in an OCR application, by using two big standard benchmark handwritten OCR datasets, MNIST and Hoda. In the beginning, a 133-element feature vector was selected from the most used features, proposed in the literature. Finally, the size of initial feature vector was reduced from 100% to 59.40% (79 elements) for the MNIST dataset, and to 43.61% (58 elements) for the Hoda dataset, in order. Meanwhile, the accuracies of OCR systems are enhanced 2.95% for the MNIST dataset, and 4.71% for the Hoda dataset. The achieved results show an improvement in the precision of the system in comparison to the rival approaches, Principal Component Analysis and Random Projection. The proposed technique can also be useful for generating decision rules in a pattern recognition system using rule-based classifiers.

1. Introduction

Pattern recognition (PR) is one of the most attractive branches in the artificial intelligence field. In all PR systems, the quantity, quality, and diversity of training data in the learning process directly affect the final result. The training dataset size has double importance in this regard, because the training phase—related to large datasets—is often a time-consuming process.

Nowadays, the emergence of the big-data issue has resulted in specific attention on size reduction and also

dimensionality reduction to save time and memory usage. In addition, the demand for employing various applications running on limited-speed and limited-memory devices, such as mobile phones and mobile scanners, is growing dramatically [1]. Hence, the necessity to find efficient techniques for reducing the volume of data in order to decrease the overall processing time and also the memory requirements is considered more important than in the past.

Two general approaches are utilized for the dataset volume reduction in the literature: size reduction and dimensionality reduction.

The available size reduction techniques try to reduce the number of objects or observations in a dataset. They find and remove two groups of samples from a dataset: samples far from a class centroid (outlier samples or support vector samples) [2–5] and samples close to each class centroid (e.g., using k -means clustering technique) [6]. However, the outlier and support vector samples are necessary to evaluate system efficiency and system functionality. In addition, the samples close to a class centroid include important information about the various characteristics of a class and are necessary to make a system model.

The dimensionality reduction techniques have been used to identify and remove the less important features, extracted from the dataset samples. They are widely employed in different areas, such as biological data clustering [7], EMG signal feature reduction [8], face recognition [9], blog visualization reduction [10], and gene expression databases reduction [11]. Principal Component Analysis (PCA), Singular Value Decomposition (SVD), and Random Projection (RP) are some of the well-known methods in this part [12]. Concerning the dimensionality reduction in the Optical Character Recognition (OCR) applications, PCA has been used to compress the features space in the numeral part of the CEDAR database [13], MNIST database [14], and also the Tamil handwritten character classes [15]. Nevertheless, finding an optimal, effective, and robust feature set out of a large initial extracted feature is usually a heuristic and hard task [16].

This paper contributes to the corpus of knowledge in dimensionality reduction as follows: (1) introducing a new preprocessing method to stick disconnected parts of an image and (2) proposing a novel two-stage spectrum-based method to reduce the number of features in the feature space. The mentioned method is based on analysing one-dimensional and two-dimensional spectrum diagrams for standard deviation and minimum to maximum distribution corresponding to the existing features in the initial feature vector. Unlike the other available techniques for dimensionality reduction, such as PCA, the proposed method can keep every feature in the final feature vector, based on some characteristics of a specific feature or even based on user opinion. To investigate the efficiency of the proposed dimensionality reduction technique, two large standard benchmark handwritten OCR datasets—MNIST and Hoda—are employed. The empirical results have shown the effectiveness of the proposed method. Although the results have been reported for OCR application databases, the salient point of the proposed approach is that it can also be used for other datasets with numerical feature vectors.

The rest of the paper is organized as follows. Section 2 discusses the background of the research topic and introduces the related works for dimensionality reduction operation. Thereafter, Section 3 presents the proposed dimensionality reduction technique, including the definition of one- and two-dimensional standard deviations and also minimum to maximum spectrum diagrams. Section 4 presents the research method and the experimental results and comparisons, and, finally, Section 5 presents the conclusion of the paper.

2. Background and Related Works

2.1. Feature Extraction in OCR Application. Among the various stages in PR systems, feature extraction plays a vital role in building system models, the recognition process, and system accuracy [17]. Feature extraction is a task to detect/extract the maximum amount of various desired attributes and characteristics from the input data. The features are the information that is fed to the recognizer to build a system model [18]. They should be insensitive to irrelevant variability in the input as much as possible, limited in number to permit for effective computation of discriminant functions and should not be similar, redundant, or repetitive. Various kinds of features can be found and/or calculated for an object in a PR system. Usually, features are categorized into global transformations [19], structural [20], statistical [21], and template-based matching [22].

The structural features describe the geometrical and topological characteristics of patterns using their global and local properties [23]. They are the most popular features investigated by researchers in OCR systems, because they are intuitive aspects of writing [24]. Structural features are less influenced by sources of distortion, but they are highly dependent on the style of writing. They may be extracted from each row, column, skeleton, or contour of an image. However, extracting this kind of features from the pattern images is not usually an easy task.

The statistical features are derived from the statistical distribution of the image's pixels and describe the characteristic measurements of a pattern. They include numerical values computed from a part or a whole of an image. Although these features are easy to extract, they can lead the system to the wrong way, because most of them are very sensitive to noise, scale, rotation, and other changes in the patterns.

Transformation features are derived from the transformed representation of an image in a new space by transformation operators. The transformation process maps an image from one space to another space, and it usually reduces the dimensionality and order of computing in the new space. They provide features that are invariant to global deformation like translation, dilation, and rotation [24].

Template-based features are usually created by matching predefined templates on graphical input data. However, they are completely data dependent, and they cannot transfer from a PR system to another.

Some of the most used structural features [14, 17, 20, 25, 26], statistical features [18, 21, 27, 28], and transformation features in OCR applications are shown in Table 1.

2.2. Dimensionality Reduction in OCR Application. Various features are computed and/or extracted in the feature extraction block of an OCR system. However, it is possible that some of the extracted features correspond to very small details of the patterns or that some of them are a combination of other features (nonorthogonal features), while others might not have any efficacy in the recognition stage. Irrelevant or redundant features may degrade the recognition results, reduce the speed of learning algorithms, and significantly increase the time complexity of the recognition

TABLE 1: Some of the most used features in OCR applications.

| Type of features | Some of the most used features |
|------------------|---|
| Structural | <ul style="list-style-type: none"> (i) Simple, double, and complex loops (ii) Loops positions, their types, and their relative locations (iii) Hills and valleys (iv) Open curves in different directions (v) Ascenders and descenders (vi) Numbers and locations of dots in each character (vii) Location of dots relevant to baselines (viii) Starting, ending, branching, crossing, turning, and corner points in image skeleton (ix) Curvature and length of the image's segments (x) Length of a character segment relative to other segments (xi) Location of a character segment relative to center of mass image's skeleton |
| Statistical | <ul style="list-style-type: none"> (i) Normalized central, Zernike, pseudo Zernike, fast Zernike, Legendre, orthogonal Fourier-Mellin, rotational, and complex moments extracted from the whole body, only from the main body, or only from the secondary parts of an image (ii) Fourier and gradient descriptors (iii) Image area and image perimeter (iv) Pixels distribution in left, right, up, and down halves of the image (v) Image density (vi) Aspect Ratio (vii) Mean, mode, variance, and 2D standard deviation (viii) Average and variance of X and Y changes in portions of the image skeleton (ix) Local maxima points in horizontal and vertical projection histograms (x) Ratio of horizontal variance histogram to vertical variance histogram (xi) Ratio of upper half variance to lower half variance of an image (xii) Thinness ratio (xiii) The ratio of pixel distribution between two or more parts of the image (xiv) Center of mass (COM) (center of gravity) (xv) Centroid distance (xvi) Radial coding (xvii) Top, bottom, left, and right profile histograms (xviii) Number of horizontal (row) and vertical (column) transitions (xix) Number of modified horizontal and vertical transitions (xx) Outer and inner contour directional chain code histograms (xxi) Normalized contour chain code descriptors (xxii) Modified contour chain code descriptors (xxiii) Fractal, shadow code descriptors (xxiv) Energy of original image (xxv) Number of specific points such as end, branch, and cross points (xxvi) Number of connected components (xxvii) Relative location of start and end points of an image skeleton (xxviii) Pen width and line height (xxix) Baselines positions (xxx) Histogram of slopes along contour (xxxi) Skeleton-based N-degree directional descriptors |
| Transformation | <ul style="list-style-type: none"> (i) M-band packet wavelet coefficients (ii) Fourier, DCT, and radon coefficients |

process [29]. Hence, following the feature extraction process, the issue of dimensionality reduction (feature selection) arises.

Feature selection is typically a search problem to find an optimal subset with m features out of the original M features, to exclude irrelevant and redundant features from the initial feature vector. It reduces problem dimensionality, reduces system complexity and processing time, and increases the system accuracy [30]. In this respect, some feature subset

selection algorithms have been proposed. According to the criterion function used for finding one m member subset out of 2^M possible subsets (M is the number of initial features), two general categories are introduced for this important task: Wrapper algorithm and Filtering algorithm. In the Wrapper algorithm, the classifier performance is used to evaluate the performance of a feature subset. In the Filtering algorithm, the features evaluation function is used rather than optimizing the classifier performance. In this category, by

using a special feature evaluation function, the best individual features are found one by one. However, the m best features are not the best m features [31]. Usually, the Wrapper methods are slower but perform better than the Filtering methods [32].

Based on the removing strategies, the feature selection methods are categorized into three groups. The first category is the Sequential Backward Selection (SBS) technique [33]. In this approach, the features are deleted one by one and the systems' performance is measured to determine feature performance. However, finding the correct sequence to delete the features one by one is also very important. This means that a system's derived efficiency after deleting features A , B , and C is not equal to the same system's derived efficiency after deleting the features in order A , C , and B or B , C , and A , and so on. Due to their nature, some features are relevant to others from different points of view. In this case, the SBS technique does not help to find the best subset of features. El-Glaly and Quek [33] extracted 4 set features S_1 , S_2 , S_3 , and S_4 to use them in an OCR application. They trained an OCR system with those 4 feature sets separately. After that, sets S_1 to S_4 were delivered to a Principal Component Analysis (PCA) algorithm, which rearranged the features in each set, based on their importance in the recognition system. The results showed that a feature (e.g., X) in rank 23 in set S_3 (23rd feature in feature vector S_3 after applying PCA) took rank 7 in set S_1 (7th feature in feature vector S_1 after applying PCA), and so on. This experiment indicates that if feature X is deleted for the sake of feature reduction, it may cause a large error in the final results.

The second group of feature selection methods comprises the random search methods, such as Genetic Algorithms (GA), which keep a set of the best answers in a population. Bahmani et al. [34] used GA in a handwritten OCR system. The initial number of features in their proposed system was 81 and the accuracy was 77%. After applying GA, the number of features was reduced to 55 and the accuracy increased from 77% to 80%. Soryani and Rafat [35] employed GA to carry out feature subset selection in a typical printed OCR system. They tested the proposed method to identify 5 fonts and 5 sizes of printed alphabet characters, achieved a reduction in the number of features from the initial 256 loci features to 146, and enhanced the system accuracy by 4.07% as well. The GA methods always select chromosomes one by one with the best recognition percentage and move those chromosomes (features) to the next stage. However, it is possible that when a good characteristic feature gets combined with another feature, the overall performance is not as good as the individual performances.

The third method for feature selection operation is a group of methods that have been applied to find important patterns in high-dimensional input data such as Principal Component Analysis (PCA) [12] and Random Projection (RP) [36]. PCA is a statistical method that tries to convert a correlated features space into a new uncorrelated features space. In the new space, features are reordered in decreasing variance value such that the first transformed feature accounts for the most variability in the data. Hence, PCA overcomes the problem of high dimensionality and colinearity [37]. RA technique is a simple powerful dimension

reduction technique that uses Random Projection matrices to map the data from a huge-dimensional space to a lower space one [36]. To achieve this aim, a mapping matrix R is used, where the columns of R are realizations of independent zero-mean normal variables, scaled to have unit length.

Some researchers employed PCA for various applications in PR systems. Gesualdi and Seixas [38] used PCA for data compression in the feature extraction part of a licence plate recognition system. They reduced the number of features from 30 to 4. The reported results showed that the achieved accuracy for digits recognition was acceptable but that the accuracy for characters recognition was degraded significantly. Ziaratban et al. [39] extracted a set of feature points including terminal, two-way, and three-way branch points from the skeletons of the characters in an OCR application. Finally, each skeleton was decomposed into some primitives, which are curved lines between any two successive feature points. Since the number of primitives varies from character to character, they used PCA algorithm to reduce and equalize the lengths of the feature vector. Using a postprocessing stage, they achieved 93.15% accuracy on a dataset with 7647 test samples. To recognize isolated handwritten Arabic letters, Abandah et al. [16] extracted 95 features from all feature categories. After that, only the first 40 features were selected from the PCA process result. Finally, five different classifiers were employed and, on average, 87% accuracy was achieved in the best case. Although PCA has been widely utilized in different PR applications, it suffers from high computational cost. In addition, after applying PCA, finding the order of the most effective to least effective features in the initial feature vector is not possible.

3. The Proposed Model

Figure 1 depicts the overall structure of the proposed dimensionality reduction model. First of all, some common preprocessing operations were carried out on the input images to enhance the quality of the input patterns. Then, the most used features—in the literature—were extracted as the initial feature vector denoted by *Initial_S*. Dimension reduction was taken in two stages. In the first stage, using the proposed tools, One-Dimensional Standard Deviation (1D.SD) and One-Dimensional Minimum to Maximum (1D.MM) spectrums, the number of features was decreased from n in the *Initial_S* to k in the first reduced version of feature vector S_1 . Then, in the second stage, and by employing the proposed tools Two-Dimensional Standard Deviation (2D.SD) and Two-Dimensional Minimum to Maximum (2D.MM) spectrums, the number of features was decreased again from k in feature vector S_1 to p in the final reduced version of feature vector S_2 . The operations mentioned are demonstrated in the following subsections.

3.1. Preprocessing. The performance of an OCR system depends very much upon the quality of the original data. In this context, we took into consideration that the proposed algorithm should be nonsensitive with respect to the scaling, rotation, and transformation of patterns. Hence, some

important preprocessing operations, such as noise removal, dimension normalization, and slant correction using common powerful techniques, are first performed on the samples.

We applied a median filter with a 3×3 window and also morphological opening and closing operators using dilation and erosion techniques for high-frequency noise removal [23]. The image size was normalized without making any changes to the image Aspect Ratio, and, as a result, the width or height (or both) was changed to 50 pixels and the image was located in the center of a 50×50 pixels bounding box.

The body of every English and Arabic/Farsi digit is constructed using only one component. Thus, after the preprocessing operations, if it is found that there is still more than one group of connected pixels in the image of the digit, the extra blocks are considered as noise or separate components of the initial image. To find and remove the rest of the noise, the pen width is estimated using three different methods, and then the average of those values is considered to be the final pen width. To achieve this, we compute

- (a) the mode of image vertical projection;
- (b) (the value of image density)/(the number of image skeleton pixels);
- (c) $\{((\text{the value of image density})/(\text{the number of image outer profile pixels})) * 2.$

The results of the experiments showed that the average of three values is a more accurate estimate of the pen width than each of the values alone. After finding the pen width, all small components with a pixel density that is less than two times of the pen width can be considered to be noise and can be deleted from the input image. The threshold of 2 was obtained experimentally. The rest of the connected components are considered as broken parts of the digit image.

In order to connect the broken image segments together, we used a new approach. By using connected component analysis, we named the biggest available part as the main part M of the image. The outer contour of the main part M was then extracted and the coordinates of its pixels were saved in array $MAIN$. Thereafter, for all of the rest secondary components S_i (which are smaller than the main part M), we found the outer contour and saved the pixels coordinate of those outer contours in another array SEC . Then, we computed the Euclidean distance between all elements of array $MAIN$ with all elements of array SEC . The smallest value of the computed distance indicates the shortest path between contour M and one of the secondary contours S_k . Finally, we drew a line with thickness equal to the estimated pen width along the shortest path between M and S_k . As a result, the main part M is connected to a secondary part S_k . This process was repeated until there is not another secondary component. A new version of main part M is used in each iteration of the algorithm, because, in each iteration, one secondary part is connected to the old version of main part. Algorithm 1 demonstrates the pseudocode for this process.

The images in the Hoda dataset are in binary format, while the images in MNIST are in grey level format. Hence, by analysing the grey level histogram for the image and using the standard global Otsu's method [23], the MNIST samples

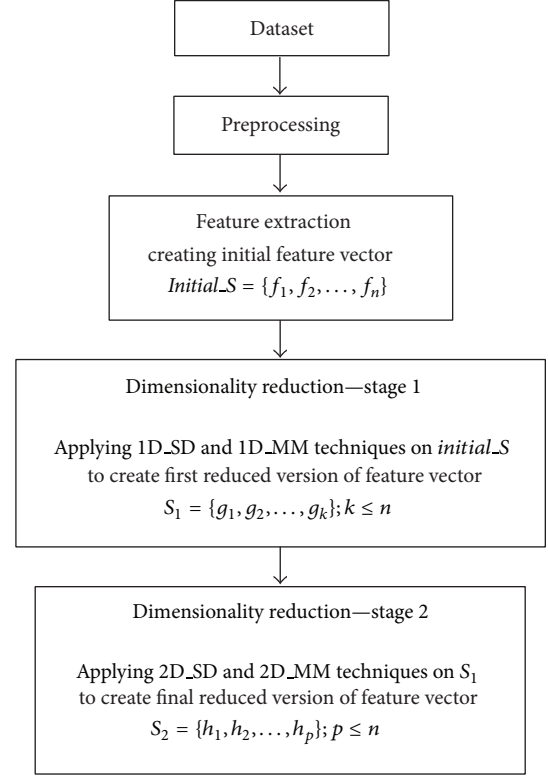


FIGURE 1: The 2-stage proposed model for dimensionality reduction (feature selection).

were also changed to bilevel images. The method proposed by Hanmandlu et al. [40] was used to correct the slant angle of each image. First, an image is divided into upper and lower halves. Afterwards, the centres of mass points for these two parts are calculated. The slope of a line which connects these two mass point centres is considered to be the slant angle and the image is rotated in the reverse direction of this value.

We applied the proposed method on the digits part of the MNIST dataset [41] and the Hoda dataset [42] to connect the broken parts of the digit images. The results were encouraging as we were able to achieve 95.11% and 97.16% successful connections for the MNIST and the Hoda datasets, respectively. Figures 2(a) and 2(b) show examples of the above-mentioned preprocessing operations on two sets of training digits from the Hoda and MNIST datasets, in order.

3.2. Feature Extraction. Due to the vast diversity in writing style, handwritten characters are placed in the high-dimensional dataset category. Hence, finding an optimal, effective, and robust feature set to utilize in the recognition phase of an OCR system is usually a complex task. In this research, based on the literature, an initial feature vector, $Initial_S$, including 133 of the most used features for both English and Arabic/Farsi digits, was extracted from the input images. Some of the extracted features are

- (i) Aspect Ratio;
- (ii) image area, perimeter, diameter, extent, eccentricity, and solidity;


```

while (there is another secondary component in input image) do
{
    find outer contour of the main part  $M$ ;
    save the pixels coordinate of  $M$  in array  $MAIN$ ;
    repeat
    {
        find outer contour of an image secondary part  $S$ ;
        save the pixels coordinate of  $S$  in array  $SEC$ ;
    } until (there is not another secondary parts in image);
    for (each pixel  $A$  in array  $MAIN$ )
    {
        for (each pixel  $B$  in array  $SEC$ )
        {
            compute the distance  $d$  between pixels  $A$  and  $B$ ;
            save ( $d$ , coordinate of pixel  $A$ , coordinate of pixel  $B$ ) in array  $D$ ;
        }
    }
     $d\_min$  = smallest value  $d$  in array  $D$ ;
     $A\_min$  = coordinate of pixel  $A$ , corresponding to  $d\_min$ ;
     $B\_min$  = coordinate of pixel  $B$ , corresponding to  $d\_min$ ;
    draw (a straight line with pen_width thickness from  $A\_min$  to  $B\_min$ );
}

```

ALGORITHM 1: The proposed procedure to connect the broken image segments.

| Initial shape | Enhanced shape | Initial shape | Enhanced shape |
|---------------|----------------|---------------|----------------|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

(a) Arabic/Farsi digits

(b) English digits

FIGURE 2: Applying proposed connecting procedure on input images.

- (iii) Euler's number;
- (iv) centre of mass (COM) and centroid distance;
- (v) pixel distribution density in up, down, left, and right halves of the normalized image;

- (vi) pixel distribution density in upper and lower main diagonals of the normalized image;
- (vii) ratio of pixel distribution in different quarters of the normalized image to each other;
- (viii) ratio of horizontal variance histogram to vertical variance histogram;
- (ix) ratio of upper half variance to lower half variance of an image;
- (x) normalized horizontal and vertical transitions;
- (xi) maximum horizontal and vertical crossing counts;
- (xii) average of multiplication distances X and Y from COM;
- (xiii) average of distances X and Y from boundary;
- (xiv) ratio of major to minor axes lengths;
- (xv) convex area;
- (xvi) number and location of start, end, branch, corner, and crossing points in image skeleton;
- (xvii) normalized invariant moments to order 7;
- (xviii) discrete cosine transform coefficients related to the main image to order 9;
- (xix) top, down, left, and right concavities in the image skeleton;
- (xx) number of modified horizontal and vertical transitions;
- (xxi) average distance and average angular distance of each foreground pixel in a subimage from a virtual origin.

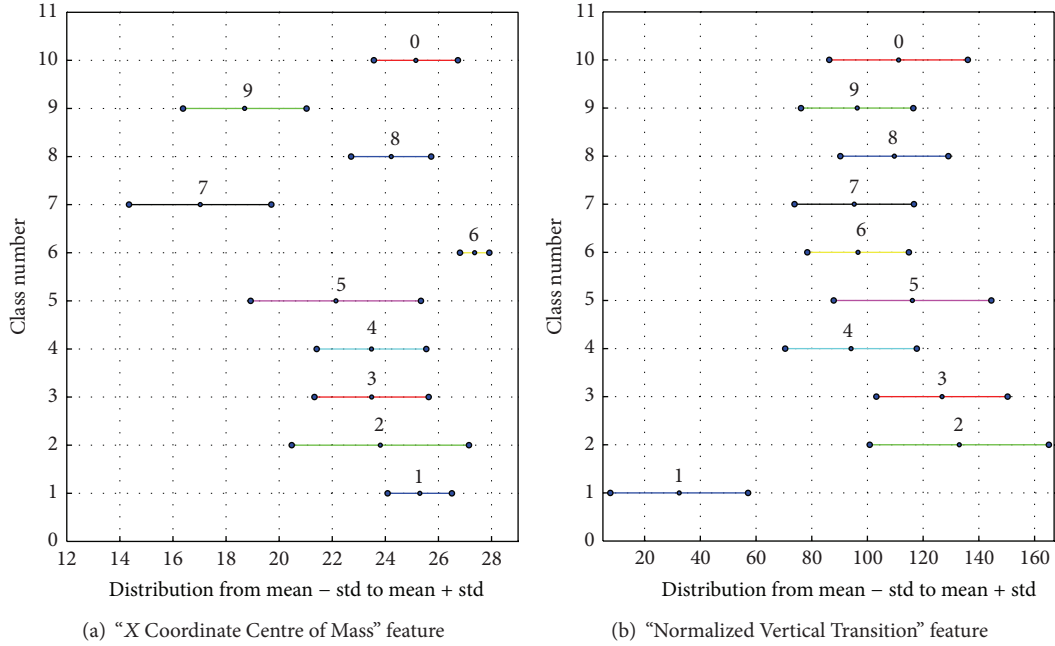


FIGURE 3: 1D_SD spectrum distribution diagrams for the English digits set.

3.3. Dimensionality Reduction (Feature Selection). In our approach, we proposed one-dimensional and two-dimensional spectrum diagrams for standard deviation and minimum to maximum distribution.

3.3.1. Stage 1: Reduction Using 1D_SD and 1D_MM. In this stage, some candidate features were selected through 1D_SD and 1D_MM spectrums. In digit recognition domain, there are 10 classes corresponding to 10 digits, 0 to 9. Hence, for each feature from the initial feature set, a 1D_SD diagram is plotted with 10 spectrum lines corresponding to digits 0 to 9. In the 1D_SD plot, a spectrum line, corresponding to a specific feature, is drawn from the mean $-SD$ to the mean $+SD$ for each class. Figures 3(a) and 3(b) show the 1D_SD distribution diagrams corresponding to the “X Coordinate Centre of Mass” and “Normalized Vertical Transition” features for English digits, respectively. In Figure 3(a), the majority of spectrum lines are in an overlapping range [20, 25], meaning that the “X Coordinate Centre of Mass” feature alone cannot discriminate existing classes from each other in the feature space. In Figure 3(b), the spectrum line corresponding to class (digit) 1 is completely separate from the other spectrum lines, indicating that the “Normalized Vertical Transition” feature can completely discriminate digit 1 (class 1) from other English digit sets (other classes). Therefore, it can be considered as a candidate feature in the final features vector.

Similar to Figure 3, Figures 4(a) and 4(b) show the 1D_SD distribution diagrams corresponding to the “Maximum Vertical Crossing Count” and “Aspect Ratio” feature for the Arabic/Farsi digits, respectively. In Figure 4(a), the majority of the spectrum lines are in an overlapping range [3.5, 7], meaning that the “Maximum Vertical Crossing Count” feature alone cannot discriminate the existing classes from each

other in the feature space. In Figure 4(b), the spectrum line corresponding to class (digit) 1 is completely separate from other spectrum lines, indicating that the “Aspect Ratio” feature can completely discriminate digit 1 (class 1) from other Arabic/Farsi digits set (other classes). Therefore, it can be considered as a candidate feature in the final features vector.

Finding a set of separated spectrum lines using only 1D_SD distribution diagrams is not enough to create an optimum feature vector, because the outlier samples in each class are not placed in the range of 1D_SD spectrum lines. Indeed they are put in One-Dimensional Minimum to Maximum (1D_MM) spectrums.

In the 1D_MM plot, a spectrum line corresponding to a specific feature is drawn from the minimum to the maximum value of that specific feature for each class. A shorter spectrum line corresponding to a specific feature indicates that the existing samples in a particular class have more similarity (less diversity) to each other in respect to that feature. Hence, a shorter spectrum line is better than a longer one. In addition, a distribution diagram with class centres (locations of the means of the classes) further apart is better than one with closer class centres. In this case, a classifier separates the existing clusters better.

Figures 5(a) and 5(b) illustrate 1D_MM spectrum lines for the same features, “Normalized Vertical Transition” (Figure 3(b)) and “Aspect Ratio” (Figure 4(b)), for the English and Arabic/Farsi digit sets, respectively. It is obvious that, in Figure 5(a), some samples of class 1 overlap with some samples in all the rest of the classes. This means that, in the recognition phase, these samples may be misclassified as belonging to other classes and vice versa, if only the “Normalized Vertical Transition” feature is employed. In addition, in Figure 5(b), some samples of class 1 overlap

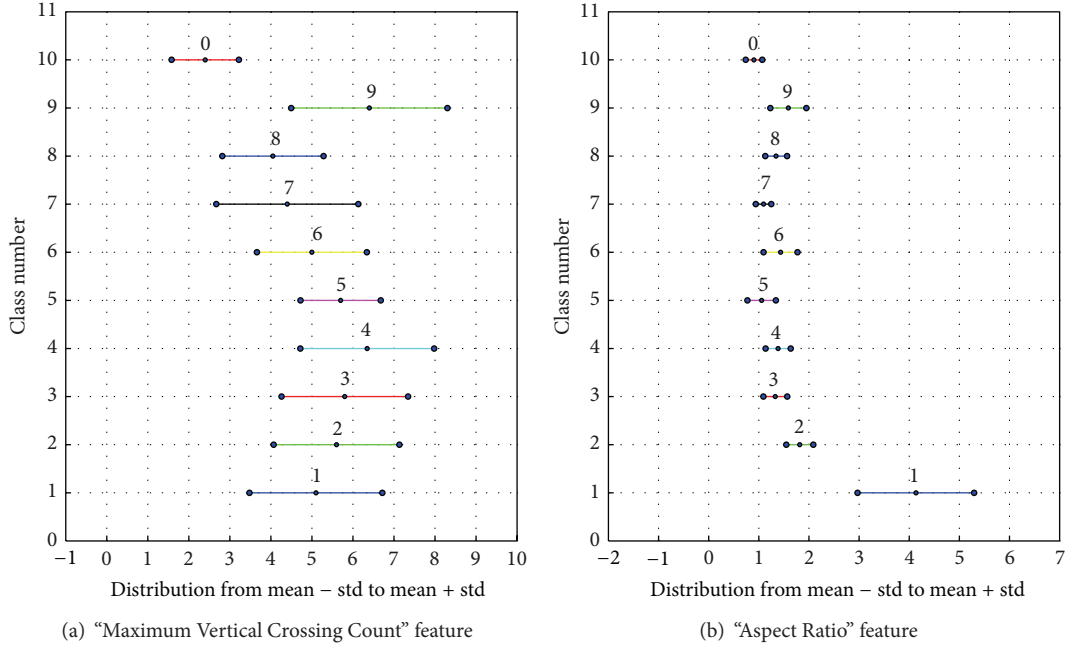


FIGURE 4: 1D_SD spectrum distribution diagrams for the Arabic/Farsi digits set.

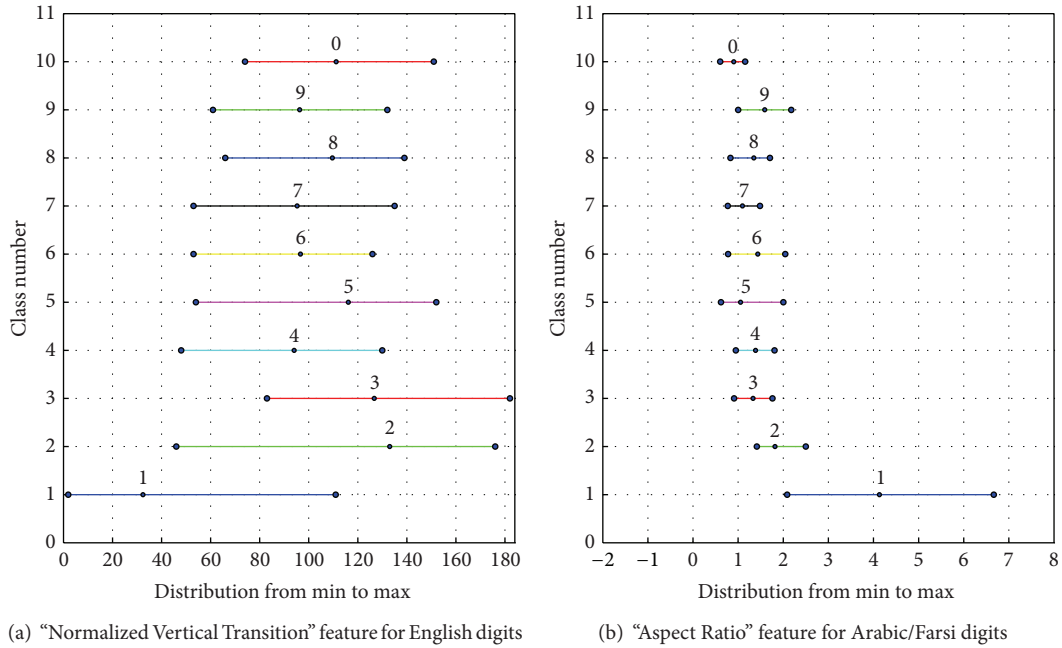


FIGURE 5: 1D_MM spectrum distribution diagrams.

with some samples in classes 2 or 9. In other words, in the recognition phase, it is possible that some samples of class 1 are misclassified into classes 2 or 9 and vice versa, if only the "Aspect Ratio" feature is utilized.

In our proposed dimensionality reduction method, 1D_MM is utilized to find the maximum allowable overlapping threshold T_1 to create the first reduced feature vector S_1 from the initial features set, $Initial_S$. By investigating the overlapping values of the spectrum lines in the 1D_MM

diagram for each feature in $Initial_S$, the value of threshold T_1 is selected. In this study, the T_1 threshold was set to 30%, experimentally.

3.3.2. Stage 2: Reduction Using 2D_SD and 2D_MM. Similar to the 1D_SD and 1D_MM distribution diagrams, the Two-Dimensional Standard Deviation (2D_SD) distribution diagram and the Two-Dimensional Minimum to Maximum

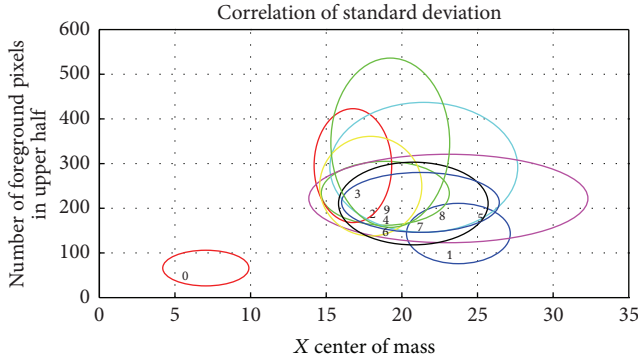


FIGURE 6: 2D_SD spectrum distribution diagrams for Arabic/Farsi digits: “X Coordinate Centre of Mass” feature versus “Number of Foreground Pixels in Upper Half of Image” feature.

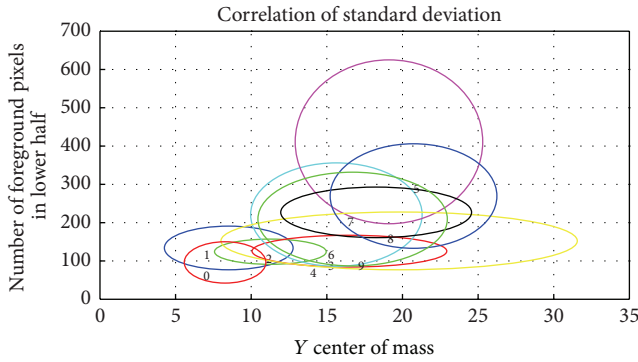


FIGURE 7: 2D_SD spectrum distribution diagrams for Arabic/Farsi digits: “Y Coordinate Centre of Mass” feature versus “Number of Foreground Pixels in Lower Half of Image” feature.

(2D_MM) spectrum for two features are made by mapping one feature on the X-axis and another feature on the Y-axis. In these cases, an ellipse (or rectangular) is plotted for each couple of features.

In 2D_SD, the main ellipse diagonals (or the length and width in the rectangular case) are plotted from the mean $-SD$ to the mean $+SD$ for two features. In 2D_MM, the main ellipse diagonals (or the length and width in the rectangular case) are plotted from the minimum value to the maximum value for these two features. As such, the $[n * (n + 1)/2]$ 2D_SD (or 2D_MM) distribution diagram can be generated for n independent features.

Figure 6 shows a 2D_SD distribution diagram for two features, namely, “X Coordinate Centre of Mass” and “Number of Foreground Pixels in Upper Half of Image,” for the Arabic/Farsi digits set. As can be seen, the ellipse for class (digit) 0 is completely distinct from the other ellipses. Hence, the feature pair (X Coordinate Centre of Mass and Number of Foreground Pixels in Upper Half of Image) is a good choice for membership in the final features vector (to distinguish class (digit) 0 from other classes (digits)).

Figure 7 shows another 2D_SD distribution for two features, “Y Coordinate Centre of Mass” and “Number of

Foreground Pixels in Upper Half of Image,” of the Arabic/Farsi digits set. It is completely clear that in this case the mentioned features are highly correlated, and, therefore, they are not a suitable feature pair for membership in the final features vector.

In our proposed dimensionality reduction method, 2D_MM is utilized to find a maximum allowable overlapping threshold T_2 to create the final reduced feature vector S_2 from the first reduced feature vector S_1 . By investigating the overlapping values of the spectrum ellipse (rectangular) in the 2D_MM diagram for the pair features in S_1 , the value of the threshold T_2 is selected. In this study, the T_2 threshold was set to 20%, experimentally.

3.3.3. Creating Final Feature Vector. For the dimensionality reduction process, we defined the value of a specific feature as $f_k(S_{i,j})$, where f_k is the value of the k th feature from the initial feature vector, $Initial_S$, and $S_{i,j}$ represents the j th sample of class i . Subsequently, using all samples in the training part of each class, the values of the minimum, maximum, mean, and standard deviation for all the features in the initial feature vector were computed.

To find the first reduced subset of feature vector, the 1D_SD distribution diagram along with the 1D_MM spectrums was generated for all 133 features in initial features set, $Initial_S$. The system selected every feature for which its 1D_SD spectrum line had a maximum of 30% overlapping (threshold T_1) with the other 1D_SD spectrum lines of the other classes. The output of this stage was the first reduced version of the feature vector S_1 , which satisfied the criteria necessary for membership in the final feature vector.

Finally, by using the 2D_SD distribution diagrams and also the 2D_MM spectrums on the first reduced version of the feature set S_1 , the final reduced versions of feature vectors S_2 were selected. In this stage, a couple of features were selected, if the 2D_SD had a maximum of 20% overlapping (threshold T_2) with the other 2D_SD distribution diagrams. In stage 2, it was possible that a feature was added to S_2 more than once. Hence, in the final step, the repetitive features in S_2 were removed to create the smallest size of S_2 .

In our proposed method, the 1D_SD and 2D_SD spectrums were utilized to decide whether or not a feature was suitable for including in the final feature vector. 1D_MM and 2D_MM were used to find the best value for thresholds T_1 and T_2 . It is completely clear that these threshold values are dependent on characteristics of training dataset samples. Algorithm 2 explains the proposed method for the dimensionality reduction operation.

The mentioned operations created feature vectors $Initial_S$, $E-S_1$, and $E-S_2$ for the English dataset MNIST and feature vectors $Initial_S$, $A/F-S_1$, and $A/F-S_2$ for the Arabic/Farsi dataset Hoda. Table 2 shows the number of features in each stage for these datasets.

The following are among the selected features in the final feature vectors $E-S_2$ and $A/F-S_2$:

- (i) X Coordinate Centre of Mass;
- (ii) Number of Foreground Pixels in Upper Half;

Extract most-used features (in literature) from input dataset, and **Create** initial features set $Initial_S$;
 $n :=$ Number of features in initial features set $Initial_S$;
first reduced version of feature vector $S_1 :=$ null;
Stage 1:
Choosing threshold T_1 , by investigating 1D-MM spectrum diagrams.
for ($k = 1 : n$)
{
 Compute the coordinate of all 1D-SD spectrum lines corresponding to feature f_k ;
 for ($c = 1 : \text{number of classes}$)
 {
 If (overlapping of spectrum line of class c with all the rest spectrum lines has the value less than threshold T_1) **then**
 {
 Insert feature f_k to S_1 ;
 goto L1;
 }
 }
 L1: **continue**;
}
 $m :=$ Number of features in first reduced version of features vector S_1 ;
final reduced versions of feature vectors $S_2 :=$ null;
Stage 2:
Choosing threshold T_2 , by investigating 2D-MM spectrum diagrams.
for ($k = 1 : m$)
 for ($h = 1 : m$)
 {
 Compute the coordinate of all 2D-SD spectrum ellipses corresponding to features pair (f_k, f_h) ;
 for ($c = 1 : \text{number of classes}$)
 {
 if (overlapping of spectrum ellipses of class c with all the rest spectrum ellipses has the value less than threshold T_2) **then**
 {
 Insert feature pair f_k and f_h to S_2 ;
 goto L2;
 }
 }
 }
 L2: **continue**;
}
delete the repetitive features from S_2 ;

ALGORITHM 2: Two-stage dimensionality reduction procedure.

TABLE 2: Number of features in initial feature vector, first reduced version of feature vector, and final reduced version of feature vector.

| | Initial feature vector | First reduced versions of feature vectors S_1 | Final reduced versions of feature vectors S_2 |
|---------------------------|------------------------|---|---|
| English dataset MNIST | 133 | $E-S_1$: 103 | $E-S_2$: 79 |
| Arabic/Farsi dataset Hoda | 133 | $A/F-S_1$: 94 | $A/F-S_2$: 58 |

- | | |
|--|--|
| (iii) Number of Foreground Pixels in Lower Half; | (x) variance for vertical histogram; |
| (iv) ratio of foreground pixels to area of bounding box; | (xi) solidity; |
| (v) ratio of number of foreground pixels upper main diagonal to number of foreground pixels under main diagonal; | (xii) perimeter; |
| (vi) Aspect Ratio; | (xiii) ratio of major to minor axes lengths; |
| (vii) normalized horizontal transition; | (xiv) convex area; |
| (viii) maximum horizontal crossing count; | (xv) number of end points; |
| (ix) Normalized Vertical Transition; | (xvi) number of end points in different zones of image bounding box; |
| | (xvii) number of branch points; |

TABLE 3: Distribution of digits in the MNIST and Hoda datasets.

| Digit | MNIST | | Hoda | |
|-------|--------|--------|--------|--------|
| | Train | Test | Train | Test |
| 0 | 5,923 | 980 | 6,000 | 2,000 |
| 1 | 6,742 | 1,135 | 6,000 | 2,000 |
| 2 | 5,958 | 1,032 | 6,000 | 2,000 |
| 3 | 6,131 | 1,010 | 6,000 | 2,000 |
| 4 | 5,842 | 982 | 6,000 | 2,000 |
| 5 | 5,421 | 892 | 6,000 | 2,000 |
| 6 | 5,918 | 958 | 6,000 | 2,000 |
| 7 | 6,265 | 1,028 | 6,000 | 2,000 |
| 8 | 5,851 | 974 | 6,000 | 2,000 |
| 9 | 5,949 | 1,009 | 6,000 | 2,000 |
| Total | 60,000 | 10,000 | 60,000 | 20,000 |

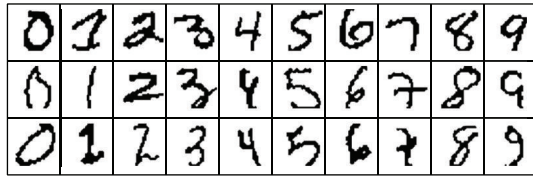


FIGURE 8: Some sample digits in the MNIST dataset.

- (xviii) some discrete cosine transform coefficients, such as (1,1), (1,2), (1,4), (1,5), and (2,1);
- (xix) some discrete cosine transform coefficients of image profile, such as (1,4), (2,3), (2,5), and (3,3);
- (xx) some discrete cosine transform coefficients of outer boundary, such as (2,1), (2,7), (3,3), and (3,4).

4. Experimental Results and Comparison

4.1. Datasets. In recent years, researchers have produced some standard benchmark datasets in order to encourage other researchers to follow their investigation in the PR field and also to compare the functionality of PR systems in the same conditions.

This research has been specifically conducted on handwritten digit OCR datasets. Some of English handwritten standard datasets, including the numeral part, are MNIST, CEDAR, CENPRMI, and IRONOFF, and some of Arabic/Farsi handwritten standard datasets, including numeral parts, are Al-Isra, ARABASE, IFHCDB, CENPARMI, Hadaf, LMCA, and Hoda.

In order to test the effectiveness of the proposed method, the digit parts of two big handwritten standard benchmark datasets were utilized, namely, MNIST, for English numerals, and Hoda, for Arabic/Farsi numerals. The following subsections demonstrate these datasets briefly.

4.1.1. MNIST Dataset. The Modified National Institute of Standards and Technology (MNIST) dataset contains 60,000 training and 10,000 test samples [41]. This dataset is an unbalanced dataset. This means that the sample frequencies

for different classes in training—and also the testing part—are not equal (Table 3). All the digits have been stored in 28×28 image pixels, with intensities from 0 to 255. Figure 8 shows some sample digits from this dataset.

4.1.2. Hoda Dataset. The Hoda dataset is a very large corpus of Arabic/Farsi handwritten alphanumeric characters [42]. It has two parts: digits and characters. The digit section of the Hoda dataset was prepared in 2007 by extracting the images of the digits from 11,942 registration forms related to university entrance forms. Those forms were scanned at 200 dpi in 24-bit colour format. The digits were extracted from the postal code, national code, record number, identity certificate number, and phone number fields of each form. The digit section of the Hoda dataset has 80,000 samples and has been divided into two parts, namely, training (60,000 samples) and testing (20,000 samples). This dataset is a balanced dataset. It includes 6,000 and 2,000 samples for each digit in the training and testing parts, respectively. Figure 9 shows some sample digits from this dataset.

Table 3 includes the distribution of digits in training and testing parts of MNIST and Hoda datasets.

4.2. Proposed Method. In this research, the same operations were carried out in the preprocessing step on the training and testing samples. The outputs were noise filtered, reslanted, relocated, and dimension normalized.

Several experiments were carried out to test the effectiveness of the proposed method for dimensionality reduction.

In the first part, we applied the proposed approach for the recognition of handwritten English digits. In all experiments, a multilayer perceptron neural network with backpropagation was trained with 103 (or 79) neurons in the input layer (corresponding to the number of features in sets $E-S_1$ and $E-S_2$), 30 neurons (found experimentally) in the hidden layer, and 10 neurons (corresponding to 10 different classes of digits 0 to 9) in the output layer, respectively.

In the first experiment, a neural network was employed with 103 (number of features in set $E-S_1$) neurons in the input layer. The network was trained with all 60,000 samples from the training part of the MNIST dataset and was then tested with all 10,000 samples from the testing part of the MNIST dataset. This operation was repeated 10 times, and, finally, 93.17% accuracy was achieved on average in this stage.

To compare the performance of the proposed method against other well-known feature selection techniques, a general Principal Component Analysis (PCA) technique and a Random Projection (RP) technique [36] were applied on the initial reduced feature set $E-S_1$ with 103 features. PCA changed the order of the features in the new orthogonal feature space—based on the derived eigenvectors—and generated a new reordered feature set $E-S_3$ with the same 103 features. The complete reordered feature set $E-S_3$ was fed into the same MLP-NN and a final accuracy of 93.82% was achieved. This result was 0.65% higher than the 93.17% result, portraying the superiority of the PCA technique for feature selection. Similarly, we employed RP dimensionality reduction technique to create a new smaller feature space

TABLE 4: Recognition rate corresponding to different feature vectors for the English dataset MNIST and Arabic/Farsi dataset Hoda.

| | Feature set | Number of features in feature vector | Feature selection method | | | Accuracy % |
|---------------------------|---------------------------------|--------------------------------------|--------------------------|------------|-----------|--------------|
| | | | Proposed method | PCA method | RP method | |
| English dataset MNIST | <i>Initial_S</i> | 133 | — | — | — | 91.93 |
| | <i>E-S₁</i> | 103 | * | | | 93.17 |
| | <i>E-S₂</i> | 79 | * | | | 94.88 |
| | <i>E-S₃</i> | 103 | | * | | 93.82 |
| | <i>E-S₄</i> | 79 | | * | | 90.71 |
| | <i>E-S₅</i> | 103 | | | * | 93.51 |
| | <i>E-S₆</i> | 79 | | | * | 88.39 |
| Arabic/Farsi dataset Hoda | <i>Initial_S</i> | 133 | — | — | — | 90.41 |
| | <i>A/F-S₁</i> | 94 | * | | | 92.60 |
| | <i>A/F-S₂</i> | 58 | * | | | 95.12 |
| | <i>A/F-S₃</i> | 94 | | * | | 94.04 |
| | <i>A/F-S₄</i> | 58 | | * | | 89.00 |
| | <i>A/F-S₅</i> | 94 | | | * | 91.07 |
| | <i>A/F-S₆</i> | 58 | | | * | 83.66 |

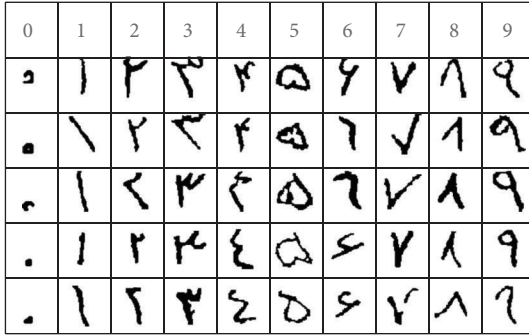


FIGURE 9: Some sample digits in the Hoda dataset.

(with 103 features) from the initial feature set, *Initial_S*. The output of this stage was the new feature set *E-S₅*. The feature set *E-S₅* was fed into the same MLP-NN, too. In this experiment, a final accuracy of 93.51% was achieved that it is 0.31% lower than the achieved result by using PCA.

In the second experiment, the system was trained with the proposed final version of feature set *E-S₂* with only 79 features. On average, the correct recognition rate increases from 93.17% to 94.88%, clearly indicating the superiority of the reduced feature set *E-S₂* of the proposed method against the initial reduced feature set *E-S₁* with 103 features. To find the superiority of the our proposed feature set, *E-S₂*, compared to the other subsets of *Initial_S*, which have 79 members, we made set *E-S₄* with the first 79 members of set *E-S₃* (generated by PCA) and set *E-S₆* with 79 members using RP technique. The recognition rate declined dramatically from 93.82% to 90.71% by using set *E-S₄* and from 93.51% to 88.39% by using set *E-S₆*. These results obviously show the effectiveness and superiority of our proposed technique in comparison with PCA and RP as two of the popular techniques for feature selection operation. The obtained results also show the superiority of PCA technique compared to RP

technique for dimensionality reduction purpose. However, it is worth mentioning that some researchers have shown the superiority of RP against PCA—for dimension reduction purpose—in high-dimensional feature space condition [43, 44]. The outcome results are reported in rows 2 to 8 of Table 4.

In the second part, we repeated all the experiments in the first part, for the Arabic/Farsi digits, using the Hoda dataset. *A/F-S₁*, *A/F-S₂*, *A/F-S₃*, *A/F-S₄*, *A/F-S₅*, and *A/F-S₆* were the first reduced feature vectors with 94 members, the final reduced feature vector created by our proposed method including 58 features, the reordered initial reduced feature vector by PCA including 94 features, the first 58 features from the *A/F-S₃*, the smaller version of feature space with 94 features created by using RP technique, and also another smaller version of feature space with 58 features created by using RP technique, in order. It is worth mentioning that not only the trend was completely similar to previous experiments for the English digits dataset, but also the achieved results for the Arabic/Farsi dataset Hoda were better than the achieved results for the English dataset MNIST. In this case, using the proposed feature selection method, the number of features was reduced from 113 to 58; meanwhile, the final accuracy was increased from 90.41% to 95.12%. In this experiment, when we used the PCA technique for feature selection, the accuracy was decreased significantly from 94.04% to 89.00%, and when we employed the RP technique for feature selection, the accuracy was decreased from 91.07% to 83.66%. Here, we only employed 1/2 of training and testing samples from Hoda dataset. The corresponding results are shown in the last seven rows of Table 4.

5. Conclusion

In this paper, a new method for dimensionality reduction (feature selection) in pattern recognition systems was introduced. To begin with, an initial set of the most used features was extracted from the training patterns of two handwritten

digit standard datasets: MNIST for the English digits and Hoda for the Arabic/Farsi digits. Then, by using the proposed 1D_SD and 1D_MM distribution diagrams methods, the initial feature vector was reduced to a smaller version based on the maximum allowable overlap between the spectrum lines using the threshold T_1 . Thereafter, by using the proposed 2D_SD and 2D_MM spectrums, a final reduced feature vector was selected. In this stage, another threshold, T_2 , was used to guarantee that the overlaps between the spectrum diagrams were not more than the maximum threshold.

The mentioned algorithm was implemented in an OCR application system to reduce the dimension of the initial feature vector. For English MNIST dataset, the feature vector was decreased from 100% (133 elements) to 59.40% (79 elements); meanwhile, the accuracy was increased 2.95% (from 91.93% to 94.88%). The accuracy was 4.17% more than the outcome results in a similar experiment that used PCA (90.71%) and 6.49% more than outcome results in a similar experiment that used RP (88.39%) as two of the common techniques for the feature selection operation. All of the 60,000 training samples and 10,000 testing samples were used in the operation.

Utilizing Arabic/Farsi dataset Hoda, the feature vector was decreased from 100% (133 elements) to 43.61% (58 elements); meanwhile, the accuracy was increased 4.71% (from 90.41% to 95.12%). The accuracy was 6.12% more than the outcome results in a similar experiment that used PCA (89.00%) and 11.46% more than outcome results in a similar experiment that used RP (83.66%) as two of the common techniques for the feature selection operation. For this experiment, we only employed 1/2 of the training samples (30,000 samples) and 1/2 of the testing samples (10,000 samples).

The results clearly indicate the superiority of the proposed method for dimensionality reduction (feature selection). According to the results, the proposed technique is completely effective for OCR application as a subcategory of PR systems. Nevertheless, the proposed new method can be used for other PR systems with different database types.

Along with proposing a new method for dimensionality reduction, this paper introduced a new method to connect the broken parts of an image in the preprocessing stage of an OCR system. This technique estimates the pen width in three different ways. By utilizing the connected component analysis, it traverses the outer contour of the separated blocks in an image and connects them together.

The proposed feature selection method can be considered as an approach to infer the appropriate rules for creating a decision tree classifier in a PR system. In other words, using the 1D_MM and 2D_MM spectrum diagrams, the necessary rules in a decision tree classifier are found more accurate and faster. This is a salient feature of our proposed approach.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This paper is supported by University of Malaya Research Grant BK026-2013.

References

- [1] Z. Sanaei, S. Abolfazli, A. Gani, and R. Buyya, "Heterogeneity in mobile cloud computing: taxonomy and open challenges," *IEEE Communications Surveys & Tutorials*, vol. 99, pp. 1–24, 2013.
- [2] A. M. Urmanov, A. A. Bougaev, and K. C. Gross, "Reducing the size of a training set for classification," Tech. Rep., US Patent Application Publication, 2007.
- [3] S. V. N. Vishwanathan and M. N. Murty, "Use of multi category proximal SVM for data set reduction," *International Journal of Studies in Fuzziness and Soft Computing*, vol. 140, pp. 3–20, 2004.
- [4] Z. Wu, W. Xie, J. Yu, and X. Gao, "Reduction of training datasets via fuzzy entropy for support vector machines," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '04)*, pp. 2381–2385, October 2004.
- [5] J. Cervantes, X. Li, and W. Yu, "Support vector classification for large data sets by reducing training data with change of classes," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '08)*, pp. 2609–2614, October 2008.
- [6] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proceedings 21th International Conference on Machine Learning (ICML '04)*, pp. 225–232, July 2004.
- [7] D. H. Milone, G. Stegmayer, L. Kamenetzky, M. López, and F. Carrari, "Clustering biological data with SOMs: on topology preservation in non-linear dimensional reduction," *Expert System With Applications*, vol. 40, no. 9, pp. 3841–3845, 2013.
- [8] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, 2012.
- [9] A. Bansal, K. Mehta, and S. Arora, "Face recognition using PCA and LDA algorithm," in *Proceedings of the 2nd International Conference on Advanced Computing and Communication Technologies (ACCT '12)*, pp. 251–254, January 2012.
- [10] F. S. Tsai, "Dimensionality reduction techniques for blog visualization," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2766–2773, 2011.
- [11] H. B. Borges and J. C. Nievola, "Comparing the dimensionality reduction methods in gene expression databases," *Expert Systems with Applications*, vol. 39, pp. 10780–10795, 2012.
- [12] M. Song, H. Yang, S. H. Siadat, and M. Pechenizkiy, "A comparative study of dimensionality reduction techniques to enhance trace clustering performances," *Expert Systems With Applications*, vol. 40, pp. 3722–3737, 2013.
- [13] H.-C. Kim, D. Kim, and S. Yang Bang, "A numeral character recognition using the PCA mixture model," *Pattern Recognition Letters*, vol. 23, no. 1–3, pp. 103–111, 2002.
- [14] P. Zhang, C. Y. Suen, and T. D. Bui, "Multi-modal nonlinear feature reduction for the recognition of handwritten numerals," in *Proceedings of the 1st Canadian Conference on Computer and Robot Vision*, pp. 393–400, May 2004.
- [15] V. Deepu, S. Madhvanath, and A. G. Ramakrishnan, "Principal component analysis for online handwritten character recognition," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, pp. 327–330, August 2004.
- [16] G. A. Abandah, K. S. Younis, and M. Z. Khedher, "Handwritten Arabic character recognition using multiple classifiers based

- on letter form,” in *Proceedings of the 5th IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA '08)*, pp. 128–133, February 2008.
- [17] M. Karic and G. Martinovic, “Improving offline handwritten digit recognition using concavity-based features,” *International Journal of Computers, Communications and Controls*, vol. 8, no. 2, pp. 220–234, 2013.
- [18] M. T. Parvez and S. A. Mahmoudi, “Offline arabic handwritten text recognition: a survey,” *ACM Computing Survey*, vol. 45, no. 2, p. 23, 2013.
- [19] J. H. Al-Khateeb, J. Jiang, J. Ren, F. Khelifi, and S. S. Ipson, “Multiclass classification of unconstrained handwritten arabic words using machine learning approaches,” *The Open Signal Processing Journal*, vol. 2, pp. 21–28, 2009.
- [20] X. Peng, H. Cao, S. Setlur, V. Govindaraju, and P. Natarajan, “Multilingual OCR research and applications: an overview,” in *Proceedings of the 4th International Workshop on Multilingual OCR*, 2013.
- [21] M. Shah and G. B. Jethava, “A literature review on hand written character recognition,” *Indian Streams Research Journal*, vol. 3, no. 2, pp. 1–19, 2013.
- [22] D. R. Ahmed and M. J. Nordin, “Offline OCR system for machine-printed Turkish using template matching,” *Advanced Materials Research*, vol. 341–342, pp. 565–569, 2012.
- [23] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLABed*, Gatesmark Publishing, 2nd edition, 2009.
- [24] M. S. Khorsheed, “Off-line Arabic character recognition—a review,” *Pattern Analysis and Applications*, vol. 5, no. 1, pp. 31–45, 2002.
- [25] S. Impedovo, *More Than Twenty Years of Advancements on Frontiers in Handwriting Recognition*, Pattern Recognition, 2013.
- [26] O. P. Sharma, M. K. Ghose, K. B. Shah, and B. K. Thakur, “Recent trends and tools for feature extraction in OCR technology,” *International Journal of Soft Computing and Engineering*, vol. 2, no. 6, pp. 220–223, 2013.
- [27] R. Chen, C. Beijun, M. R. Asharif, and K. Yamashita, “Character recognition algorithm based on the central location and polar projection,” in *Proceedings of the 2nd International Congress on Image and Signal Processing (CISP '09)*, October 2009.
- [28] G. Abandah and N. Anssari, “Novel moment features extraction for recognizing handwritten Arabic letters,” *Journal of Computer Science*, vol. 5, no. 3, pp. 226–232, 2009.
- [29] M. A. Shayegan and S. Aghabozorgi, “A new dataset size reduction approach for PCA-based classification in OCR application,” *Mathematical Problems in Engineering*, vol. 2014, Article ID 537428, 14 pages, 2014.
- [30] I. Guyon and A. Elisseeff, “An Introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, no. 1, pp. 1157–1182, 2003.
- [31] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [32] M. Dash and H. Liu, “Feature selection for classification,” *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.
- [33] Y. El-Glaly and F. Quek, *Isolated Handwritten Arabic Character Recognition Using Multilayer Perceptron and K Nearest Neighbor Classifiers*, 2011.
- [34] Z. Bahmani, F. Alamdar, R. Azmi, and S. Haratizadeh, “Off-line Arabic/Farsi handwritten word recognition using RBF neural network and genetic algorithm,” in *Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS '10)*, pp. 352–357, October 2010.
- [35] M. Soryani and N. Rafat, “Application of genetic algorithms to feature subset selection in a Farsi OCR,” *World Academy of Science, Engineering and Technology*, vol. 18, pp. 113–116, 2006.
- [36] I. K. Fodor, *A Survey of Dimension Reduction Techniques*, 2002.
- [37] H. Wang, “Structural two-dimensional principal component analysis for image recognition,” *Machine Vision and Applications*, vol. 22, no. 2, pp. 433–438, 2011.
- [38] D. R. Gesualdi and J. M. Seixas, *Character Recognition in Car License Plates Based on Principal Components and Neural Processing*, The VII Brazilian Symposium on Neural Networks, 2002.
- [39] M. Ziaratban, K. Faez, and M. Ezoji, “Use of legal amount to confirm or correct the courtesy amount on Farsi bank checks,” in *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR '07)*, pp. 1123–1127, September 2007.
- [40] M. Hanmandlu, K. R. Murali Mohan, S. Chakraborty, S. Goyal, and D. R. Choudhury, “Unconstrained handwritten character recognition based on fuzzy logic,” *Pattern Recognition*, vol. 36, no. 3, pp. 603–623, 2003.
- [41] The MNIST dataset, <http://yann.lecun.com/exdb/mnist/index.html>.
- [42] H. Khosravi and E. Kabir, “Introducing a very large dataset of handwritten Farsi digits and a study on their varieties,” *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1133–1141, 2007.
- [43] S. Deegalla and H. Boström, “Reducing high-dimensional data by principal component analysis vs. random projection for nearest neighbor classification,” in *Proceedings of the 5th International Conference on Machine Learning and Applications (ICMLA '06)*, pp. 245–250, December 2006.
- [44] D. Fradkin and D. Madigan, “Experiments with random projections for machine learning,” in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '03)*, pp. 517–522, August 2003.

Research Article

Performance Comparison of Wavelet and Multiwavelet Denoising Methods for an Electrocardiogram Signal

Balambigai Subramanian,¹ Asokan Ramasamy,² and Kamalakannan Rangasamy¹

¹ Department of Electronics and Communication Engineering, Kongu Engineering College, Perundurai, Erode District, Tamil Nadu 638052, India

² Kongunadu College of Engineering and Technology, Thottiyam, Trichy District, Tamil Nadu 621215, India

Correspondence should be addressed to Balambigai Subramanian; bharathian007@yahoo.com

Received 22 January 2014; Revised 14 April 2014; Accepted 17 April 2014; Published 11 May 2014

Academic Editor: Feng Gao

Copyright © 2014 Balambigai Subramanian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The increase in the occurrence of cardiovascular diseases in the world has made electrocardiogram an important tool to diagnose the various arrhythmias of the heart. But the recorded electrocardiogram often contains artefacts like power line noise, baseline noise, and muscle artefacts. Hence denoising of electrocardiogram signals is very important for accurate diagnosis of heart diseases. The properties of wavelets and multiwavelets have better denoising capability compared to conventional filtering techniques. The electrocardiogram signals have been taken from the MIT-BIH arrhythmia database. The simulation results prove that there is a 29.7% increase in the performance of multiwavelets over the performance of wavelets in terms of signal to noise ratio (SNR).

1. Introduction

In modern medicine, there are many methods to diagnose heart disease such as electrocardiogram (ECG), ultrasound, magnetic resonance imaging (MRI), and computer tomography (CT). Among these methods, diagnosis using electrocardiogram has the advantages of convenience and low cost so that it can be used in a wide area. However, certain arrhythmia (a fast, slow, or irregular heartbeat) which can cause abnormal symptoms may occur only sporadically or may occur only under certain conditions such as stress. Arrhythmia of this type is difficult to obtain on an electrocardiogram tracing that runs only for a few minutes. The electrocardiogram is the record of variation of bioelectric potential with respect to time as the human heart beats. Due to its ease of use and noninvasiveness, electrocardiogram plays an important role in patient monitoring and diagnosis.

The change in solar activity including electrocardiographic data with variations in galactic cosmic rays, geomagnetic activity, and atmospheric pressure suggests the possibility of links among these physical environmental variations and health risks, such as myocardial infarctions and ischemic strokes. An increase in the incidence of myocardial infarction

in association with magnetic storms has been reported by Cornélissen et al. [1].

Magnetic storms are found to decrease heart rate variability (HRV) indicating a possible mechanism since a reduced HRV is an important factor for coronary artery disease and myocardial infarction. An increase of 5% in mortality during years of maximal solar activity is found when compared with years of minimal solar activity. These chronodiagnostics are particularly important for those venturing into regions away from hospitals.

Goudarzi et al. [2] made an effort to find the optimum multiwavelet for compression of ECG signals to be used along with SPIHT codec. This work examined different multiwavelets on 24 sets of ECG data with entirely different characteristics selected from MIT-BIH database and assessed the functionality of the different multiwavelets in compressing electrocardiogram signals and their simulation results showed the cardinal balanced multiwavelet (cardbal2) by the means of identity (Id) prefiltering method to be the best effective transformation and the most efficient multiwavelet was applied for SPIHT coding algorithm on the transformed signal by this multiwavelet.

Kania et al. [3] studied the application of wavelet denoising in noise reduction of multichannel high resolution ECG signals. The influence of the selection of wavelet function and the choice of decomposition level on efficiency of denoising process was considered and whole procedures of noise reduction were implemented in MATLAB environment using the fast wavelet transform. The denoising method was found to be advantageous since noise level was decreased in ECG signals, in which noise reduction by averaging had limited application, that is, in case of arrhythmia.

Helenprabha and Natarajan [4] proposed a technique used for measuring electrical signals generated by foetal heart as measured from multichannel potential recordings on the mother's body surface. They proposed a new class of adaptive filter that combines the attractive properties of finite impulse response (FIR) filter with infinite impulse response (IIR) filter. The maternal ECG and foetal signals were simulated using MATLAB. The gamma filter design was implemented in FPGA Spartan 2E which was programmed using VHDL. Their results have solved the complex situations more reliably than normal adaptive methods used earlier for recovering foetal signals.

Chang et al. [5] proposed measures to make the optimal filter design under different constraints possible for ECG signal processing. Experiments have been conducted by them with artificially and practically corrupted ECG signals for PLI adaptive filtering technique. The assessments included the convergence time, the frequency tracking efficiency, the execution time, and the relative statistics in time and frequency domain. The results demonstrated that there is no universal optimum approach for this application thus far.

Alfaouri and Daqrouq [6] performed wavelet transform thresholding technique for ECG signal denoising. They decomposed the signal into five levels of wavelet transform using the Daubechies wavelet (db4) and determined a threshold through a loop to find the value where minimum error was achieved between the detailed coefficients of threshold noisy signal and the original signal. The threshold value was accomplished experimentally after using a loop of calculating a minimum error between the denoised wavelet subsignals and the original free of noise subsignals. The experimental application of the threshold result was better than Donoho's threshold particularly in ECG signal denoising.

Zhidong and Chan [7] proposed a novel method for the removal of power line frequency from ECG signals based on empirical mode decomposition (EMD) and adaptive filter. A data-driven adaptive technique called EMD was used to decompose ECG signal into a series of intrinsic mode functions (IMFs). The adaptive power line cancellation filter was designed to remove the power line interference, the reference signal of which was produced by selective reconstruction of IMFs. Clinic ECG signals were used to evaluate the performance of the filter. Results indicated that the power line interference of ECG was removed effectively by the new method.

Kaur and Singh [8] proposed a combination method for power line interference reduction in ECG. The methods were moving average method and using the IIR notch characteristics. Their results showed reduction in the power

line noise in the ECG signal using the proposed filter that has fewer coefficients and hence lesser computation time for real time processing.

Haque et al. [9] used wavelet method to detect the small variations of ECG features. They simulated standard ECG signals as well as the simulated noise corrupted signal using FFT and wavelet for proper feature extraction. They found wavelet to be superior to the conventional FFT method in finding the small abnormalities in electrocardiogram signals.

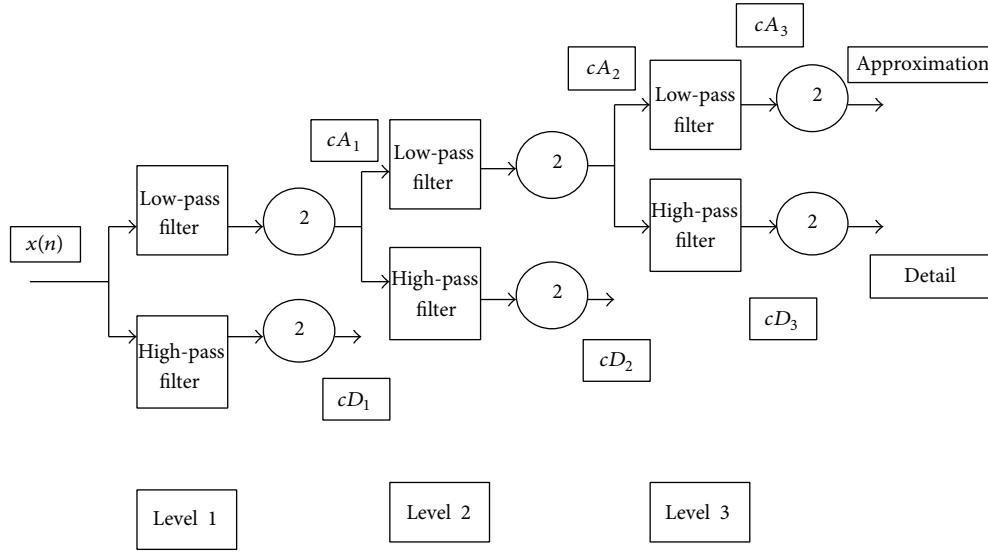
Tan and Lei [10] used wavelet transform to filter out noise interferences of electrocardiogram signals for the filtering of the myoelectric interference, the power frequency interference, and the baseline drift. Firstly Coif4 wavelet was adapted to decompose electrocardiogram signals containing noises. Secondly, the soft and hard threshold value quantified high-frequency coefficients of every scale and finally the electrocardiogram were reconstructed using high-frequency coefficients of every scale which were quantified by the threshold value. Experiments showed that wavelet transform had good real time filtering effect and it had more advantages than traditional methods.

2. Materials and Methods

2.1. Wavelet Method. A wavelet is simply a small wave which has energy concentrated in time. It is compactly supported and has finite energy function. It can satisfy admissibility condition and could be amendable for multiresolution analysis. The wavelet transform is a mathematical tool for decomposing a signal into a set of orthogonal waveforms localized both in time and frequency domains. The wavelet transform is a suitable tool to analyse the electrocardiogram signal, which is characterized by a cyclic occurrence of patterns with different frequency content (*P* wave, QRS complex, and *T* wave). It is a decomposition of the signal as a combination of a set of basic functions, obtained by means of dilation (*a*) and translation (*b*) of a single prototype wavelet; there are several wavelet functions (mother wavelet with different properties) like the Morlet or Mexican Hat wavelets or complex frequency *B* spline wavelets that are used in study.

Wavelet analysis is done by the breaking up of a signal into a shifted and scaled version of the original wavelet. A continuous wavelet transform can be defined as the sum of overall time of the signal multiplied by a scaled and shifted version of the wavelet function. The greater the scale factor "*a*" is, the wider the basis function is and, consequently, the corresponding coefficient gives information about lower frequency components of the signal and vice versa.

The wavelet transform is designed to address the problem of nonstationary signals such as electrocardiogram signals. It involves representing a time function in terms of simple, fixed building blocks, and termed wavelets. The next step is the selection of number of decomposition levels of signal $x_i(t)$; see Figure 1. First decomposition level is obtained by using two complementary high- and low-pass filters and then half of the samples are eliminated. The filters cut frequency is equal to half of the bandwidth of analysed signal.

FIGURE 1: Diagram of multiresolution analysis of signal $x_i(t)$.

Such algorithm, which is amplification of discrete wavelet transform, is known as fast wavelet transform.

For analysis the following mother wavelet was used:

$$\Psi_{m,n}(t) = 2^{-m/2} \Psi(2^{-m}t - n), \quad (1)$$

where n is coefficient of time translation and m is coefficient of scale (compression).

In the first step threshold values for detail coefficients at every level of decomposition are determined according to the following relationship:

$$\text{THR}_j = \sqrt{2 \log \|cD_j\|}. \quad (2)$$

The next step is the modification of values of the j th level detail coefficients basis of appointed threshold. This method is called soft thresholding procedure as follows:

$$cD_j(t) = \begin{cases} \text{sgn}(cD_j(t))(|x| - \text{THR}_j); & cD_j(t) > \text{THR}_j, \\ 0; & cD_j(t) \leq \text{THR}_j. \end{cases} \quad (3)$$

The final step of the analysis is reconstruction of signal $x_i(t)$ based of approximation coefficients chosen i th level of decomposition (cA_i) and modified detail coefficients from i th (cD_i) as well as higher levels of decomposition

$$x_i(t) = \sum_n cA_{m,n} \Phi_k(2^{-m}t - n) + \sum_{m=m_0}^{m_k} \sum_n cD_{m,n} \Psi_{m,n}(2^{-m}t - n), \quad (4)$$

where $\Phi_k(t)$ is scaling function from k th level of decomposition and $\Psi_{m,n}(t)$ are wavelet functions for $m = m_0, \dots, m_k$ levels of decomposition.

The advantages of wavelet methods are possibility of receiving good quality signal for beat to beat analysis and

possibility of having high quality signal while averaging technique is impossible, as causing morphology distortion of electrocardiogram signals, it provides a way for analysing waveforms bounded in both frequency and duration, it allows signals to be stored more efficiently than by the Fourier transform, it can lead to better approximate real-world signals and it is well-suited for approximating data with sharp discontinuities. The disadvantage of wavelet method is that the wavelet transforms ignore polynomial components of the signal up to the approximation order of the basis.

2.2. Equation for Continuous Wavelet Transform. The wavelet transform equation is given by

$$\text{CWT}_x^\Psi(\tau, s) = \Psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) \Psi^*\left(\frac{t-\tau}{s}\right) dt, \quad (5)$$

where $x(t)$ = given signal, τ = translation parameter, s = scaling parameter = $1/f$, and $\Psi(t)$ = mother wavelet.

2.3. Multiwavelet Method. Multiwavelets constitute a new chapter which has been added to wavelet theory in recent years. Recently, much interest has been generated in the study of the multiwavelets where more than one scaling functions and mother wavelet are used to represent a given signal. The first construction for polynomial multiwavelets was given by Albert, who used them as a basis for the representation of certain operators. Later, Geronimo, Hardin, and Massopust constructed a multiscaling function with 2 components using fractal interpolation.

In spite of many theoretical results on multiwavelet, their successful applications to various problems in signal processing are still limited. Unlike scalar wavelets in which Mallet's pyramid algorithm have provided a solution for good signal decomposition and reconstruction, a good framework for the application of the multiwavelet is still not available. Nevertheless, several researchers have proposed method of

how to apply a given multiwavelet filter to signal and image decomposition.

2.4. Multiscaling Functions and Multiwavelets. The concept of multiresolution analysis can be extended from the scalar case to general dimension r . A vector valued function $\varphi = [\varphi_1, \varphi_2, \dots, \varphi_r]^T$ belonging to $L^2(R)^r$ and r is called a multiscaling function if the sequence of closed spaces

$$V_j = \text{span} \{2^{j/2} \varphi_i (2^j - k) : 1 \leq i \leq r, k \in \mathbb{Z}\}. \quad (6)$$

$j \in \mathbb{Z}$ constitute a multiresolution analysis (MRA) of multiplicity r for $L^2(R)$. The multiscaling function must satisfy the two-scale dilation equation

$$\varphi(t) = \sqrt{2} \sum_k G_k \varphi(2t - k). \quad (7)$$

Now let W_j denote a complementary space of V_j in V_{j+1} . The vector valued function $\Psi = [\Psi_1, \Psi_2, \dots, \Psi_r]^T$ such that

$$W_j = \text{span} \{2^{j/2} \varphi_i (2^j - k) : 1 \leq i \leq r, k \in \mathbb{Z}\}. \quad (8)$$

$j \in \mathbb{Z}$ is called a multiwavelet. The multiscaling function must satisfy the two-scale equation

$$\Psi(t) = \sqrt{2} \sum_k H_k \varphi(2t - k). \quad (9)$$

$H_k \in L^2(\mathbb{Z})^{r \times r}$ is an $r \times r$ matrix of coefficients. The two-scale equations (2) and (4) can be realized as a multifilter bank operating on r input data streams and filtering them in two $2r$ output data streams, each of which is downsampled by a factor of two. If $x(t)$ is the given signal and it is assumed that $x(t) \in V_0$, then

$$x(t) = \sqrt{2} \sum_k V_{0,k}^T \varphi(t - k). \quad (10)$$

And the scaling coefficient $V_{1,k}^T$ of the first level can be considered as a result of low-pass multifiltering and downsampling as follows:

$$V_{1,k} = \sum_m G_{m-2k} V_{0,m}. \quad (11)$$

Analogously, the first level multiwavelet coefficients $W_{1,k}$ are obtained using high-pass multifiltering and downsampling as follows:

$$W_{1,k} = \sum_m H_{m-2k} V_{0,m}. \quad (12)$$

Full multiwavelet decomposition of the signal $x(t)$ can be found by iterative filtering of the scaling coefficient as follows:

$$\begin{aligned} V_{j,k} &= \sum_m G_{m-2k} V_{j-1,m}, \\ W_{j,k} &= \sum_m H_{m-2k} V_{j-1,m}. \end{aligned} \quad (13)$$

Note that $V_{j,k}$ and $W_{j,k}$ are $r \times 1$ column vectors.

2.5. Advantages of Wavelets and Multiwavelets Compared to Conventional Filtering Techniques

- (i) The Fourier transform fails to analyze the nonstationary signal, whereas wavelet transform allows the components of a nonstationary signal to be analyzed.
- (ii) Wavelet transform holds the property of multiresolution to give both time and frequency domain information in a simultaneous manner.
- (iii) A set of wavelets which are complementary can decompose the given data without gaps or overlap so that the decomposition process becomes mathematically reversible.

2.6. Comparison between Wavelet and Multiwavelet. Multiwavelets contain multiple scaling functions, whereas scalar wavelets contain one scaling function and one wavelet. This leads to more degrees of freedom in constructing wavelets. Therefore, opposed to scalar wavelets, properties such as compact support, orthogonality, symmetry, vanishing moments, and short support can be gathered simultaneously in multiwavelets which are fundamental in signal processing.

The increase in degree of freedom in multiwavelets is obtained at the expense of replacing scalars with matrices, scalar functions with vector functions, and single matrices with block of matrices. However, prefiltering is an essential task which should be performed for any use of multiwavelet in signal processing.

3. Results and Discussions

3.1. Data Collection

3.1.1. MIT-BIH Arrhythmia Database. MIT-BIH arrhythmia database consists of 48-half-hour electrocardiogram recordings. The recordings were digitized at 360 Hz (samples per second per channel) with 11-bit resolution over 10 mV. The simulations were carried out in MATLAB environment R2010b. Various benchmark records from the MIT-BIH database were considered for this study.

(A) Performance Analysis of Wavelet Based Denoising Method for Electrocardiogram

Wavelet Denoising Using Biorthogonal 1D Wavelet. Figure 2 shows the wavelet denoising for the record 103 m in which level 1 approximation coefficient d_1 for Biorthogonal wavelet shows that there is maximum noise in it. Hence reconstruction of the original signal to obtain the denoised electrocardiogram from d_1 coefficients will also contain the maximum noise. The level 4 decomposition contains the least noise and hence reconstruction is done using d_4 .

Original and Denoised Signals Using Biorthogonal Wavelet. Figure 3 shows the comparison of the original signal and the denoised signal using Biorthogonal wavelet at level 4 decomposition for the record 103 m. In this the signal to noise ratio obtained is 32.2094 dB and the power line noise

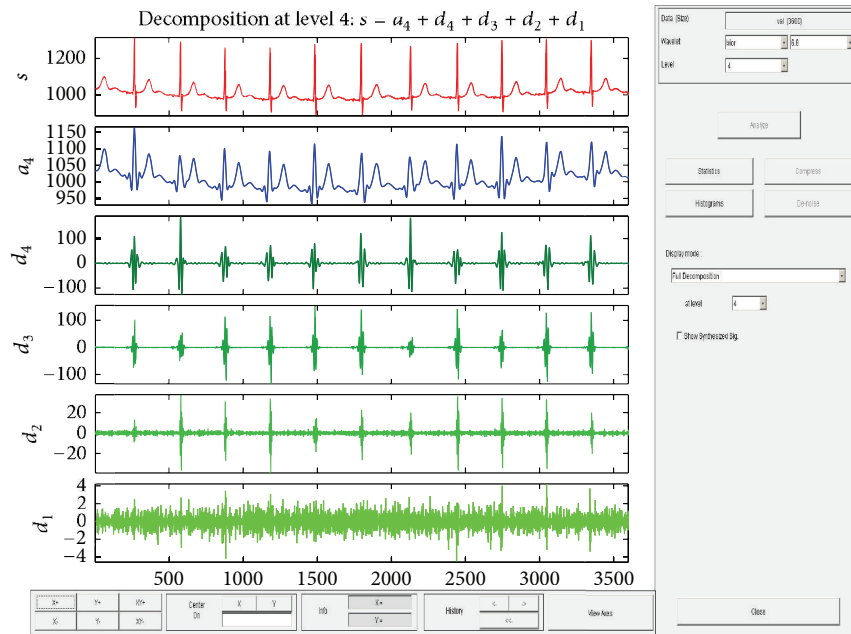


FIGURE 2: Wavelet denoising for ECG record no 103 m using Biorthogonal 1D wavelet.

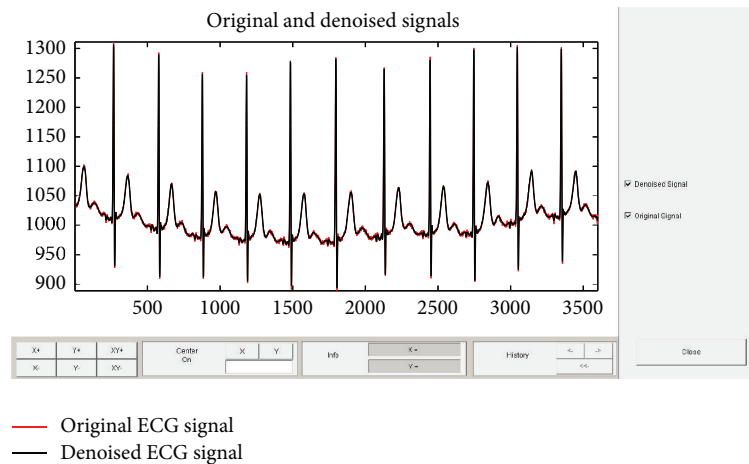


FIGURE 3: Original and denoised signals for ECG record no 103 m using Biorthogonal wavelet.

is reduced for this record. Figure 4 shows the approximation and coefficient details after the signal in record 114 m has been subjected to the Biorthogonal wavelet transform and Figure 5 shows the original and denoised electrocardiogram in the record 114 m. The signal to noise ratio obtained is 9.184 dB. This shows that the noise has to be removed further to enhance the quality of the obtained electrocardiogram.

(B) Performance Analysis of Multiwavelet Based Denoising Method for Electrocardiogram

Biorthogonal Based Multiwavelet Denoising. The multiwavelet denoising for the record 103 m shown in Figure 6 yields a signal to noise ratio of 35.5220 dB in which the power

line noise gets removed. Figure 7 shows the multiwavelet denoising for the record 114 m and the signal to noise ratio obtained is 13.4022 dB because of the removal of power line noise.

3.2. Performance Comparison of Wavelet and Multiwavelet Methods. Comparison of signal to noise ratio for wavelet and multiwavelet based denoising techniques for various electrocardiogram records.

4. Conclusion

The inference from Tables 1, 2, 3, and 4 is that the output signal to noise ratio value of multiwavelet denoising functions is

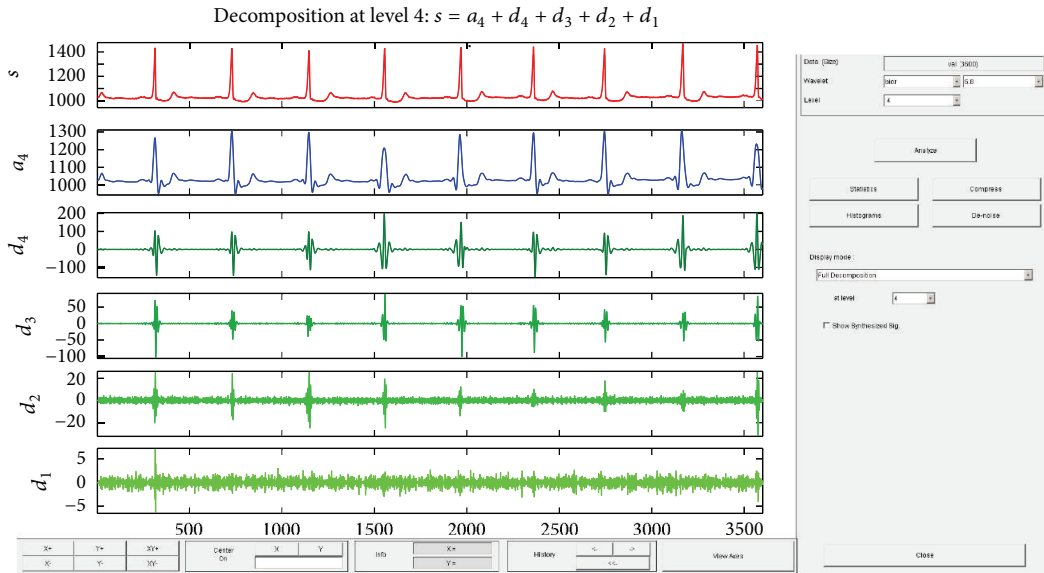


FIGURE 4: Wavelet denoising for ECG record no 114 m using Biorthogonal 1D wavelet.

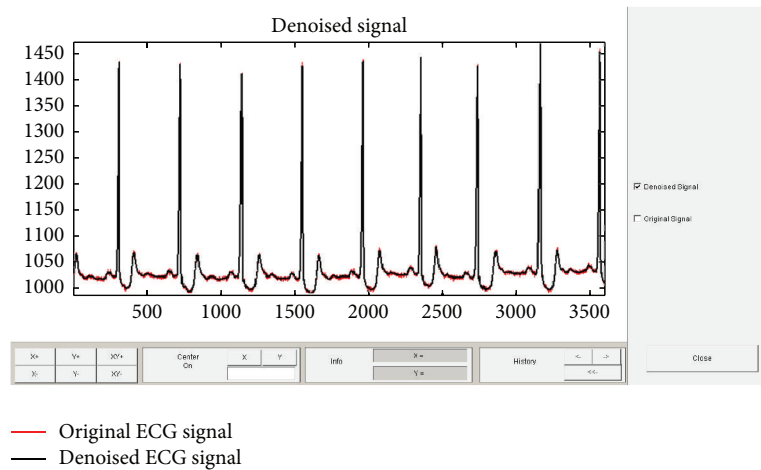


FIGURE 5: Original and denoised signals for ECG record no 114 m using Biorthogonal wavelet.

TABLE 1: Performance comparison of wavelet and multiwavelet for record 100 m.

| Wavelet family | SNR (dB) | |
|----------------|-----------------|--------------|
| | Record no 100 m | |
| | Wavelet | Multiwavelet |
| Bio 6.8 | 11.5708 | 29.8328 |
| Db 3 | 2.7306 | 21.7829 |
| Db10 | 0.561 | 7.3489 |
| Coif 4 | 2.466 | 3.6112 |
| Sym 4 | 9.0630 | 21.956 |

TABLE 2: Performance comparison of wavelet and multiwavelet for record 103 m.

| Wavelet family | SNR (dB) | |
|----------------|-----------------|--------------|
| | Record no 103 m | |
| | Wavelet | Multiwavelet |
| Bio 6.8 | 32.2094 | 35.5220 |
| Db 3 | 15.827 | 27.7849 |
| Db10 | 19.1099 | 33.1183 |
| Coif 4 | 24.1340 | 29.7007 |
| Sym 4 | 24.6220 | 36.0964 |

greater than the signal to noise value of wavelet functions. The table also indicates that the Daubechies wavelet 10 has better denoising capability than when compared to corresponding values of the wavelet denoising as the shape of this wavelet

is more close to the shape of electrocardiogram. The increase in signal to noise ratio value indicates that there is no loss in the information contained in the original electrocardiogram signal and multiwavelet has better denoising capability to

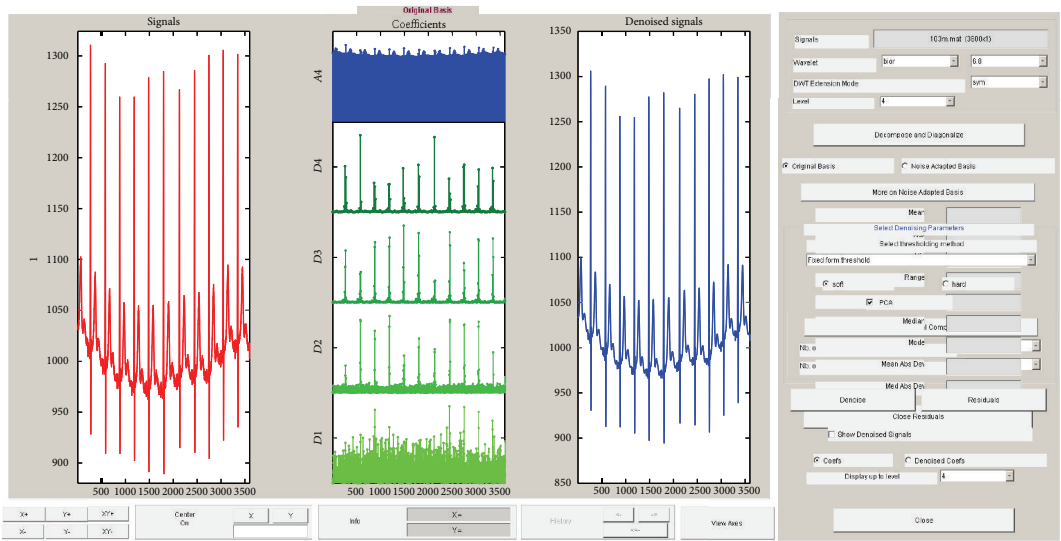


FIGURE 6: Multiwavelet denoising for ECG record no 103 m using Biorthogonal 1D wavelet.

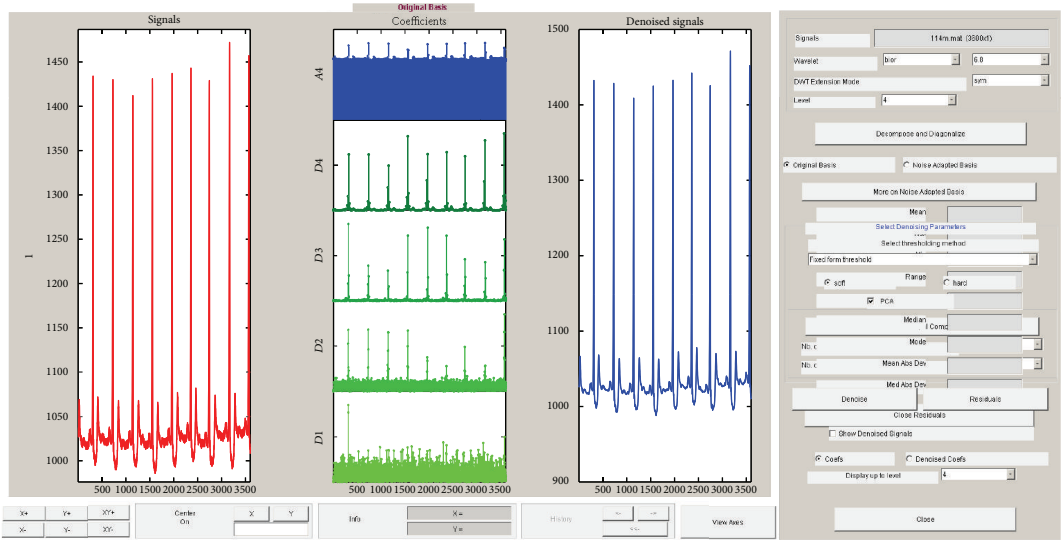


FIGURE 7: Multiwavelet denoising for ECG record no 114 m using Biorthogonal 1D wavelet.

TABLE 3: Performance comparison of wavelet and multiwavelet for record 114 m.

| Wavelet family | SNR (dB) | |
|----------------|----------|--------------|
| | Wavelet | Multiwavelet |
| Bio 6.8 | 9.184 | 13.4022 |
| Db 3 | 28.6913 | 31.0045 |
| Db10 | 18.2315 | 28.5193 |
| Coif 4 | 19.3813 | 36.6501 |
| Sym 4 | 29.6574 | 30.8971 |

TABLE 4: Performance comparison of wavelet and multiwavelet for record 201 m.

| Wavelet family | SNR (dB) | |
|----------------|----------|--------------|
| | Wavelet | Multiwavelet |
| Bio 6.8 | 11.5636 | 32.3220 |
| Db 3 | 20.1344 | 29.3001 |
| Db10 | 21.2696 | 33.7040 |
| Coif 4 | 17.8031 | 21.2874 |
| Sym 4 | 24.0735 | 25.1638 |

remove the power line noise in the various electrocardiogram records.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

References

- [1] C. Germaine, F. Halberg, T. Breus et al., "Non-photic solar associations of heart rate variability and myocardial infarction," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 64, no. 5, pp. 707–720, 2002.
- [2] M. M. Goudarzi, A. Taheri, M. Pooyan, and R. Mahboobi, "Multiwavelet and biological signal processing," *International Journal of Information Technology*, vol. 24, pp. 264–272, 2006.
- [3] M. Kania, M. Fereniec, and R. Maniewski, "Wavelet denoising for multi-lead high resolution ECG signals," *Measurement Science Review*, vol. 7, Section 2, no. 4, pp. 30–33, 2007.
- [4] K. Helenprabha and A. M. Natarajan, "FPGA implementation of gamma filter for extracting fetal electrocardiogram," *IETE Journal of Research*, no. 53, pp. 433–440, 2007.
- [5] F. C. Chang, C. K. Chang, K. Y. Chi, and Y. D. Lin, "Evaluation measures for adaptive PLI filters in ECG signal processing," *Computers in Cardiology*, vol. 34, pp. 529–532, 2007.
- [6] M. Alfaouri and K. Daqrouq, "ECG Signal Denoising by wavelet transform thresholding," *The American Journal of Applied Sciences*, vol. 53, pp. 276–281, 2008.
- [7] Z. Zhidong and M. Chan, "A novel cancellation method of powerline interference in ECG signal based on EMD and adaptive filter," in *Proceedings of the 11th IEEE International Conference on Communication Technology*, pp. 517–520, 2008.
- [8] M. Kaur and B. Singh, "Powerline interference reduction in ECG using combination of MA method and IIR notch," *International Journal of Recent Trends in Engineering*, vol. 2, no. 6, pp. 125–129, 2009.
- [9] A. K. M. F. Haque, M. H. Ali, M. A. Kiber, and M. T. Hasan, "Detection of small variations of ECG features using wavelet," *ARPJ Journal of Engineering and Applied Sciences*, vol. 4, no. 6, pp. 27–30, 2009.
- [10] Y.-F. Tan and D. Lei, "Study on Wavelet Transform in the processing for ECG signals," in *Proceedings of the WRI World Congress on Software Engineering (WCSE '09)*, vol. 9, pp. 515–518, IEEE Computer Society, May 2009.

Research Article

Capacity of 60 GHz Wireless Communications Based on QAM

Jingjing Wang,^{1,2} Na Li,³ Wei Shi,⁴ Yangyang Ma,¹ Xiulong Liang,¹ and Xinli Dong⁵

¹ College of Information Science & Technology, Qingdao University of Science & Technology, Qingdao, Shandong 266061, China

² State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, Jiangsu 210096, China

³ School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

⁴ Department of Electrical Engineering, Ocean University of China, Qingdao, Shandong 266100, China

⁵ China United Network Communications Corporation Qingdao Branch, Qingdao, Shandong 266072, China

Correspondence should be addressed to Jingjing Wang; kathy1003@163.com

Received 24 January 2014; Accepted 2 April 2014; Published 29 April 2014

Academic Editor: Feng Gao

Copyright © 2014 Jingjing Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With apparent advantages of the several GHz license-free spectrums, 10 W maximum transmit power, and so forth, 60 GHz wireless communication technology has become the first choice for Gbps level short-range wireless communications. This paper researches 60 GHz wireless communications over the additive white Gaussian noise channel. Channel capacity with quadrature amplitude modulation (QAM) is investigated for the unlicensed 59–64 GHz radio spectrum set aside by FCC. Moreover, the capacity with QAM is compared to that with phase shift keying (PSK). It is shown that QAM is capable of providing Gbps data rate and outperforms PSK especially when the modulation order is large. The results prove that QAM is an attractive scheme for 60 GHz wireless communications.

1. Introduction

The growth of wireless communications is spurred by the consumer desire for untethered access to information and entertainment. While contemporary unlicensed systems support light and moderate levels of wireless data traffic, as seen in Bluetooth and wireless local area networks (WLANs), current technology is unable to supply data rates comparable to wired standards like gigabit Ethernet and High-Definition Multimedia Interface (HDMI) [1]. An abundance of unlicensed spectrum surrounding the 60 GHz operating frequency has the ability to support these high-rate communications.

The 60 GHz band is an excellent choice for high-speed Internet, data, and voice communications since it offers benefits such as several GHz license-free spectrums, 10 W maximum transmit power, virtually interference-free operation, high level of frequency reuse enabled, and highly secure operation [2]. However, the 60 G wireless channel shows 20 to 40 dB increased free space path loss and suffers from 15 (up to 30) dB/km atmospheric absorption depending on

the atmospheric conditions. Multipath effects, except for indoor reflections, are vastly reduced at 60 G making non-line-of-sight (NLOS) communication very difficult [1, 3]. While the high path loss seems to be disadvantage at 60 GHz, it, however, confines the 60 GHz operation to within a room in an indoor environment. Hence, the effective interference levels for 60 GHz are less severe than those systems located in the congested 2–2.5 GHz and 5–5.8 GHz regions [2]. The oxygen absorption also enables higher “frequency reuse” since radiation from one particular 60 GHz radio link is quickly reduced to a level that will not interfere with other 60 GHz links operating in the same geographic vicinity [3]. Federal Communications Commission (FCC) set aside the 59–64 GHz frequency band for general unlicensed applications [4]. The effect of the antenna directionality to 60 GHz channel capacity is studied in [5]. The throughput of wireless mobile ad hoc networks with directional antennas at 60 GHz unlicensed band is investigated in [6]. The capacity analysis of 60 GHz wireless communications based on PSK modulation is given in [7, 8]. Quadrature amplitude modulation (QAM) is widely used for the high-speed data transmission [9–11].

Compared with other digital modulation techniques like PSK or PAM, QAM modulation has better anti-noise performance and could make full use of the bandwidth.

In this paper, we investigate the capacity of 60 GHz wireless communication system over AGWN channel under the FCC rules. The major modulation method used here is QAM, and capacity comparison between QAM and PSK is also simply illustrated. The rest of the paper is organized as follows. Section 2 presents the general used QAM constellations and makes a comparison between two different constellations for 8-QAM. Section 3 calculates channel capacity over AWGN channel in 60 GHz wireless communication system. Section 4 conducts Monte Carlo simulations to illustrate the channel capacity. And Section 5 gives a conclusion.

2. QAM Constellations

QAM can be viewed as combined amplitude and phase modulation. When the requirement of data transfer rate exceeds the upper limit 8-PSK can provide, QAM is generally used. Because the QAM constellation points are much more disperse than PSK constellation points and the distances between the constellation points are much bigger with the same ary. So QAM modulation could provide a better transmission performance.

QAM signal waveforms may be expressed as [10]

$$S_m(t) = A_{mc}g(t) \cos 2\pi f_c t - A_{ms}g(t) \sin 2\pi f_c t, \quad (1)$$

where A_{mc} and A_{ms} are the information-bearing signal amplitudes of the quadrature carriers and $g(t)$ is the signal pulse. The vector representation of these waveforms is

$$S_m = \left[A_{mc} \sqrt{\frac{1}{2}} \varepsilon_g \quad A_{ms} \sqrt{\frac{1}{2}} \varepsilon_g \right], \quad (2)$$

where ε_g is the energy of the basic signal pulse $g(t)$.

M-QAM constellations can be constructed in many different ways, and they have different capacity and error characteristics. Although rectangular, circle, and star signal constellations are common in practice, a certain kind of constellation can be designed to achieve the best communication performance, under some specific premises [12, 13].

Figures 1 and 2 present two 8-QAM constellations. Figure 1 is a rectangular 8-QAM constellation and Figure 2 is a circular 8-QAM constellation.

Assuming that the signal points are equally probable, the average transmitted signal power is [4]

$$P_{av} = \frac{1}{M} \sum_{m=1}^M (A_{mc}^2 + A_{ms}^2) = \frac{A^2}{M} \sum_{m=1}^M (a_{mc}^2 + a_{ms}^2), \quad (3)$$

where (a_{mc}, a_{ms}) are the coordinates of the signal points, normalized by A .

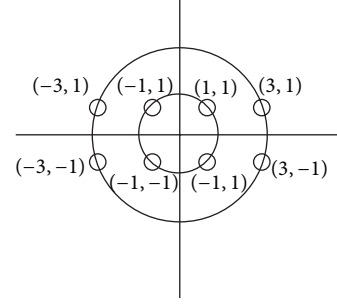


FIGURE 1: Rectangular 8-QAM constellation.

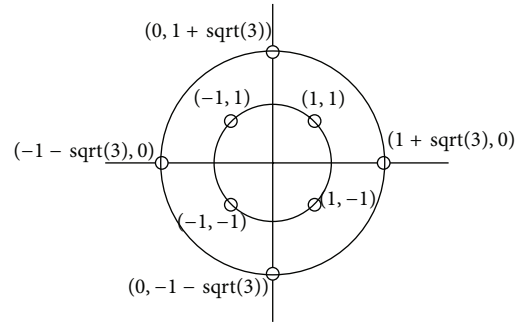


FIGURE 2: Circle 8-QAM constellation.

As can be seen from the above figures, the minimum distances between the constellation points for (a) and (b) are, respectively,

$$\begin{aligned} d_a &= 2\sqrt{\frac{1}{6}}, \\ d_b &= 2\sqrt{\frac{1}{3 + \sqrt{3}}}. \end{aligned} \quad (4)$$

Comparing both,

$$\text{Ratio} = \frac{d_b}{d_a} = \sqrt{\frac{6}{3 + \sqrt{3}}} \approx 1.126, \quad (5)$$

$$\text{Ratio}_{dB} = 20 \log(1.126) \approx 1.03.$$

Minimum distance of signal set shown in Figure 1 is approximately 1 dB less than that shown in Figure 4 with the same average transmitted power. The more the distance between the constellations, the less the chance of a constellation point getting decoded incorrectly. Actually, the second signal constellation is the optimal one for 8-QAM because it has the largest minimum Euclidean distance between signal points for a given transmitted power. At the same time, as shown in Figure 5, signal set with the circle constellation for 8-QAM provides a higher data rate.

Rectangular QAM signal constellations have distinct advantage of being easily generated and transmitted as two PAM signals impressed on phase-quadrature carriers. In addition, they are easily demodulated. Although, it is

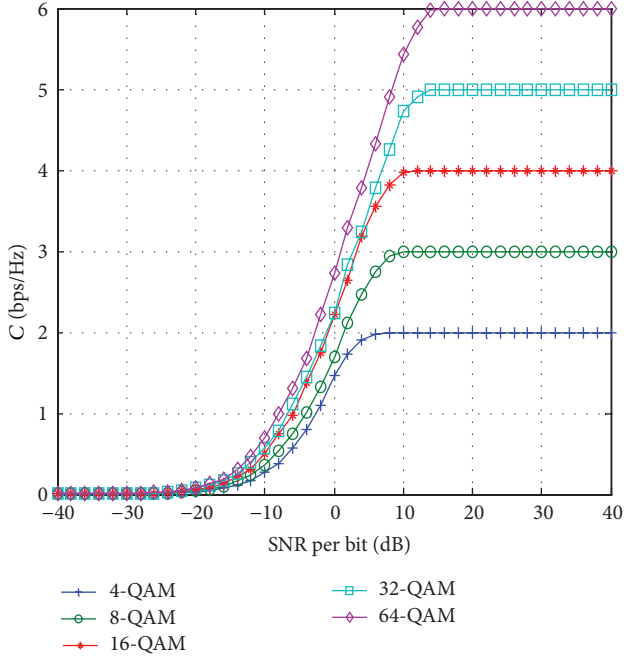


FIGURE 3: Channel capacity of M -ary QAM system over AGWN channels.

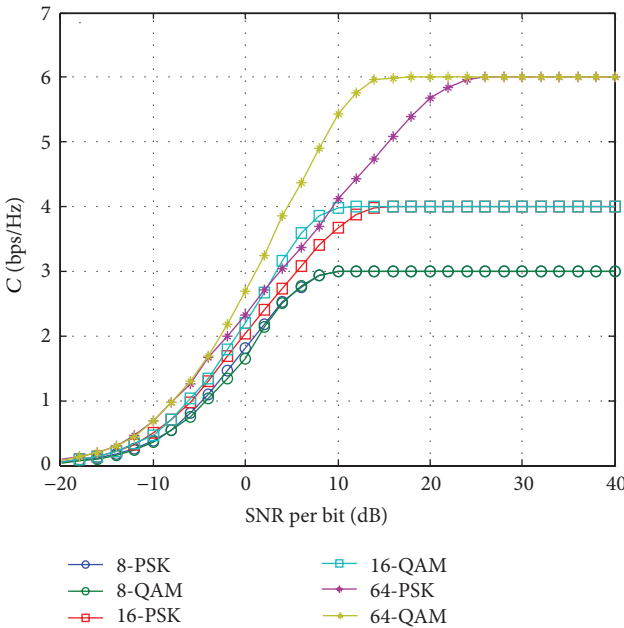


FIGURE 4: Relations between capacities of M -ary PSK and QAM with the same average power.

generally a sub-optimal modulation scheme, compared to other M -QAM constellations, in the sense that they do not maximally space the constellation points for a given energy. For $M \geq 16$, the minimum distance required to achieve a given average transmitted power is only slightly smaller than the minimum distance required for the optimal M -ary

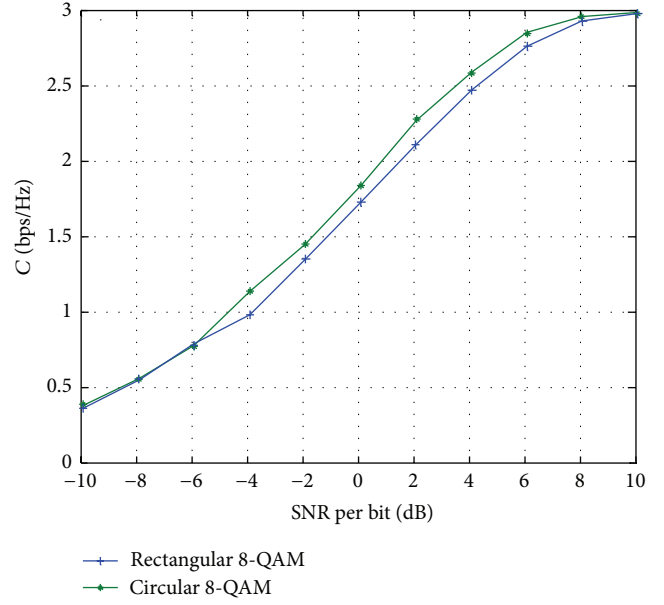


FIGURE 5: Capacity comparison between two different 8-QAM signal constellations.

QAM signal constellation. For these reasons, rectangular M -ary QAM signals are most frequently used in practice [10]. And they are also adopted in this paper.

3. Channel Capacity

In general, the channel capacity is a function of the channel realization, transmitted signal power and noise. For AWGN channel, the shannon capacity is normalized with respect to the bandwidth and expressed in bps, that is, normalized with respect to the bandwidth, is

$$C = W \log_2 (1 + \text{SNR}), \quad (6)$$

where W is the system band width and SNR is the receive signal to noise ratio, defined by ε_b/N_0 , where ε_b is the energy per bit [9].

The Shannon capacity predicts the channel capacity C for an AWGN channel with continuous-valued inputs and outputs. However, a channel employing multilevel/phase modulation, for example, PAM, PSK, or QAM modulation, has discrete-valued inputs and continuous-valued outputs, which impose an additional constraint on the capacity calculation [10].

We consider the modulation channels with discrete-input X and continuous-output Y , which is defined as [7, 9]

$$Y = X + W, \quad (7)$$

where W is a zero-mean Gaussian random variable with variance σ^2 and $X = x_k$, $k = 0, \dots, q - 1$. For a given X , it follows that Y is Gaussian with mean x_k and variance σ^2 . That is,

$$p\left(\frac{y}{x_k}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-x_k)^2/2\sigma^2}. \quad (8)$$

The capacity of this channel in bits per channel use is the maximum average mutual information between the discrete input $X = \{x_0, x_1, \dots, x_{q-1}\}$ and the output $Y = \{-\infty, \infty\}$. That is,

$$C = \max_{p(x_i)} \sum_{i=0}^{q-1} \int_{-\infty}^{\infty} p(y | x_i) p(x_i) \log_2 \frac{p(y | x_i)}{p(y)} dy, \quad (9)$$

where

$$p(y) = \sum_{k=0}^{q-1} p(y | x_k) p(x_k). \quad (10)$$

Assuming an equal a priori probability real or complex signal constellation, that is, $p(x_i) = 1/q$, the channel capacity of an AWGN channel with q -ary modulation is then [10]

$$\begin{aligned} C &= \log_2(q) - \frac{1}{q} \sum_{k=0}^{q-1} \mathbb{E}_{y|x_k} \left\{ \log_2 \frac{\sum_{i=0}^{q-1} p(y | x_i)}{p(y | x_k)} \right\} \\ &= \log_2(q) \\ &\quad - \frac{1}{q} \sum_{k=0}^{q-1} \mathbb{E}_{y|x_k} \left\{ \log_2 \sum_{i=0}^{q-1} \exp \left[-\frac{|x_k + w - x_i|^2 - |w|^2}{2\sigma^2} \right] \right\}, \end{aligned} \quad (11)$$

where $\mathbb{E}[\cdot]$ is the expected value operator and w is the complex white Gaussian noise, modeled as a Gaussian distributed random variable with zero mean and variance σ^2 in each real dimension. Equation (11) is a universal formula applied to q -ary PAM/PSK/QAM and can be evaluated by Monte Carlo simulation. With normalized signal energy, the relationships between channel capacity and SNR can be evaluated by (11).

4. Experimental Results and Analysis

In this paper, Monte Carlo simulations are conducted to present the channel capacity of 60 GHz over AWGN channels under FCC regulations.

Figure 3 shows the normalized channel capacity for M -ary QAM system over AGWN channels. It is shown that the achievable data rate is 8.42 Gbps for a 60 GHz with 5 GHz bandwidth at a SNR of 0 dB for 8-QAM system. And for 16-QAM and 64-QAM, the data rate can be 11.14 Gbps and 13.19 Gbps. Hence QAM has potential to support Gbps data transmission in the 60 GHz system.

Figure 4 shows comparison of channel capacities for M -ary PSK and QAM systems with the same average transmitted power. It shows that the data rate for M -QAM is higher than that for M -PSK, especially when $M > 8$. For 16-QAM, the data rate can improve 6.8% at a SNR of 0 dB and for 64-QAM the improvement can reach 14.1%. That is to say, QAM can achieve a higher data rate even at a lower SNR. We can conclude from this figure that when $M > 8$, capacity performance of the QAM system is better than that of the PSK system. The superiority of QAM is obvious because it has the largest minimum Euclidean distance between signal points for a given transmitted power.

Figure 5 demonstrates the difference among channel capacities for different M -QAM constellations. It shows that, for 8-QAM, the circle constellation provides a higher data rate. This confirms our analysis in Section 2. Moreover, the capacity advantage of QAM with circle constellation over that with rectangular constellation is quite small, whereas the latter is much easier to implement in practice. Hence rectangular QAM modulation is more preferable for 60 GHz wireless communications.

5. Conclusions

PSK is the common used modulation for 60 GHz currently, because of its advantages in bandwidth and SNR. However, the data rate for M -PSK is obviously lower than that of M -QAM, with the same average transmitted power. Moreover, as M increases, the distance between the adjacent phases gradually decreases, which reduces the noise tolerance and makes it difficult to guarantee the error rate, while QAM can improve the noise tolerance and provide a lower error rate.

For M -QAM, many different signal constellations can be designed and conducted, from which we can select an optimal one to meet our specific requirements in 60 GHz wireless communication.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the referees and editors for providing very helpful comments and suggestions. This project was supported by Key Laboratory of Universal Wireless Communications (Beijing University of Posts and Telecommunications), Ministry of Education, China (no. KFKT-2013102), National Natural Science Foundation of China (no. 61304222), Natural Science Foundation of Shandong Province (no. ZR2012FQ021), Shandong Province Higher Educational Science and Technology Program (no. J12LN88), and Open Project of State Key Laboratory of Millimeter Waves (no. K201321).

References

- [1] R. C. Daniels and R. W. Heath Jr., "60 GHz wireless communications: emerging requirements and design recommendations," *IEEE Vehicular Technology Magazine*, vol. 2, no. 3, pp. 41–50, 2007.
- [2] P. Cheolhee and T. S. Rappaport, "Short-range wireless communications for next-generation networks: UWB 60 GHz millimeter-wave wpan, and ZigBee," *IEEE Wireless Communications*, vol. 14, no. 4, pp. 70–78, 2007.
- [3] S. K. Yong and C.-C. Chong, "An overview of multigigabit wireless through millimeter wave technology: potentials and technical challenges," *Eurasip Journal on Wireless Communications and Networking*, vol. 2007, Article ID 78907, 2007.

- [4] Federal Communications Commission, *Amendment of Parts 2, 15 and 97 of the Commission's Rules To Permit Use of Radio Frequencies above 40 GHz For New Radio Applications*, 1995.
- [5] A. Seyedi, "On the capacity of wideband 60GHz channels with antenna directionality," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 4532–4536, November 2007.
- [6] M. Alimadadi, A. Mohammadi, and M. D. Soltani, "Throughput analysis of Ad-Hoc networks with directional antenna at 60GHz," *Journal of Electromagnetic Waves and Applications*, vol. 28, no. 2, pp. 228–241, 2014.
- [7] H. Zhang and T. A. Gulliver, "On the capacity of 60 GHz wireless communications," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE '09)*, pp. 936–939, May 2009.
- [8] J. Wang, H. Zhang, T. Lv, and G. T. Aaron, "Capacity of 60 GHz wireless communication systems over fading channels," *Journal of Networks*, vol. 7, no. 1, pp. 203–209, 2012.
- [9] K. Chris, *The Benefits of 60 GHz Unlicensed Wireless Communications*, YDI Wireless Whitepaper, 2002.
- [10] J. G. Proakis, *Digital Communications*, Publishing House of Electronics Industry, 2006.
- [11] H. Zhang and T. A. Gulliver, "Capacity and error probability analysis for orthogonal space-time block codes over fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 808–819, 2005.
- [12] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Transactions on Communications*, vol. 22, no. 1, pp. 28–38, 1974.
- [13] C.-E. W. Sundberg, W. C. Wong, and R. Steele, "Logarithmic PCM weighted QAM transmission over Gaussian and Rayleigh fading channels," *IEE Proceedings F: Communications, Radar and Signal Processing*, vol. 134, no. 6, pp. 557–570, 1987.

Research Article

The Determinants, Inverses, Norm, and Spread of Skew Circulant Type Matrices Involving Any Continuous Lucas Numbers

Jin-jiang Yao and Zhao-lin Jiang

Department of Mathematics, Linyi University, Linyi, Shandong 276000, China

Correspondence should be addressed to Zhao-lin Jiang; jzh1208@sina.com

Received 6 January 2014; Accepted 17 February 2014; Published 17 April 2014

Academic Editor: Feng Gao

Copyright © 2014 J.-j. Yao and Z.-l. Jiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the skew circulant and skew left circulant matrices with any continuous Lucas numbers. Firstly, we discuss the invertibility of the skew circulant matrices and present the determinant and the inverse matrices by constructing the transformation matrices. Furthermore, the invertibility of the skew left circulant matrices is also discussed. We obtain the determinants and the inverse matrices of the skew left circulant matrices by utilizing the relationship between skew left circulant matrices and skew circulant matrix, respectively. Finally, the four kinds of norms and bounds for the spread of these matrices are given, respectively.

1. Introduction

Circulant and skew-circulant matrices are appearing increasingly often in scientific and engineering applications. Briefly, scanning the recent literature, one can see their utility is appreciated in the design of digital filters [1–3], image processing [4–6], communications [7], signal processing [8], and encoding [9]. They have been put on firm basis with the work of Davis [10] and Jiang and Zhou [11].

The skew circulant matrices as preconditioners for linear multistep formulae- (LMF-) based ordinary differential equations (ODEs) codes. Hermitian and skew-Hermitian Toeplitz systems are considered in [12–15]. Lyness and Sørøvik employed a skew circulant matrix to construct s -dimensional lattice rules in [16]. Spectral decompositions of skew circulant and skew left circulant matrices were discussed in [17]. Compared with cyclic convolution algorithm, the skew cyclic convolution algorithm [8] is able to perform filtering procedure in approximate half of computational cost for real signals. In [2] two new normal-form realizations are presented which utilize circulant and skew circulant matrices as their state transition matrices. The well-known second-order coupled form is a special case of the skew circulant form. Li et al. [18] gave the style spectral decomposition

of skew circulant matrix firstly and then dealt with the optimal backward perturbation analysis for the linear system with skew circulant coefficient matrix. In [3], a new fast algorithm for optimal design of block digital filters (BDFs) was proposed based on skew circulant matrix.

Besides, some scholars have given various algorithms for the determinants and inverses of nonsingular circulant matrices [10, 11]. Unfortunately, the computational complexity of these algorithms is very amazing with the order of matrix increasing. However, some authors gave the explicit determinants and inverse of circulant and skew circulant involving some famous numbers. For example, Jaiswal evaluated some determinants of circulant whose elements are the generalized Fibonacci numbers [19]. Lind presented the determinants of circulant and skew circulant involving the Fibonacci numbers [20]. Dazheng [21] gave the determinant of the Fibonacci-Lucas quasicyclic matrices. Shen et al. considered circulant matrices with the Fibonacci and Lucas numbers and presented their explicit determinants and inverses by constructing the transformation matrices [22]. Gao et al. [23] gave explicit determinants and inverses of skew circulant and skew left circulant matrices with the Fibonacci and Lucas numbers. Jiang et al. [24, 25] considered the skew circulant

and skew left circulant matrices with the k -Fibonacci numbers and the k -Lucas numbers and discussed the invertibility of these matrices and presented their determinant and the inverse matrix by constructing the transformation matrices, respectively.

Recently, there are several papers on the norms of some special matrices. Solak [26] established the lower and upper bounds for the spectral norms of circulant matrices with the classical Fibonacci and Lucas numbers entries. İpek [27] investigated an improved estimation for spectral norms of these matrices. Shen and Cen [28] gave upper and lower bounds for the spectral norms of r -circulant matrices in the forms of $A = C_r(F_0, F_1, \dots, F_{n-1})$, $B = C_r(L_0, L_1, \dots, L_{n-1})$, and they also obtained some bounds for the spectral norms of Kronecker and Hadamard products of matrix A and matrix B . Akbulak and Bozkurt [29] found upper and lower bounds for the spectral norms of Toeplitz matrices such that $a_{ij} \equiv F_{i-j}$ and $b_{ij} \equiv L_{i-j}$. The convergence in probability and in distribution of the spectral norm of scaled Toeplitz, circulant, reverse circulant, symmetric circulant, and a class of k -circulant matrices is discussed in [30].

Beginning with Mirsky [31], several authors [32–38] have obtained bounds for the spread of a matrix.

The purpose of this paper is to obtain the explicit determinants, explicit inverses, norm, and spread of skew circulant type matrices involving any continuous Lucas numbers. And we generalize the result [23]. In passing, the norm and spread of skew circulant type matrices have not been researched. It is hoped that this paper will help in changing this. More work continuing the present paper is forthcoming.

In the following, let r be a nonnegative integer. We adopt the following two conventions $0^0 = 1$, and, for any sequence $\{a_n\}$, $\sum_{k=i}^n a_k = 0$ in the case $i > n$.

The Lucas sequences are defined by the following recurrence relations [21–23, 27–29]:

$$L_{n+1} = L_n + L_{n-1}, \quad \text{where } L_0 = 2, L_1 = 1, \quad (1)$$

for $n \geq 0$. The first few values of the sequences are given by the following table:

$$\begin{array}{c|cccccccccc} n & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline L_n & 2 & 1 & 3 & 4 & 7 & 11 & 18 & 29 & 47 & 76 \end{array}. \quad (2)$$

The $\{L_n\}$ is given by the formula

$$L_n = \alpha^n + \beta^n, \quad (3)$$

where α and β are the roots of the characteristic equation $x^2 - x - 1 = 0$.

Definition 1 (see [17]). A skew circulant matrix over C with the first row (a_1, a_2, \dots, a_n) is meant a square matrix of the form

$$\begin{pmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ -a_n & a_1 & a_2 & \dots & a_{n-1} \\ \vdots & -a_n & a_1 & \ddots & \vdots \\ -a_3 & \vdots & \ddots & \ddots & a_2 \\ -a_2 & -a_3 & \dots & -a_n & a_1 \end{pmatrix}_{n \times n}, \quad (4)$$

denoted by $\text{SCirc}(a_1, a_2, \dots, a_n)$.

Definition 2 (see [17]). A skew left circulant matrix over C with the first row (a_1, a_2, \dots, a_n) is meant a square matrix of the form

$$\begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_2 & a_3 & \dots & a_n & -a_1 \\ a_3 & \ddots & \ddots & \ddots & \vdots \\ \vdots & a_n & -a_1 & \dots & -a_{n-2} \\ a_n & -a_1 & \dots & -a_{n-2} & -a_{n-1} \end{pmatrix}_{n \times n}, \quad (5)$$

denoted by $\text{SLCirc}(a_1, a_2, \dots, a_n)$.

Lemma 3 (see [10, 17]). Let $A = \text{SCirc}(a_1, a_2, \dots, a_n)$ be skew circulant matrix; then

(i) A is invertible if and only if the eigenvalues of A

$$\lambda_k = f(\omega^k \eta) \neq 0, \quad (k = 0, 1, 2, \dots, n-1), \quad (6)$$

where $f(x) = \sum_{j=1}^n a_j x^{j-1}$, $\omega = \exp(2\pi i/n)$, and $\eta = \exp(\pi i/n)$;

(ii) if A is invertible, then the inverse of A is a skew circulant matrix.

Lemma 4 (see [17]). Let $A = \text{SLCirc}(a_1, a_2, \dots, a_n)$ be skew left circulant matrix and let n be odd; then

$$\lambda_j = \pm \left| \sum_{k=1}^n a_k \omega^{(j-(1/2))(k-1)} \right|, \quad \left(j = 1, 2, \dots, \frac{n-1}{2} \right), \quad (7)$$

$$\lambda_{(n+1)/2} = \sum_{k=1}^n |a_k (-1)^{k-1}|,$$

where λ_j , $j = 1, 2, \dots, (n-1)/2$, $(n+1)/2$ are the eigenvalues of A .

Lemma 5 (see [23]). With the orthogonal skew left circulant matrix

$$\Theta := \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & -1 \\ 0 & 0 & \dots & -1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & -1 & \dots & 0 & 0 \end{pmatrix}_{n \times n}, \quad (8)$$

it holds that

$$\text{SCirc}(a_1, a_2, \dots, a_n) = \Theta \text{SLCirc}(a_1, a_2, \dots, a_n). \quad (9)$$

Lemma 6 (see [23]). If

$$[\text{SCirc}(a_1, a_2, \dots, a_n)]^{-1} = \text{SCirc}(b_1, b_2, \dots, b_n), \quad (10)$$

then

$$[\text{SLCirc}(a_1, a_2, \dots, a_n)]^{-1} = \text{SLCirc}(b_1, -b_n, \dots, -b_2). \quad (11)$$

Lemma 7 (see [27, 28]). Let $\{L_n\}$ be the Lucas numbers; then

$$(i) \sum_{i=0}^{n-1} L_i = L_{n+1} - 1, \quad (12)$$

$$(ii) \sum_{i=0}^{n-1} L_i^2 = L_n L_{n-1} + 2, \quad (13)$$

$$(iii) \sum_{i=0}^{n-1} i L_i = (n-1) L_{n+1} - L_{n+2} + 4. \quad (14)$$

Definition 8 (see [29]). Let $A = (a_{ij})$ be an $n \times n$ matrix. The maximum column sum matrix norm, the spectral norm, the Euclidean (or Frobenius) norm, and the maximum row sum matrix norm of matrix A are, respectively,

$$\begin{aligned} \|A\|_1 &= \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|, \\ \|A\|_2 &= \left(\max_{1 \leq i \leq n} \lambda_i(A^* A) \right)^{1/2}, \\ \|A\|_F &= \left(\sum_{i,j=1}^n |a_{ij}|^2 \right)^{1/2}, \\ \|A\|_\infty &= \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|, \end{aligned} \quad (15)$$

where A^* denotes the conjugate transpose of A .

Lemma 9 (see [30]). If A is an $n \times n$ real symmetric or normal matrix, then one has

$$\|A\|_2 = \max_{1 \leq i \leq n} |\lambda_i|, \quad (16)$$

where λ_i ($i = 1, 2, \dots, n$) are the eigenvalues of A .

Definition 10 (see [31, 32]). Let $A = (a_{ij})$ be an $n \times n$ matrix with eigenvalues λ_i , $i = 1, 2, \dots, n$. The spread of A is defined as

$$s(A) = \max_{i,j} |\lambda_i - \lambda_j|. \quad (17)$$

Beginning with Mirsky [31], several authors [32–38] have obtained bounds for the spread of a matrix.

Lemma 11. Let $A = (a_{ij})$ be an $n \times n$ matrix. An upper bound for the spread due to Mirsky [31] states that

$$s(A) \leq \sqrt{2\|A\|_F^2 - \frac{2}{n}|\text{tr } A|^2}, \quad (18)$$

where $\|A\|_F$ denotes the Frobenius norm of A and $\text{tr } A$ is trace of A .

Lemma 12 (see [38]). Let $A = (a_{ij})$ be an $n \times n$ matrix; then

(i) if A is real and normal, then

$$s(A) \geq \frac{1}{n-1} \left| \sum_{i \neq j} a_{ij} \right|, \quad (19)$$

(ii) and if A is Hermitian, then

$$s(A) \geq 2 \max_{i \neq j} |a_{ij}|. \quad (20)$$

2. Determinant and Inverse of Skew Circulant Matrix with the Lucas Numbers

In this section, let $A_{r,n} = \text{SCirc}(L_{r+1}, \dots, L_{r+n})$ be skew circulant matrix. Firstly, we give a determinant explicit formula for the matrix $A_{r,n}$. Afterwards, we prove that $A_{r,n}$ is an invertible matrix for $n \geq 2$, and then we find the inverse of the matrix $A_{r,n}$.

In the following, let $x = -((L_r + L_{r+n})/(L_{r+1} + L_{r+n+1}))$, $t = L_{r+2}/L_{r+1}$, $c = L_{r+1} + L_{r+n+1}$, $d = L_r + L_{r+n}$, $l_n = L_{r+1} + tL_{r+n} + \sum_{k=1}^{n-2} (tL_{r+k+1} - L_{r+k+2}) \cdot x^{n-(k+1)}$ and $l'_n = \sum_{k=1}^{n-1} L_{r+k+1} \cdot x^{n-(k+1)}$.

Theorem 13. Let $A_{r,n} = \text{SCirc}(L_{r+1}, \dots, L_{r+n})$ be skew circulant matrix; then

$$\begin{aligned} \det A_{r,n} &= L_{r+1} \left[L_{r+1} + tL_{r+n} \right. \\ &\quad \left. + \sum_{i=1}^{n-2} (tL_{r+i+1} - L_{r+i+2}) x^{n-(i+1)} \right] \cdot c^{n-2}, \end{aligned} \quad (21)$$

where L_{r+n} is the $(r+n)$ th Lucas number. Specially, when $r = 0$, one gets the result of [23].

Proof. Obviously, $\det A_{r,1} = L_{r+1}$ satisfies the equation. In the case $n > 1$, let

$$\Sigma = \begin{pmatrix} 1 & & & & & \\ t & & & & & \\ 1 & & & 1 & -1 & \\ 0 & & 0 & 1 & -1 & -1 \\ \vdots & & & & \ddots & \ddots \\ 0 & & 1 & \ddots & \ddots & \ddots \\ 0 & 1 & -1 & \ddots & \ddots & 0 \\ 0 & 1 & -1 & -1 & \ddots & \ddots \end{pmatrix}, \quad (22)$$

$$\Omega_1 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & x^{n-2} & 0 & \cdots & 0 & 1 \\ 0 & x^{n-3} & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & x & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

be two $n \times n$ matrices; then we have

$$\Sigma A_{r,n} \Omega_1 = \begin{pmatrix} L_{r+1} & l'_n & c_{13} & \cdots & c_{1,n-1} & c_{1n} \\ 0 & l'_n & c_{23} & \cdots & c_{2,n-1} & c_{2n} \\ 0 & 0 & c & & & \\ 0 & 0 & d & \ddots & & \\ \vdots & \vdots & & \ddots & c & \\ 0 & 0 & & & d & c \end{pmatrix}, \quad (23)$$

where

$$c_{1j} = L_{r+n+2-j}, \quad c_{2j} = tL_{r+n+2-j} - L_{r+n+3-j}, \quad (j = 3, 4, \dots, n). \quad (24)$$

So it holds that

$$\begin{aligned} & \det \Sigma \det A_{r,n} \det \Omega_1 \\ &= L_{r+1} \left[L_{r+1} + tL_{r+n} \right. \\ & \quad \left. + \sum_{k=1}^{n-2} (tL_{r+k+1} - L_{r+k+2}) x^{n-(i+1)} \right] \\ & \quad \cdot (L_{r+1} + L_{r+n+1})^{n-2}. \end{aligned} \quad (25)$$

While taking $\det \Sigma = \det \Omega_1 = (-1)^{(n-1)(n-2)/2}$, we have

$$\begin{aligned} & \det A_{r,n} \\ &= L_{r+1} \left[L_{r+1} + tL_{r+n} \right. \\ & \quad \left. + \sum_{k=1}^{n-2} (tL_{r+k+1} - L_{r+k+2}) x^{n-(i+1)} \right] \\ & \quad \cdot (L_{r+1} + L_{r+n+1})^{n-2}. \end{aligned} \quad (26)$$

This completes the proof. \square

Theorem 14. Let $A_{r,n} = \text{SCirc}(L_{r+1}, \dots, L_{r+n})$ be skew circulant matrix; then $A_{r,n}$ is an invertible matrix. Specially, when $r = 0$, one gets the result of [23].

Proof. Taking $n = 2$ in, Theorem 13, we have $\det A_{r,2} = L_{r+1}^2 + L_{r+2}^2 \neq 0$. Hence $A_{r,2}$ is invertible. In the case $n > 2$, since $L_{r+n} = \alpha^{r+n} + \beta^{r+n}$, where $\alpha + \beta = 1$, $\alpha\beta = -1$, we have

$$\begin{aligned} f(\omega^k \eta) &= \sum_{j=1}^n L_{r+j} (\omega^k \eta)^{j-1} \\ &= \sum_{j=1}^n (\alpha^{r+j} + \beta^{r+j}) (\omega^k \eta)^{j-1} \\ &= \frac{\alpha^{r+1} (1 + \alpha^n)}{1 - \alpha \omega^k \eta} + \frac{\beta^{r+1} (1 + \beta^n)}{1 - \beta \omega^k \eta} \\ &= \frac{L_{r+1} + L_{r+n+1} + (L_r + L_{r+n}) \omega^k \eta}{1 - \omega^k \eta - \omega^{2k} \eta^2} \\ & \quad (k = 1, 2, \dots, n-1), \end{aligned} \quad (27)$$

where $\omega = \exp(2\pi i/n)$, $\eta = \exp(\pi i/n)$. If there exists $\omega^l \eta$ ($l = 1, 2, \dots, n-1$) such that $f(\omega^l \eta) = 0$, we obtain $L_{r+1} + L_{r+n+1} + (L_r + L_{r+n}) \omega^l \eta = 0$, for $1 - \omega^l \eta - \omega^{2l} \eta^2 \neq 0$, and hence it follows

that $\omega^l \eta = -((L_{r+1} + L_{r+n+1})/(L_r + L_{r+n}))$ is a real number. Since

$$\begin{aligned} \omega^l \eta &= \exp \frac{(2l+1)\pi i}{n} \\ &= \cos \frac{(2l+1)\pi}{n} + i \sin \frac{(2l+1)\pi}{n}, \end{aligned} \quad (28)$$

it yields that $\sin((2l+1)\pi/n) = 0$, so we have $\omega^l \eta = -1$ for $0 < (2l+1)\pi/n < 2\pi$. Since $x = -1$ is not the root of the equation $L_{r+1} + L_{r+n+1} + (L_r + L_{r+n})x = 0$ ($n > 2$). We obtain $f(\omega^k \eta) \neq 0$, for any $\omega^k \eta$ ($k = 1, 2, \dots, n-1$), while

$$\begin{aligned} f(\eta) &= \sum_{j=1}^n L_j \eta^{j-1} \\ &= \frac{L_{r+1} + L_{r+n+1} + (L_r + L_{r+n}) \eta}{1 - \eta - \eta^2} \neq 0. \end{aligned} \quad (29)$$

It follows from Lemma 3 that the conclusion holds. \square

Lemma 15. Let the matrix $\mathcal{H} = [h_{ij}]_{i,j=1}^{n-2}$ be of the form

$$h_{ij} = \begin{cases} L_{r+1} + L_{r+n+1} = c, & i = j, \\ L_r + L_{r+n} = d, & i = j+1, \\ 0, & \text{otherwise.} \end{cases} \quad (30)$$

Then the inverse $\mathcal{H}^{-1} = [h'_{ij}]_{i,j=1}^{n-2}$ of the matrix \mathcal{H} is equal to

$$h'_{ij} = \begin{cases} \frac{(-d)^{i-j}}{c^{i-j+1}}, & i \geq j, \\ 0, & i < j. \end{cases} \quad (31)$$

Specially, when $r = 0$, one gets the result of [23].

Proof. Let $e_{ij} = \sum_{k=1}^{n-2} h_{ik} h'_{kj}$. Obviously, $e_{ij} = 0$ for $i < j$. In the case $i = j$, we obtain $e_{ii} = h_{ii} h'_{ii} = (L_{r+1} + L_{r+n+1}) \cdot (1/(L_{r+1} + L_{r+n+1})) = 1$. For $i \geq j+1$, we obtain

$$\begin{aligned} e_{ij} &= \sum_{k=1}^{n-2} h_{ik} h'_{kj} = h_{i,i-1} h'_{i-1,j} + h_{ii} h'_{ij} \\ &= d \cdot \frac{(-d)^{i-j-1}}{c^{i-j}} + c \cdot \frac{(-d)^{i-j}}{c^{i-j+1}} = 0. \end{aligned} \quad (32)$$

Hence, we get $\mathcal{H} \mathcal{H}^{-1} = I_{n-2}$, where I_{n-2} is $(n-2) \times (n-2)$ identity matrix. Similarly, we can verify that $\mathcal{H}^{-1} \mathcal{H} = I_{n-2}$. Thus, the proof is completed. \square

Theorem 16. Let $A_{r,n} = \text{SCirc}(L_{r+1}, \dots, L_{r+n})$ be skew circulant matrix; then

$$(A_{r,n})^{-1} = \frac{1}{l_n} \cdot \text{SCirc}(y'_1, y'_2, \dots, y'_n), \quad (33)$$

where

$$\begin{aligned}
 y'_1 &= 1 - \left[(L_{r+3} - tL_{r+2}) \cdot \frac{(-d)^{n-3}}{c^{n-2}} \right. \\
 &\quad \left. + \sum_{i=1}^{n-3} (L_{r+n+2-i} - tL_{r+n+1-i}) \cdot \frac{(-d)^{i-1}}{c^i} \right], \\
 y'_2 &= -t - \sum_{i=1}^{n-2} (L_{r+n+1-i} - tL_{r+n-i}) \cdot \frac{(-d)^{i-1}}{c^i}, \\
 y'_3 &= -(L_{r+3} - tL_{r+2}) \cdot \frac{1}{c}, \\
 y'_4 &= -\sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{i-1}}{c^i}, \\
 y'_k &= -\sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{k-5+i}}{c^{k-4+i}} \\
 &\quad (k = 5, 6, \dots, n).
 \end{aligned} \tag{34}$$

Specially, when $r = 0$, one gets the result of [23].

Proof. Let

$$\Omega_2 = \begin{pmatrix} 1 & -\frac{l'_n}{L_{r+1}} & \omega_{13} & \omega_{14} & \cdots & \omega_{1n} \\ 0 & 1 & \omega_{23} & \omega_{24} & \cdots & \omega_{2n} \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \tag{35}$$

where

$$\begin{aligned}
 \omega_{1j} &= \frac{1}{L_{r+1}} \left[\frac{l'_n}{l_n} (tL_{r+n+2-j} - L_{r+n+3-j}) - L_{r+n+2-j} \right] \\
 \omega_{2j} &= \frac{1}{l_n} \cdot (L_{r+n+3-j} - tL_{r+n+2-j}) \quad (j = 3, 4, \dots, n).
 \end{aligned} \tag{36}$$

Then, we have

$$\Sigma A_{r,n} \Omega_1 \Omega_2 = \begin{pmatrix} L_{r+1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & l_n & 0 & 0 & \cdots & 0 \\ 0 & 0 & c & 0 & \cdots & 0 \\ 0 & 0 & d & c & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & c \end{pmatrix}, \tag{37}$$

so $\Sigma A_{r,n} \Omega_1 \Omega_2 = \mathcal{D} \oplus \mathcal{H}$, where $D = \text{diag}(L_{r+1}, l_n)$ is a diagonal matrix and $\mathcal{D} \oplus \mathcal{H}$ is the direct sum of \mathcal{D} and \mathcal{H} . If we denote $\Omega = \Omega_1 \Omega_2$, then we obtain $A_{r,n}^{-1} = \Omega(\mathcal{D}^{-1} \oplus \mathcal{H}^{-1})\Sigma$.

Since the last row elements of the matrix Ω are $(0, 1, \omega_{23}, \omega_{24}, \dots, \omega_{2,n-1}, \omega_{2n})$, then the last row elements of the matrix $\Omega(\mathcal{D}^{-1} \oplus \mathcal{H}^{-1})$ are $(0, 1/l_n, T_{23}, T_{24}, \dots, T_{2n})$, where

$$\begin{aligned}
 T_{23} &= \sum_{i=1}^{n-2} \omega_{2,2+i} \cdot \frac{(-d)^{i-1}}{c^i}, \\
 T_{2k} &= \sum_{i=1}^{n+1-k} \omega_{2,k-1+i} \cdot \frac{(-d)^{i-1}}{c^i} \quad (k = 3, 4, \dots, n).
 \end{aligned} \tag{38}$$

Hence, it follows from Lemma 15 that letting $A_{r,n}^{-1} = \text{SCirc}(y_1, y_2, \dots, y_n)$, then its last row elements are $(-y_2, -y_3, \dots, -y_n, y_1)$ which are given by the following equations:

$$\begin{aligned}
 -y_2 &= \frac{t}{l_n} + T_{23} \\
 &= \frac{t}{l_n} + \frac{1}{l_n} \sum_{i=1}^{n-2} (L_{r+n+1-i} - tL_{r+n-i}) \cdot \frac{(-d)^{i-1}}{c^i}, \\
 -y_3 &= T_{2,n} = \frac{1}{l_n} (L_{r+3} - tL_{r+2}) \cdot \frac{1}{c}, \\
 -y_4 &= T_{2,n-1} - T_{2n} \\
 &= \frac{1}{l_n} \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{i-1}}{c^i}, \\
 -y_5 &= T_{2,n-2} - T_{2n-1} - T_{2n} \\
 &= \frac{1}{l_n} \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^i}{c^{i+1}}, \\
 -y_k &= T_{2,n-k+3} - T_{2,n-k+4} - T_{2,n-k+5} \\
 &= \frac{1}{l_n} \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{k-5+i}}{c^{k-4+i}}, \\
 &\vdots \\
 -y_n &= T_{23} - T_{24} - T_{25} \\
 &= \sum_{i=1}^{n-2} \omega_{2,2+i} \cdot \frac{(-d)^{i-1}}{c^i} - \sum_{i=1}^{n-3} \omega_{2,3+i} \cdot \frac{(-d)^{i-1}}{c^i} \\
 &\quad - \sum_{i=1}^{n-4} \omega_{2,4+i} \cdot \frac{(-d)^{i-1}}{c^i} \\
 &= \frac{1}{l_n} \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{n-5+i}}{c^{n-4+i}},
 \end{aligned}$$

$$\begin{aligned}
y_1 &= \frac{1}{l_n} - T_{23} - T_{24} \\
&= \frac{1}{l_n} - \frac{1}{l_n} \left[(L_{r+3} - tL_{r+2}) \cdot \frac{(-d)^{n-3}}{c^{n-2}} \right. \\
&\quad \left. + \sum_{i=1}^{n-3} (L_{r+n+2-i} - tL_{r+n+1-i}) \cdot \frac{(-d)^{i-1}}{c^i} \right].
\end{aligned} \tag{39}$$

Hence, we obtain

$$\begin{aligned}
y_1 &= \frac{1}{l_n} - \frac{1}{l_n} \left[(L_{r+3} - tL_{r+2}) \cdot \frac{(-d)^{n-3}}{c^{n-2}} \right. \\
&\quad \left. + \sum_{i=1}^{n-3} (L_{r+n+2-i} - tL_{r+n+1-i}) \cdot \frac{(-d)^{i-1}}{c^i} \right], \\
y_2 &= -\frac{t}{l_n} - \frac{1}{l_n} \sum_{i=1}^{n-2} (L_{r+n+1-i} - tL_{r+n-i}) \cdot \frac{(-d)^{i-1}}{c^i}, \\
y_3 &= -\frac{1}{l_n} (L_{r+3} - tL_{r+2}) \cdot \frac{1}{c}, \\
y_4 &= -\frac{1}{l_n} \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{i-1}}{c^i}, \\
y_5 &= -\frac{1}{l_n} \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^i}{c^{i+1}}, \\
y_k &= -\frac{1}{l_n} \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{k-5+i}}{c^{k-4+i}}, \\
&\vdots \\
y_n &= \frac{1}{l_n} \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{n-5+i}}{c^{n-4+i}}, \\
A_{r,n}^{-1} &= \frac{1}{l_n} \cdot \text{SCirc}(y'_1, y'_2, \dots, y'_n),
\end{aligned} \tag{40}$$

where

$$\begin{aligned}
y'_1 &= 1 - \left[(L_{r+3} - tL_{r+2}) \cdot \frac{(-d)^{n-3}}{c^{n-2}} \right. \\
&\quad \left. + \sum_{i=1}^{n-3} (L_{r+n+2-i} - tL_{r+n+1-i}) \cdot \frac{(-d)^{i-1}}{c^i} \right], \\
y'_2 &= -t - \sum_{i=1}^{n-2} (L_{r+n+1-i} - tL_{r+n-i}) \cdot \frac{(-d)^{i-1}}{c^i}, \\
y'_3 &= -(L_{r+3} - tL_{r+2}) \cdot \frac{1}{c},
\end{aligned}$$

$$\begin{aligned}
y'_4 &= -\sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{i-1}}{c^i}, \\
y'_k &= -\sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{k-5+i}}{c^{k-4+i}}, \quad (k = 5, 6, \dots, n).
\end{aligned} \tag{41}$$

This completes the proof. \square

3. Norm and Spread of Skew Circulant Matrix with the Lucas Numbers

Theorem 17. Let $A_{r,n} = \text{SCirc}(L_{r+1}, \dots, L_{r+n})$ be skew circulant matrix; then three kinds of norms of $A_{r,n}$ are given by

$$\|A_{r,n}\|_1 = \|A_{r,n}\|_{\infty} = L_{r+n+2} - L_{r+2}, \tag{42}$$

$$\|A_{r,n}\|_F = \sqrt{n(L_{r+n}L_{r+n+1} - L_rL_{r+1})}. \tag{43}$$

Proof. By Definition 8 and (12), we have

$$\|A_{r,n}\|_1 = \|A_{r,n}\|_{\infty} = \sum_{i=1}^n L_{r+i} = L_{r+n+2} - L_{r+2}. \tag{44}$$

By Definition 8 and (13), we have

$$\begin{aligned}
(\|A_{r,n}\|_F)^2 &= \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \\
&= n \sum_{i=1}^n L_{r+i}^2 \\
&= n \left(\sum_{i=0}^{r+n} L_i^2 - \sum_{i=0}^r L_i^2 \right) \\
&= n(L_{r+n}L_{r+n+1} - L_rL_{r+1}).
\end{aligned} \tag{45}$$

Thus

$$\|A_{r,n}\|_F = \sqrt{n(L_{r+n}L_{r+n+1} - L_rL_{r+1})}. \tag{46} \quad \square$$

Theorem 18. Let

$$A'_{r,n} = \text{SCirc}(L_{r+1}, -L_{r+2}, \dots, -L_{r+n-1}, L_{r+n}) \tag{47}$$

be an odd-order alternative skew circulant matrix and let n be odd. Then

$$\|A'_{r,n}\|_2 = \sum_{i=1}^n L_{r+i} = L_{r+n+2} - L_{r+2}. \tag{48}$$

Proof. By Lemma 3, we have

$$\lambda_j(A'_{r,n}) = \sum_{i=1}^n (-1)^{i-1} L_{r+i} (\omega^j \eta)^{i-1}. \tag{49}$$

So

$$\begin{aligned} |\lambda_j(A'_{r,n})| &\leq \sum_{i=1}^n |(-1)^{i-1} L_{r+i}| \cdot |(\omega^j \eta)^{i-1}| \\ &= \sum_{i=1}^n L_{r+i}, \end{aligned} \quad (50)$$

for all $j = 0, 1, \dots, n-1$.

Since n is odd, $\sum_{i=1}^n L_{r+i}$ is an eigenvalue of $A'_{r,n}$; that is,

$$\begin{aligned} &\begin{pmatrix} L_{r+1} & -L_{r+2} & \vdots & L_{r+n} \\ -L_{r+n} & L_{r+1} & & -L_{r+n-1} \\ L_{r+n-1} & -L_{r+n} & & L_{r+n-2} \\ \vdots & \vdots & \vdots & \vdots \\ L_{r+2} & -L_{r+3} & & L_{r+1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \sum_{i=1}^n L_{r+i} \cdot \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \\ 1 \end{pmatrix}. \end{aligned} \quad (51)$$

To sum up, we have

$$\max_{0 \leq j \leq n-1} |\lambda_j(A'_{r,n})| = \sum_{i=1}^n L_{r+i}. \quad (52)$$

Since all skew circulant matrices are normal, by Lemma 9 and (12), and (52), we have

$$\|A'_{r,n}\|_2 = \sum_{i=1}^n L_{r+i} = L_{r+n+2} - L_{r+2}, \quad (53)$$

which completes the proof. \square

Theorem 19. Let $A_{r,n} = \text{SCirc}(L_{r+1}, \dots, L_{r+n})$ be skew circulant matrix; then the bounds for the spread of $A_{r,n}$ are

$$\begin{aligned} s(A_{r,n}) &\leq \sqrt{2n(L_{r+n}L_{r+n+1} - L_{r+1}L_{r+2})}, \\ s(A_{r,n}) &\geq \frac{1}{n-1} |2L_{r+n+3} - (n-2)L_{r+n+2} - nL_{r+3} - 2L_{r+4}|. \end{aligned} \quad (54)$$

Proof. The trace of $A_{r,n}$, $\text{tr } A_{r,n} = nL_{r+1}$. By (18) and (43), we have

$$s(A_{r,n}) \leq \sqrt{2n(L_{r+n}L_{r+n+1} - L_{r+1}L_{r+2})}. \quad (55)$$

Since

$$\begin{aligned} \sum_{i \neq j} a_{ij} &= \sum_{k=2}^n (n - (k-1)) L_{r+k} - \sum_{k=2}^n (k-1) L_{r+k} \\ &= (n+2) \sum_{k=2}^n L_{r+k} - 2 \sum_{k=2}^n k L_{r+k} \end{aligned}$$

$$\begin{aligned} &= (n+2)(L_{r+n+2} - L_{r+3}) \\ &\quad - 2 \left[\sum_{k=2}^n (r+k) L_{r+k} - \sum_{k=2}^n r L_{r+k} \right], \end{aligned} \quad (56)$$

by (12) and (14),

$$\sum_{i \neq j} a_{ij} = 2L_{r+n+3} - (n-2)L_{r+n+2} - nL_{r+3} - 2L_{r+4}. \quad (57)$$

By (19), we have

$$\begin{aligned} s(A_{r,n}) &\geq \frac{1}{n-1} |2L_{r+n+3} - (n-2)L_{r+n+2} \\ &\quad - nL_{r+3} - 2L_{r+4}|. \end{aligned} \quad (58)$$

\square

4. Determinant and Inverse of Skew Left Circulant Matrix with the Lucas Numbers

In this section, let $A''_{r,n} = \text{SLCirc}(L_{r+1}, \dots, L_{r+n})$ be skew left circulant matrix. By using the obtained conclusions in Section 2, we give a determinant explicit formula for the matrix $A''_{r,n}$. Afterwards, we prove that $A''_{r,n}$ is an invertible matrix for any positive interger n . The inverse of the matrix $A''_{r,n}$ is also presented.

According to Lemmas 5 and 6 and Theorems 13, 14, and 16, we can obtain the following theorems.

Theorem 20. Let $A''_{r,n} = \text{SLCirc}(L_{r+1}, \dots, L_{r+n})$ be skew left circulant matrix; then

$$\begin{aligned} \det A''_{r,n} &= (-1)^{n(n-1)/2} L_{r+1} \\ &\quad \times \left[L_{r+1} + tL_{r+n} + \sum_{k=1}^{n-2} (tL_{r+1+i} - L_{r+2+i}) x^{n-1-i} \right] \\ &\quad \cdot c^{n-2}, \end{aligned} \quad (59)$$

where L_{r+n} is the $(r+n)$ th Lucas number.

Theorem 21. Let $A''_{r,n} = \text{SLCirc}(L_{r+1}, \dots, L_{r+n})$ be skew left circulant matrix; then $A''_{r,n}$ is an invertible matrix.

Theorem 22. Let $A''_{r,n} = \text{SLCirc}(L_{r+1}, \dots, L_{r+n})$ be skew left circulant matrix; then

$$(A''_{r,n})^{-1} = \frac{1}{l_n} \text{SLCirc}(y''_1, y''_2, \dots, y''_n), \quad (60)$$

where

$$\begin{aligned} y''_1 &= 1 - \left[(L_{r+3} - tL_{r+2}) \frac{(-d)^{n-3}}{c^{n-2}} \right. \\ &\quad \left. + \sum_{i=1}^{n-3} (L_{r+n+2-i} - tL_{r+n+1-i}) \cdot \frac{(-d)^{i-1}}{c^i} \right], \end{aligned}$$

$$\begin{aligned}
y_k'' &= -y_{n-k+2}' \\
&= \sum_{i=1}^2 (L_{r+1+i} - tL_{r+i}) \cdot \frac{(-d)^{n-k-3+i}}{c^{n-k-2+i}}, \\
&\quad (k = 2, 3, \dots, n-2). \\
y_{n-1}'' &= -y_3' = (L_{r+3} - tL_{r+2}) \cdot \frac{1}{c}, \\
y_n'' &= -y_2' \\
&= t + \sum_{i=1}^{n-2} (L_{r+n+1-i} - tL_{r+n-i}) \cdot \frac{(-d)^{i-1}}{c^i}.
\end{aligned} \tag{61}$$

5. Norm and Spread of Skew Left Circulant Matrix with the Lucas Numbers

Theorem 23. Let $A_{r,n}'' = \text{SLCirc}(L_{r+1}, \dots, L_{r+n})$ be skew left circulant matrix. Then three kinds of norms of $A_{r,n}''$ are given by

$$\begin{aligned}
\|A_{r,n}''\|_1 &= \|A_{r,n}\|_\infty = L_{r+n+2} - L_{r+2}, \\
\|A_{r,n}''\|_F &= \sqrt{n(L_{r+n}L_{r+n+1} - L_rL_{r+1})}.
\end{aligned} \tag{62}$$

Proof. Using the method in Theorem 17 similarly, the conclusion is obtained. \square

Theorem 24. Let

$$A_{r,n}''' = \text{SLCirc}(L_{r+1}, -L_{r+2}, \dots, -L_{r+n-1}, L_{r+n}) \tag{63}$$

be an odd-order alternative skew left circulant matrix; then

$$\|A_{r,n}'''\|_2 = \sum_{i=1}^n L_{r+i} = L_{r+n+2} - L_{r+2}. \tag{64}$$

Proof. According to Lemma 4,

$$\lambda_j(A_{r,n}''') = \pm \left| \sum_{i=1}^n (-1)^{i-1} L_{r+i} \omega^{(j-(1/2))(k-1)} \right|, \tag{65}$$

for $j = 1, 2, \dots, (n-1)/2$, and

$$\lambda_{(n+1)/2}(A_{r,n}''') = \sum_{i=1}^n L_{r+i}. \tag{66}$$

So

$$\begin{aligned}
|\lambda_j(A_{r,n}''')| &\leq \sum_{i=1}^n |(-1)^{i-1} L_{r+i} (-1)^{i-1}| \\
&= \sum_{i=1}^n L_{r+i}, \quad \left(j = 1, 2, \dots, \frac{n+1}{2}\right).
\end{aligned} \tag{67}$$

By (66) and (67), we have

$$\max_{0 \leq i \leq (n+1)/2} |\lambda_i(A_{r,n}''')| = \sum_{i=1}^n L_{r+i}. \tag{68}$$

Since all skew left circulant matrices are symmetrical, by Lemma 9 and (12) and (68), we obtain

$$\|A_{r,n}'''\|_2 = L_{r+n+2} - L_{r+2}. \tag{69}$$

\square

Theorem 25. Let $A_{r,n}'' = \text{SLCirc}(L_{r+1}, \dots, L_{r+n})$ be skew left circulant matrix; the bounds for the spread of $A_{r,n}''$ are

$$2L_{r+n} \leq s(A_{r,n}'') \leq \begin{cases} \sqrt{M - \frac{2}{n}N^2}, & \text{if } n \text{ is odd,} \\ \sqrt{M}, & \text{if } n \text{ is even,} \end{cases} \tag{70}$$

where

$$\begin{aligned}
M &= 2n(L_{r+n}L_{r+n+1} - L_{r+1}L_r), \\
N &= L_{r+n-1} + L_{r-1}.
\end{aligned} \tag{71}$$

Proof. Since $A_{r,n}''$ is a symmetric matrix, by (20),

$$s(A_{r,n}'') \geq 2 \max_{i \neq j} |a_{ij}| = 2L_{r+n}. \tag{72}$$

The trace of $A_{r,n}''$ is, if n is odd,

$$\begin{aligned}
\text{tr}(A_{r,n}'') &= L_{r+1} - L_{r+2} + L_{r+3} - \dots + L_{r+n} \\
&= L_{r+1} + L_{r+1} + L_{r+3} + \dots + L_{r+n-2} \\
&= 2L_{r+1} + L_{r+1} + L_{r+2} + \dots + L_{r+n-3} \\
&= 2L_{r+1} + \sum_{i=1}^{n-3} L_{r+i}.
\end{aligned} \tag{73}$$

By (12), we have

$$\text{tr}(A_{r,n}'') = L_{r+n-1} + L_{r-1} = N. \tag{74}$$

Let $M = 2n(L_{r+n}L_{r+n+1} - L_{r+1}L_r)$; then, by (18), (62), and (74), we obtain

$$s(A_{r,n}'') \leq \sqrt{M - \frac{2}{n}N^2}. \tag{75}$$

If n is even, then

$$\begin{aligned}
\text{tr}(A_{r,n}'') &= L_{r+1} - L_{r+1} + L_{r+3} \\
&\quad - L_{r+3} \dots - L_{r+n-1} = 0.
\end{aligned} \tag{76}$$

By (18), (62), and (76), we have

$$s(A_{r,n}'') \leq \sqrt{M}. \tag{77}$$

So the result follows. \square

6. Conclusion

We discuss the invertibility of the skew circulant type matrices with any continuous Lucas numbers and present the determinant and the inverse matrices by constructing the transformation matrices. The four kinds of norms and bounds for the spread of these matrices are given, respectively. In [3], a new fast algorithm for optimal design of block digital filters (BDFs) is proposed based on skew circulant matrix. The reason why we focus our attention on skew circulant is to explore the application of skew circulant in the related field in medicine image, image encryption, and real-time tracking. On the basis of existing application situation [4], we conjecture that SVD decomposition of skew circulant matrix will play an important role in CT-perfusion imaging of human brain. On the basis method of [8] and ideas of [5], we will exploit real-time tracking with kernel matrix of skew circulant structure. A novel chaotic image encryption scheme based on the time-delay Lorenz system is presented in [6] with the description of circulant matrix. We will exploit chaotic image encryption algorithm based on skew circulant operation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the Natural Science Foundation of Shandong Province (Grant no. ZR2011FL017), the National Nature Science Foundation of China (Grant no. F020701), and the AMEP of Linyi University, China.

References

- [1] A. Daher, E. H. Baghious, and G. Burel, "Fast algorithm for optimal design of block digital filters based on circulant matrices," *IEEE Signal Processing Letters*, vol. 15, pp. 637–640, 2008.
- [2] V. C. Liu and P. P. Vaidyanathan, "Circulant and skew-circulant matrices as new normal-form realization of IIR digital filters," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 6, pp. 625–635, 1988.
- [3] D. Q. Fu, Z. L. Jiang, Y. F. Cui, and S. T. Jhang, "A new fast algorithm for optimal design of block digital filters by skew-cyclic convolution," *IET Signal Processing*, p. 6, 2014.
- [4] H.-J. Wittsack, A. M. Wohlschläger, E. K. Ritzl et al., "CT-perfusion imaging of the human brain: advanced deconvolution analysis using circulant singular value decomposition," *Computerized Medical Imaging and Graphics*, vol. 32, no. 1, pp. 67–77, 2008.
- [5] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Computer Vision—ECCV 2012*, vol. 7575 of *Lecture Notes in Computer Science*, pp. 702–715, Springer, Berlin, Germany, 2012.
- [6] X. Huang, G. Ye, and K.-W. Wong, "Chaotic image encryption algorithm based on circulant operation," *Abstract and Applied Analysis*, vol. 2013, Article ID 384067, 8 pages, 2013.
- [7] Y. Jing and H. Jafarkhani, "Distributed differential space-time coding for wireless relay networks," *IEEE Transactions on Communications*, vol. 56, no. 7, pp. 1092–1100, 2008.
- [8] M. J. Narasimha, "Linear convolution using skew-cyclic convolutions," *IEEE Signal Processing Letters*, vol. 14, no. 3, pp. 173–176, 2007.
- [9] T. A. Gulliver and M. Harada, "New nonbinary self-dual codes," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 415–417, 2008.
- [10] P. J. Davis, *Circulant Matrices*, John Wiley & Sons, New York, NY, USA, 1979.
- [11] Z. L. Jiang and Z. X. Zhou, *Circulant Matrices*, Chengdu Technology University Publishing, Chengdu, China, 1999.
- [12] D. Bertaccini and M. K. Ng, "Skew-circulant preconditioners for systems of LMF-based ODE codes," in *Numerical Analysis and Its Applications*, vol. 1988 of *Lecture Notes in Computer Science*, pp. 93–101, Springer, Berlin, Germany, 2001.
- [13] R. H. Chan and X.-Q. Jin, "Circulant and skew-circulant preconditioners for skew-Hermitian type Toeplitz systems," *BIT Numerical Mathematics*, vol. 31, no. 4, pp. 632–646, 1991.
- [14] R. H. Chan and K.-P. Ng, "Toeplitz preconditioners for Hermitian Toeplitz systems," *Linear Algebra and Its Applications*, vol. 190, pp. 181–208, 1993.
- [15] T. Huckle, "Circulant and skew circulant matrices for solving Toeplitz matrix problems," *SIAM Journal on Matrix Analysis and Applications*, vol. 13, no. 3, pp. 767–777, 1992.
- [16] J. N. Lyness and T. Sørensen, "Four-dimensional lattice rules generated by skew-circulant matrices," *Mathematics of Computation*, vol. 73, no. 245, pp. 279–295, 2004.
- [17] H. Karner, J. Schneid, and C. W. Ueberhuber, "Spectral decomposition of real circulant matrices," *Linear Algebra and Its Applications*, vol. 367, pp. 301–311, 2003.
- [18] J. Li, Z. L. Jiang, N. Shen, and J. W. Zhou, "On optimal backward perturbation analysis for the linear system with skew circulant coefficient matrix," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 707381, 7 pages, 2013.
- [19] D. V. Jaiswal, "On determinants involving generalized Fibonacci numbers," *The Fibonacci Quarterly*, vol. 7, pp. 319–330, 1969.
- [20] D. A. Lind, "A Fibonacci circulant," *The Fibonacci Quarterly*, vol. 8, no. 5, pp. 449–455, 1970.
- [21] L. Dazheng, "Fibonacci-Lucas quasi-cyclic matrices," *The Fibonacci Quarterly*, vol. 40, no. 3, pp. 280–286, 2002.
- [22] S.-Q. Shen, J.-M. Cen, and Y. Hao, "On the determinants and inverses of circulant matrices with Fibonacci and Lucas numbers," *Applied Mathematics and Computation*, vol. 217, no. 23, pp. 9790–9797, 2011.
- [23] Y. Gao, Z. L. Jiang, and Y. P. Gong, "On the determinants and inverses of skew circulant and skew left circulant matrices with Fibonacci and Lucas numbers," *WSEAS Transactions on Mathematics*, vol. 12, no. 4, pp. 472–481, 2013.
- [24] X. Y. Jiang, Y. Gao, and Z. L. Jiang, "Determinants and inverses of skew and skew left circulant matrices involving the k -Fibonacci numbers in communications-I," *Far East Journal of Mathematical Sciences*, vol. 76, no. 1, pp. 123–137, 2013.
- [25] X. Y. Jiang, Y. Gao, and Z. L. Jiang, "Determinants and inverses of skew and skew left circulant matrices involving the k -Lucas numbers in communications-II," *Far East Journal of Mathematical Sciences*, vol. 78, no. 1, pp. 1–17, 2013.
- [26] S. Solak, "On the norms of circulant matrices with the Fibonacci and Lucas numbers," *Applied Mathematics and Computation*, vol. 160, no. 1, pp. 125–132, 2005.

- [27] A. İpek, "On the spectral norms of circulant matrices with classical Fibonacci and Lucas numbers entries," *Applied Mathematics and Computation*, vol. 217, no. 12, pp. 6011–6012, 2011.
- [28] S. Shen and J. Cen, "On the bounds for the norms of r -circulant matrices with the Fibonacci and Lucas numbers," *Applied Mathematics and Computation*, vol. 216, no. 10, pp. 2891–2897, 2010.
- [29] M. Akbulak and D. Bozkurt, "On the norms of Toeplitz matrices involving Fibonacci and Lucas numbers," *Hacettepe Journal of Mathematics and Statistics*, vol. 37, no. 2, pp. 89–95, 2008.
- [30] A. Bose, R. S. Hazra, and K. Saha, "Spectral norm of circulant-type matrices," *Journal of Theoretical Probability*, vol. 24, no. 2, pp. 479–516, 2011.
- [31] L. Mirsky, "The spread of a matrix," *Mathematika*, vol. 3, pp. 127–130, 1956.
- [32] R. Sharma and R. Kumar, "Remark on upper bounds for the spread of a matrix," *Linear Algebra and Its Applications*, vol. 438, no. 11, pp. 4359–4362, 2013.
- [33] L. Mirsky, "Inequalities for normal and Hermitian matrices," *Duke Mathematical Journal*, vol. 24, no. 4, pp. 591–599, 1957.
- [34] E. R. Barnes and A. J. Hoffman, "Bounds for the spectrum of normal matrices," *Linear Algebra and Its Applications*, vol. 201, pp. 79–90, 1994.
- [35] E. Jiang and X. Zhan, "Lower bounds for the spread of a Hermitian matrix," *Linear Algebra and Its Applications*, vol. 256, pp. 153–163, 1997.
- [36] R. Bhatia and R. Sharma, "Some inequalities for positive linear maps," *Linear Algebra and Its Applications*, vol. 436, no. 6, pp. 1562–1571, 2012.
- [37] J. Wu, P. Zhang, and W. Liao, "Upper bounds for the spread of a matrix," *Linear Algebra and Its Applications*, vol. 437, no. 11, pp. 2813–2822, 2012.
- [38] C. R. Johnson, R. Kumar, and H. Wolkowicz, "Lower bounds for the spread of a matrix," *Linear Algebra and Its Applications*, vol. 71, pp. 161–173, 1985.

Research Article

Application of EMD-Based SVD and SVM to Coal-Gangue Interface Detection

Wei Liu, Kai He, Qun Gao, and Cheng-yin Liu

School of Information and Electronics Engineering, Shandong Institute of Business and Technology, Yantai 264005, China

Correspondence should be addressed to Wei Liu; liuweiyyjn@126.com

Received 1 November 2013; Accepted 21 March 2014; Published 14 April 2014

Academic Editor: Feng Gao

Copyright © 2014 Wei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Coal-gangue interface detection during top-coal caving mining is a challenging problem. This paper proposes a new vibration signal analysis approach to detecting the coal-gangue interface based on singular value decomposition (SVD) techniques and support vector machines (SVMs). Due to the nonstationary characteristics in vibration signals of the tail boom support of the longwall mining machine in this complicated environment, the empirical mode decomposition (EMD) is used to decompose the raw vibration signals into a number of intrinsic mode functions (IMFs) by which the initial feature vector matrices can be formed automatically. By applying the SVD algorithm to the initial feature vector matrices, the singular values of matrices can be obtained and used as the input feature vectors of SVMs classifier. The analysis results of vibration signals from the tail boom support of a longwall mining machine show that the method based on EMD, SVD, and SVM is effective for coal-gangue interface detection even when the number of samples is small.

1. Introduction

Today a major problem facing the mining industry is how to develop an automated top-coal caving system that can maximize the ratio of coal to gangue. The working procedure of top-coal caving is automatically controlled by an electrohydraulic system, which determines the recovery ratio of top-coal to gangue. In order to improve the recovery ratio of top-coal, a lot of work has been done on coal-gangue interface detection (CID) [1–3]. This paper proposes a new CID method based on the analysis of vibration signals due to the collapse of coal and gangue onto the tail boom of a longwall mining machine. Some significant features that differ between coal and gangue can be extracted by analyzing these vibration signals. The acquired vibration signals are usually nonlinear and nonstationary, so it is difficult to effectively extract features. Recently, the time-frequency analysis methods are widely used in the feature extraction of vibration signals [4, 5]. Among all available time-frequency analysis methods, the wavelet transform may be the best one [6, 7]. However, wavelet transform is not a self-adaptive signal processing method. Also, energy leakage will occur when wavelet transform is used to process signals, due

to the fact that it is an adjustable windowed Fourier transform in nature [8]. The empirical mode decomposition (EMD) decomposes any time-varying signal into its fundamental intrinsic oscillatory modes [9]. The EMD is a self-adaptive time-frequency analysis method that is perfectly applicable to nonlinear and nonstationary processing [10, 11].

Recently, singular value decomposition (SVD) of matrix has been widely applied to signal processing, statistical analysis, automatic control, and so forth [12]. According to the matrix theory, singular values have good stability and represent the inherent characteristics of the matrix. That is, when a slight change of matrix elements occurs, the change of matrix singular values is small.

In practice, a large number of samples are usually not available. Support vector machine (SVM) is a new machine learning method developed on the basis of statistical learning theory [13]. SVM can solve the learning problem of a smaller number of samples. Meanwhile, SVM has better generalization than artificial neural network (ANN) and guarantees that the local and global optimal solutions are exactly the same.

In this paper, the SVD technique based on EMD is applied to the feature extraction of vibration signals from

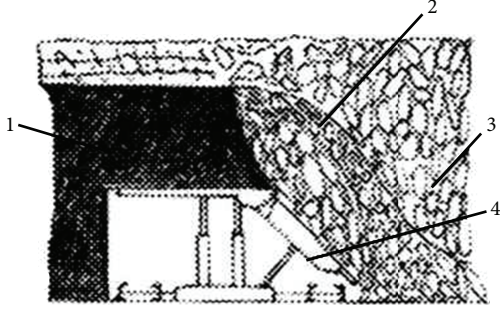


FIGURE 1: Hydraulic support and installation position of sensors (1: coal, 2: coal-gangue, 3: gangue, and 4: sensor).

coal and gangue collapse during top-coal caving. The SVM is introduced into the CID due to its high accuracy and good generalization for a smaller sample number.

This paper is organized as follows: in the next section, the feature extraction algorithm based on SVD and EMD is discussed. Section 3 briefly reviews the SVM theory. In Section 4, the basic principle of the vibration-based CID experimental system is introduced. The application of the SVM to classification of the caving states is then discussed and the results from a CID validation study are reported. The conclusions are given in the last section.

2. The SVD Technique Based on EMD

The EMD is a nonlinear and nonstationary signal analysis method proposed by Huang et al. EMD can decompose any time-varying signal into its fundamental intrinsic mode functions (IMFs), which must satisfy two conditions [9] as follows.

- (1) In the whole data set of each intrinsic mode function component, the number of extreme values and the number of zero-crossings must be equal to or differ at most by one.
- (2) At any point, the mean value of the envelope defined by local maxima and that defined by the local minima is zero.

With the definition, any time series signal $X(t)$ can be expressed as the sum of the IMF components and the residue

$$X(t) = \sum_{i=1}^n C_i(t) + R_n(t), \quad (1)$$

where $C_i(t)$ are the IMF components. Here $R_n(t)$ is the residue. The IMF includes different frequencies ranging from high to low. Acting as an adaptive data-driven filter bank, the EMD extracts the signal features of disturbances dynamically according to their different physical characteristics.

The IMFs (C_1, C_2, \dots, C_m) are chosen to construct the initial feature vector matrix A as follows:

$$A = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_m \end{pmatrix} \in C_r^{m \times n}, \quad m \leq n. \quad (2)$$

Due to the orthogonality of the EMD method, all IMFs are pairwise orthogonal. Therefore, the matrix A must be full rank. By applying the SVD to matrix A , then there exists

$$A = USV, \quad (3)$$

where $U \in R^{m \times m}$, $UU' = I$; $V \in R^{n \times n}$, $VV' = I$; $S = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. The values of λ_i are the singular values of matrix A .

As singular values can reflect the nature characteristic of the matrix, the characteristic of vibration signals of coal and gangue can be described effectively by singular values of the initial feature vector matrix. Thus, the singular values of matrix could be used as feature vectors. The SVM could be chosen as the pattern classifier to classify the caving states after the vibration feature vector has been extracted.

3. Support Vector Machine (SVM)

As a new generation learning system, SVM enables the nonlinear mapping of an n -dimensional input space into a high dimensional feature space. SVM uses a nonlinear mapping to transform an input space to a high-dimension space based on a kernel function and then looks for a nonlinear relation between inputs and outputs in the high-dimension space.

Suppose that there is a given training sample set $G = \{(x_i, y_i), i = 1, \dots, n\}$, with each sample $x_i \in R^d$, $y_i \in \{+1, -1\}$. The classification boundary can be described as follows:

$$w' \cdot x + b = 0, \quad (4)$$

where w is a weight vector and b is a bias. Therefore, the following decision function can be used to classify any data set in two classes:

$$f(x) = \text{sgn}(w' \cdot x + b). \quad (5)$$

In order to correctly classify two-class samples, the optimal hyperplane separating the samples can be obtained as a solution to the following constrained optimization problem:

minimize

$$\varphi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} w^T w \quad (6)$$

subject to

$$y_i [w^T x_i + b] - 1 \geq 0, \quad i = 1, \dots, n. \quad (7)$$

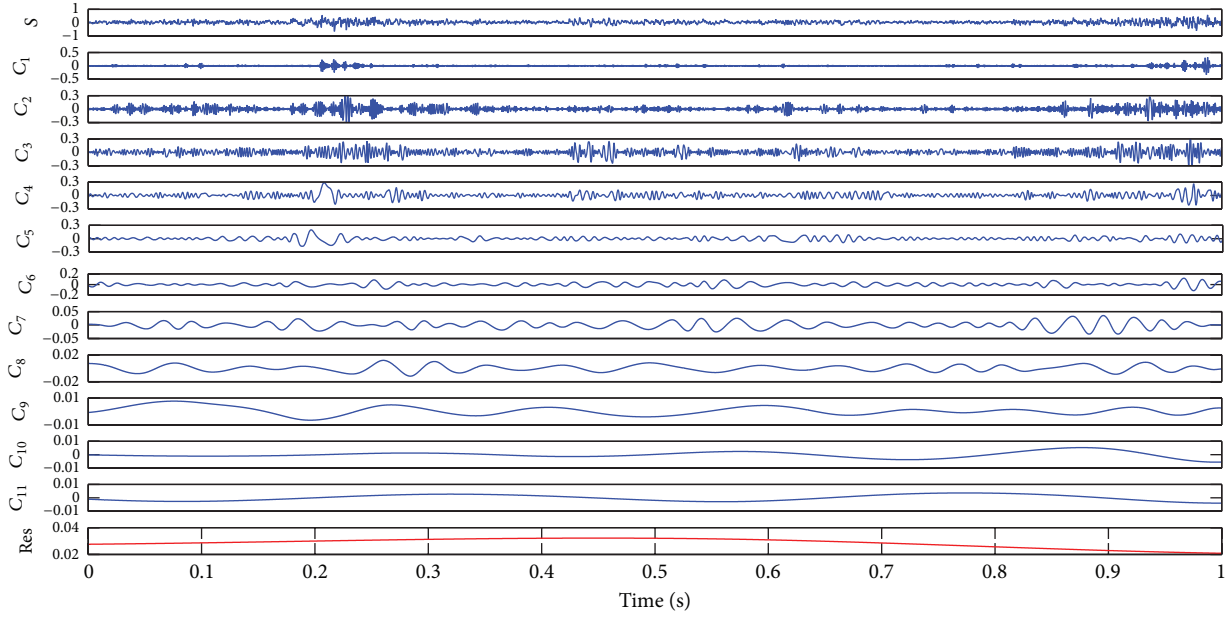


FIGURE 2: EMD results for top-coal caving.

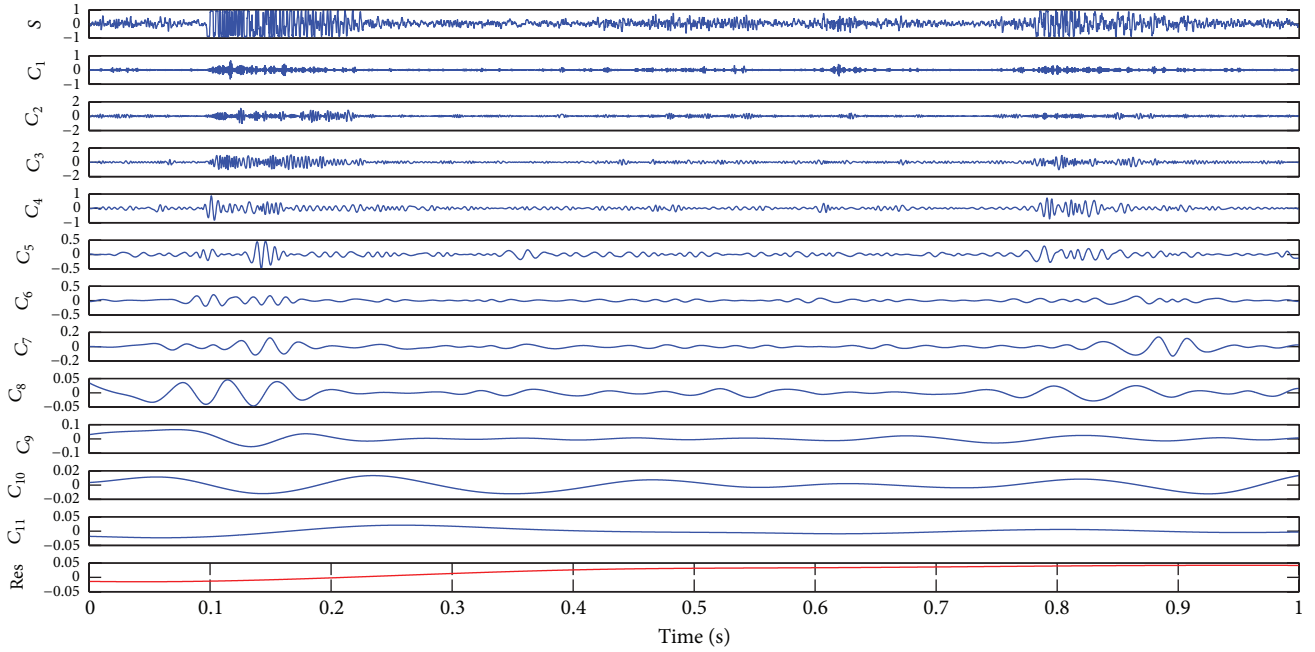


FIGURE 3: EMD results for coal-gangue.

Defining Lagrange multipliers $\alpha_i \geq 0$, the optimization problem can be converted to the following:

subject to

$$\alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0. \quad (9)$$

maximize

So the decision function can be expressed as follows:

$$P(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j (x_i \cdot x_j) \quad (8)$$

$$f(x) = \text{sgn} \left(\sum_{i=1}^n a_i^* y_i (x_i \cdot x) + b^* \right). \quad (10)$$

TABLE 1: Comparison of singular values of selected IMFs for each caving state.

| | λ_1 | λ_2 | λ_3 | λ_4 | λ_5 | λ_6 | λ_7 |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Top-coal caving | 5.8913 | 5.0226 | 4.7560 | 3.5182 | 2.7658 | 2.6858 | 1.0905 |
| Coal-gangue caving | 19.3695 | 13.7673 | 12.2830 | 6.8331 | 6.5870 | 4.2883 | 3.0371 |
| IMF component | C_3 | C_2 | C_4 | C_5 | C_1 | C_6 | C_7 |

TABLE 2: Samples for singular values of IMFs.

| Number | λ_1 | λ_2 | λ_3 | λ_4 | λ_5 | λ_6 | λ_7 | Expected output |
|--------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-----------------|
| 1 | 6.8036 | 5.1018 | 3.8767 | 2.8366 | 2.1208 | 1.8678 | 1.0817 | (1, 0) |
| 2 | 6.1150 | 4.8672 | 4.5894 | 3.3310 | 2.8880 | 2.3330 | 1.2493 | (1, 0) |
| 3 | 6.2398 | 5.8762 | 4.1749 | 3.0470 | 2.4142 | 1.9705 | 0.8274 | (1, 0) |
| 4 | 5.7488 | 5.2618 | 3.9826 | 3.1724 | 2.6175 | 2.1648 | 1.4194 | (1, 0) |
| 5 | 6.9123 | 5.8563 | 4.6258 | 3.2843 | 2.2754 | 2.1075 | 0.9596 | (1, 0) |
| 6 | 6.3855 | 5.2633 | 3.6245 | 3.2700 | 2.4509 | 2.1743 | 0.8024 | (1, 0) |
| 7 | 8.9377 | 8.6310 | 4.5831 | 4.5361 | 3.1545 | 1.8105 | 0.9570 | (0, 1) |
| 8 | 15.2788 | 11.2140 | 7.2233 | 5.1752 | 3.4570 | 1.7460 | 1.0975 | (0, 1) |
| 9 | 19.2038 | 13.6730 | 12.1804 | 6.7688 | 6.5208 | 4.1607 | 2.5960 | (0, 1) |
| 10 | 14.5418 | 12.4612 | 8.7011 | 6.1077 | 5.0121 | 3.2906 | 1.6954 | (0, 1) |
| 11 | 22.5342 | 14.4046 | 9.4929 | 5.9841 | 5.3480 | 3.6622 | 2.2097 | (0, 1) |
| 12 | 12.0520 | 10.0045 | 9.6424 | 4.9427 | 4.4479 | 3.5481 | 1.3674 | (0, 1) |

TABLE 3: Classification results of SVM.

| Caving states | Test samples | Error | Classification accuracy |
|--------------------|--------------|-------|-------------------------|
| Top-coal caving | 18 | 0 | 100% |
| Coal-gangue caving | 18 | 0 | 100% |

4. Applications

In order to investigate the EMD-based SVD technique and SVM as a means of distinguishing between top-coal and gangue caving impacts on the tail boom of a mining machine, an experiment has been carried out on number 2303 working face, Zhangcun Mine, Shanxi, China. The CID experimental system is composed of a data acquisition device, an embedded signal analysis platform, and vibration acceleration sensors, as shown in Figure 1. When coal and gangue fall down and shock the tail boom, the acceleration sensors, which are installed on the hydraulic support, acquire vibration signals from the steel plate [14].

4.1. Feature Extraction of Coal-Gangue Vibration Signals. As an example, two different vibration signals of top-coal caving and coal-gangue caving are chosen for further analysis. The sampling frequency of these signals is 8000 Hz and the sampling time is one second. Firstly, the EMD is applied to the analysis of separate vibration signals of pure coal and coal-gangue. As shown in Figures 2 and 3, the EMD decomposes the two original signals S into eleven IMFs, which contains important information correlated with the vibration signals of coal and gangue impacts. Obviously, IMFs C_1, C_2, C_3 , and C_4 have much higher frequencies than other IMFs. IMFs from C_8 to C_{11} oscillate so slowly that they only contain very low frequencies, which are composed of clutter

or noise, for example, vibration signals caused by mechanical devices themselves. As shown in Figure 3, there are two shock characteristics at the time from 0.1 s to 0.2 s and from 0.8 s to 0.9 s, which correspond with gangue falling events. Meanwhile, the frequency and the amplitude of the first seven IMFs are greater than those of other IMFs during coal-gangue caving. Therefore, the first seven IMFs are selected for further study, and the other IMFs are the residue accordingly.

For each vibration signal of each caving state, the initial feature vector matrix A can be constructed according to (2). Then SVD is applied to the initial feature vector matrix A ; namely, $A = USV$, where $A \in R^{7 \times 8000}$, $U \in R^{7 \times 7}$, $S \in R^{7 \times 8000}$, $V \in R^{8000 \times 8000}$. The singular values λ_i can be obtained, which are shown in Table 1

$$\lambda_i = [\lambda_1, \lambda_2, \dots, \lambda_7]. \quad (11)$$

It can be seen from Table 1 that the singular values of matrix are arranged in descending order of significance. The correspondence between the singular values and IMFs is also given. Especially for the coal-gangue caving, the singular values λ_1, λ_2 , and λ_3 are much bigger than the others, so the singular values could be regarded as the feature vectors to be input to the SVM classifier.

4.2. CID Based on SVM. Actually, the CID is to distinguish two caving states, that is, to solve a two-class pattern classification problem. SVM has the advantage of solving a two-class problem on the basis of searching for structural risk minimization, even in the case of few learning samples [15]. The new CID method based on SVD, EMD, and SVM is given as follows.

Step 1. Acquire N signals at the sample frequency f_s under the condition of top-coal caving and coal-gangue caving,

TABLE 4: Classification results of SVM under few samples.

| Testing sample number (from Table 2) | Real caving state | Distance to optimal hyperplane H | | Results |
|---|--------------------|------------------------------------|--------------------|------------|
| | | 126 training samples | 8 training samples | |
| 5 | Top-coal caving | 0.9473 | 0.6561 | +1 (right) |
| 6 | Top-coal caving | 1.0137 | 0.9807 | +1 (right) |
| 11 | Coal-gangue caving | -1.0223 | -0.4178 | -1 (right) |
| 12 | Coal-gangue caving | -1.0004 | -0.4182 | -1 (right) |

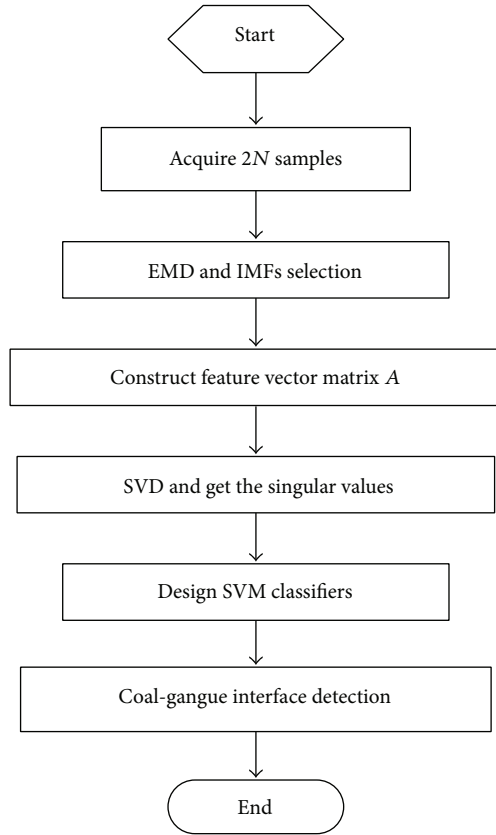


FIGURE 4: The flowchart of EMD-based SVD and SVM.

respectively. The $2N$ samples are divided into two subsets, the training samples sets and the testing samples set.

Step 2. Each signal is decomposed by EMD. Choose the first seven IMFs and construct feature vector matrix A . The singular values can be got by applying SVD to matrix A .

Step 3. Design SVM classifiers. When the feature input vector is a sample with known state of top-coal caving, the output of SVM classifier is set to 1, otherwise to -1. The singular values of the training samples are used as the input to train the SVM classifier. Then the state of caving can be distinguished after the testing samples have been input into the trained SVM classifier.

The flowchart of the proposed method is presented in Figure 4.

4.3. Validation Study. The caving state detection method based on SVD, EMD, and SVM is applied to a vibration sample set of both pure coal and coal-gangue caving. At first, a total of 126 vibration signals are acquired with a sample frequency of 8000 Hz, 63 signals for each caving state. In addition, the testing data sets consisting of 18 signals for each caving are used for validation of this detection method. Then the singular values of each signal are obtained after applying SVD based on EMD, parts of which are listed in Table 2.

Choose RBF kernel function and set $1/\sigma^2 = 0.5$, $C = 10$. It takes about 0.003 s to establish the SVM classification model. The number of support vectors is twelve, which accounts for 9.52% of the total of the training samples. The classification results are shown in Table 3. Obviously, the results are totally consistent with the real caving state.

In order for further study of the classification performance of SVM in the case of a small sample, the number of training samples decreases to eight (number 1 to number 4 and number 7 to number 10 from Table 2 and the rest as testing samples). The classification results are shown in Table 4. Table 4 shows that the SVM classifier can classify the two caving states accurately even in the case of decreasing the training samples. By comparing the distances between testing samples and the optimal hyperplane H , it is found that the overall performance of the SVM classifier weakens as the samples reduce.

5. Conclusions

The problem of coal-gangue interface detection (CID) on a fully mechanized mining face has been addressed by applying the SVD technique and EMD to extracting longwall mining machine tail boom support vibration features that can be used for top-coal and coal-gangue caving state classification. EMD is a self-adaptive analysis method that can decompose the signal into a number of IMFs. These functions provide a compact natural representation of nonstationary, nonlinear signals such as those detected by the vibration monitoring of the tail boom support of a longwall mining machine. Singular values were obtained by the application of SVD to the first seven IMFs of the example raw vibration signals (those IMFs containing key feature information), which could be used as the feature input vectors of the classifier. Based on these results, the SVM applied to the singular value vector is proposed as the classification tool for top-coal or coal-gangue caving state. The validation test had a 100% classification accuracy rate, providing strong support for the robustness of this method. Therefore, the analysis based on SVD, EMD, and

SVM for longwall mining machine tail boom vibrations offers a new method for CID.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The work for this paper was supported by the Shandong Provincial Natural Science Foundation of China (Grant nos. ZR2013EEM009 and ZR2013FL019). The authors are grateful to the anonymous reviewers for their careful reviews and detailed comments.

References

- [1] Z.-C. Wang and X.-J. Zhang, "The research on detection of rock content in coal rock mixture in top coal caving by natural gamma ray," *Chinese Journal of Sensors and Actuators*, vol. 16, no. 4, pp. 442–446, 2003.
- [2] G. L. Mowrey, "Horizon control holds key to automation," *Coal*, vol. 97, pp. 44–49, 1992.
- [3] G. L. Mowrey, "New approach to coal interface detection: the in-seam seismic technique," *IEEE Transactions on Industry Applications*, vol. 24, no. 4, pp. 660–665, 1988.
- [4] J. Zou and J. Chen, "A comparative study on time-frequency feature of cracked rotor by Wigner-Ville distribution and wavelet transform," *Journal of Sound and Vibration*, vol. 276, no. 1-2, pp. 1–11, 2004.
- [5] N. Baydar and A. Ball, "A comparative study of acoustic and vibration signals in detection of gear failures using Wigner-Ville distribution," *Mechanical Systems and Signal Processing*, vol. 15, no. 6, pp. 1091–1107, 2001.
- [6] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical Systems and Signal Processing*, vol. 18, no. 2, pp. 199–221, 2004.
- [7] G. Meltzer and N. P. Dien, "Fault diagnosis in gears operating under non-stationary rotational speed using polar wavelet amplitude maps," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 985–992, 2004.
- [8] H. T. Vincent, S. L. J. Hu, and Z. Hou, "Damage detection using empirical mode decomposition method and a comparison with wavelet analysis," in *Proceedings of the 2nd International Workshop on Structural Health Monitoring*, pp. 891–900, Stanford, Calif, USA, 1999.
- [9] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [10] N. E. Huang, "Computer implicated empirical mode decomposition method, apparatus, and article of manufacture," Patent US 6311130 B1, 1996.
- [11] W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung, "Engineering analysis of biological variables: an example of blood pressure over 1 day," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 9, pp. 4816–4821, 1998.
- [12] A.-J. van der Veen, E. F. Deprettere, and A. L. Swindlehurst, "Subspace-based signal analysis using singular value decomposition," *Proceedings of the IEEE*, vol. 81, no. 9, pp. 1277–1308, 1993.
- [13] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 2000.
- [14] W. Liu, R.-L. Wang, S.-X. Zhang, and Y. Zhang, "Real-time vibration signal acquisition of coal and rock based on DSP and ADS8509," *Coal Mine Machinery*, vol. 30, pp. 87–199, 2009.
- [15] X.-D. Zhang, *Modern Signal Processing*, Tsinghua University Press, Beijing, China, 2002.

Research Article

Palm Print Edge Extraction Using Fractional Differential Algorithm

Chunmei Chi and Feng Gao

Computer Engineering School, Qingdao Technological University, Qingdao 266033, China

Correspondence should be addressed to Feng Gao; gaofeng99@sina.com

Received 22 January 2014; Accepted 10 March 2014; Published 10 April 2014

Academic Editor: Gongzhu Hu

Copyright © 2014 C. Chi and F. Gao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Algorithm based on fractional difference was used for the edge extraction of thenar palm print image. Based on fractional order difference function which was deduced from classical fractional differential G-L definition, three filter templates were constructed to extract thenar palm print edge. The experiment results showed that this algorithm can reduce noise and detect rich edge details and has higher SNR than traditional methods.

1. Introduction

Thenar area refers to the uplift muscle which is the joint of one's wrist and thumb. Everyone has his or her unique thenar palm print patterns. But the palm print texture structure changes with age and disease. For example, roughness and thickness appear in the palm print with disease. According to the Traditional Chinese Medicine theory, one's thenar palm print image is closely related to certain diseases such as asthma. For some patients, lattice distribution and rough texture appear in their thenar palm print. Doctors tend to think that the shape of one's palm print is closely related to his or her kidney deficiency and could be an indicator of asthma and can be used for the diagnosis of some allergic disease. In the medical practice of Traditional Chinese Medicine, by observing the patient's thenar palm print direction, the distance between palm print lines, groove depth, grain interaction, and pattern characteristic, doctors usually divide the thenar palm print into 4 levels. Class A: thenar skin is moist and soft with delicate texture and furrow is very shallow; no distribution characteristic pattern can be seen. Class B: thenar skin is moist and soft, texture is clear, and lattice type distribution can be found, but the gap is narrow. Class C: thenar skin is smooth but less moist, texture is clear

and visible, lattice type distribution can be found, and the gap is wider. Class D: thenar skin is dry and rough, texture is clear and visible, large lattice type distribution can be found, and texture space is significantly widened compared with the Class C level. So far, the classification of thenar palm print of patients relies on the subjective observation of the doctors' naked eyes. In addition to medical treatment, palm print is also studied for other purposes [1–4].

In order to make the observation more objective and more accurate, we should conduct preprocessing, feature extraction and classification of the thenar palm print by using image processing technology. The computing is particularly important in order to realize the quantitative identification of thenar palm print. Thenar palm print image edge mainly refers to the texture features, such as lines, the distance between the lines, and the pattern characteristic, which is the main basis to carry out further classification. So effective edge detection plays an important role in feature extraction and quantitative identification. Edge detection (see [5–14]) method can detect the image edge to the precise location of thenar palm print and can also inhibit irrelevant details and noise. The classic methods of image edge detection like gradient method, conventional first-order differential and two-order differential method, and Laplace differential

algorithm are sensitive to noise. Their performance for the thenar palm print image edge extraction is not accurate and could not obtain the edge localization and achieve noise filtering. Generally speaking, the low frequency part of the image corresponds to the smooth area, namely, the image of the nonedge region, while the middle frequency part of the image corresponds to texture information and the high frequency part corresponds to the image edge area and noise.

With regard to the thenar palm print image processing, if an integer order differential method is employed, texture details with little gray changes in the palm print image will decay greatly. If fractional differential algorithm is employed, this information can be retained to some extent. So compared with integer order differential method, the fractional differential method is in a better position to deal with the high frequency texture image. As for the thenar palm print image, texture information is what the doctors are really interested in. So the edge detection algorithm using fractional order differential characteristics of palm print is preferred. In this paper, first, the fractional difference equation is obtained according to G-L fractional differential definition. Second, we construct some filter templates to extract the palm print image edge. The experiments show that the fractional differential edge detection algorithm is better than integer order differential methods and also has higher signal-to-noise ratio.

2. Fractional Differential Operators

2.1. Fractional Calculus. Fractional calculus should really be called noninteger order calculus. As early as nineteenth century, the concept of fractional calculus appeared together with integer order calculus. In the past two hundred years, fractional calculus has gradually developed into a system as a pure mathematical branch. But not until recent decades, and with the development of engineering technology, has fractional calculus caused the attention of both scientists and engineers. It was found that the fractional differential operator has many advantages over integer order ones in many engineering problems. Many scholars have developed different definitions of fractional calculus, among which the most well-known definitions are Riemann-Liouville's definition and Caputo's definition [15, 16]. These definitions are equivalent to each other in most cases. But Grunwald-Letnikov's definition is easier to be applied in numerical analysis. In this paper, we use Grunwald-Letnikov's definition to construct fractional order differential filter for image filtering and edge extraction of thenar palm print.

2.2. Fractional Difference. Fractional differential definition is a kind of generalization of the classical integer order derivative definition for continuous function. Usually, the n th derivative of a continuous function is defined as

$$\frac{d^n f(x)}{dx^n} = \lim_{h \rightarrow 0} \frac{1}{h^n} \sum_{i=0}^n (-1)^i \binom{n}{i} f(x - ih). \quad (1)$$

The above differential operator does not need to be constrained only to integers. For example, the $(1+i)$ th derivative of the $(1-i)$ th derivative yields the 2nd derivative.

For a general function $f(x)$ and $0 < \alpha < 1$, the complete fractional derivative is

$$D^\alpha f(x) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dx} \int_0^x \frac{f(t)}{(x-t)^\alpha} dt. \quad (2)$$

For arbitrary α , because the gamma function is undefined for arguments whose real part is a negative integer, it is necessary to apply the fractional derivative after the integer derivative has been performed. For example,

$$D^{3/2} f(x) = D^{1/2} D^1 f(x) = D^{1/2} \frac{d}{dx} f(x). \quad (3)$$

Generally, we have Riemann-Liouville fractional derivative

$$D_t^\alpha f(t) = \frac{d^n}{dt^n} D_t^{-(n-\alpha)} f(t) = \frac{d^n}{dt^n} I_t^{n-\alpha} f(t). \quad (4)$$

The corresponding derivative is calculated using Lagrange's rule for differential operators. Computing n th order derivative over the integral of order $n-\alpha$, the α order derivative is obtained. It is important to remark that n is the nearest integer bigger than α .

Caputo Fractional Derivative. There is another way for defining fractional derivatives, the Caputo fractional derivative which was introduced by M. Caputo. In contrast to the Riemann Liouville fractional derivative, when solving differential equations using Caputo's definition, it is not necessary to define the fractional order initial conditions. Caputo's definition is as follows:

$${}_a^C D_t^\alpha f(x) = \frac{1}{\Gamma(n-\alpha)} \int_a^x \frac{f^{(n)}(t)}{(x-t)^{\alpha+1-n}} dt. \quad (5)$$

Besides the above two definitions, we have also Grunwald-Letnikov's fractional derivative definition as shown below by using Γ function:

$$\frac{d^\alpha f(x)}{dx^\alpha} = \lim_{h \rightarrow 0} \frac{1}{h^\alpha} \sum_{i=0}^{\infty} (-1)^i \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-i+1)} f(x - ih), \quad (6)$$

where

$$\Gamma(n) = \int_0^{\infty} t^{n-1} \cdot e^{-t} dt = (n-1)! \quad (7)$$

and α can be noninteger. The α th difference of $f(x)$ can be expressed as follows:

$$\begin{aligned} f^\alpha(x) \approx & f(x) + (-\alpha) f(x-1) + \frac{(-\alpha)(-\alpha+1)}{2} \\ & \cdot f(x-2) + \frac{(-\alpha)(-\alpha+1)(-\alpha+2)}{3!} f(x-3) \\ & + \cdots + \frac{\Gamma(-\alpha+1)}{n! \Gamma(-\alpha+n+1)} f(x-n). \end{aligned} \quad (8)$$

The α th difference is actually the approximation of the α th derivative. Fractional calculus has been applied to many areas [15–20].

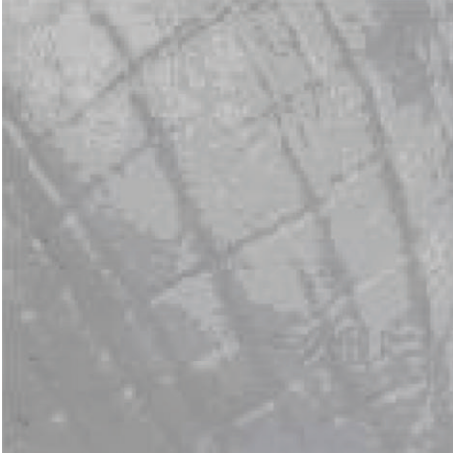


FIGURE 1: The original image.



FIGURE 3: Image got by Prewitt operator.

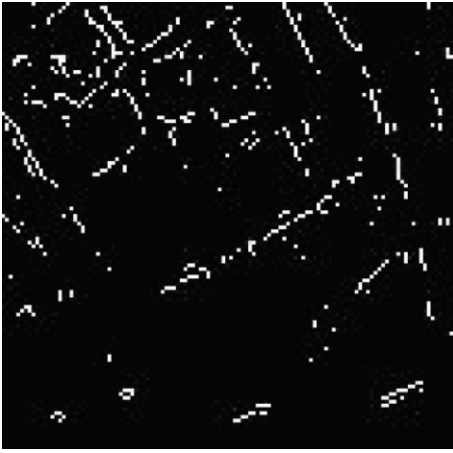


FIGURE 2: Image got by Sobel operator.



FIGURE 4: Image got by Laplace operator.

2.3. Fractional Order Filter. In general, the $M \times N$ image filter template for a gray image is given as follows:

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x + s, y + t), \quad (9)$$

where $a = (m - 1)/2$ and $b = (n - 1)/2$.

In order to get a complete picture of the filtered image, the filter template must be applied to $X = 0, 1, \dots, M - 1$ and $Y = 0, 1, \dots, N - 1$. Thus, we guarantee that all the pixels in the image are filtered. For digital images, the fractional order differential filter can be equal to the image itself in size. In order to realize the filtering and minimize the error, we can select the first three terms in the fractional difference to construct isotropic filter. We construct the below fractional order differential filter template 1, filter template 2, and filter template 3. In the structure of filter 3, we first consider the X direction and Y direction around the upper and lower side. So there are two $f(x, y)$ terms in the two directions; therefore, there are four $f(x, y)$ terms. Considering the effect of oblique direction of 45 degrees above the pixels, we add 4 oblique

TABLE 1: Filter template 1.

| | | |
|-----------|---|-----------|
| $-\alpha$ | | $-\alpha$ |
| | 4 | |
| $-\alpha$ | | $-\alpha$ |

TABLE 2: Filter template 2.

| | | |
|-----------|-----------|-----------|
| $-\alpha$ | $-\alpha$ | $-\alpha$ |
| $-\alpha$ | 8 | $-\alpha$ |
| $-\alpha$ | $-\alpha$ | $-\alpha$ |

directions. Because each oblique above contains one $f(x, y)$ term, there are 8 $f(x, y)$ terms for all the different directions.

All the fractional order differential filter templates have the same direction of rotation (see Tables 1, 2, and 3).

3. Image Edge Extraction Experiments

Experiments were carried out by using MATLAB. The thenar palm print image was taken by the digital camera in high

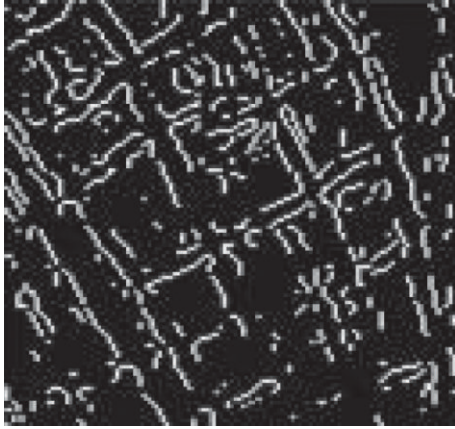


FIGURE 5: Edge extracted by Sobel operator.

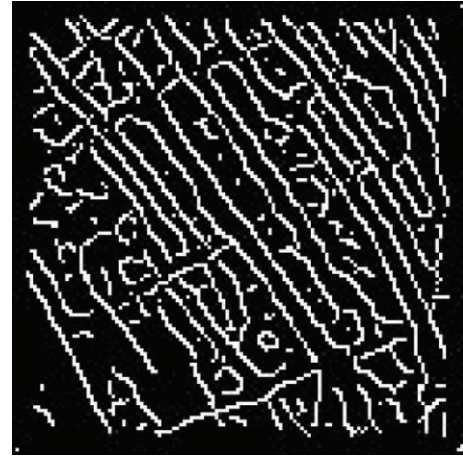


FIGURE 7: Edge extracted by Laplace operator.

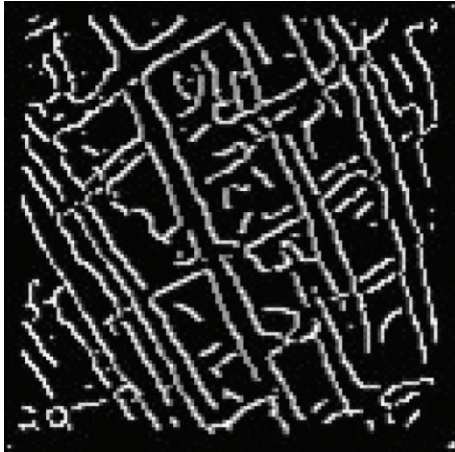


FIGURE 6: Edge extracted by Prewitt operator.

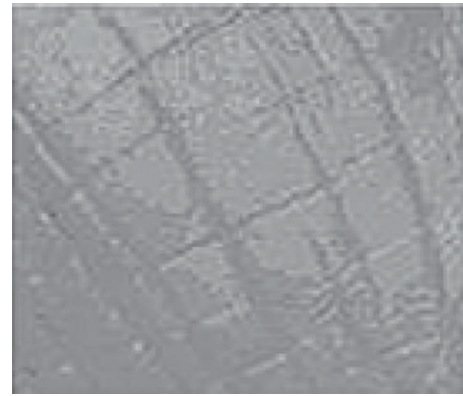


FIGURE 8: Image got by 0.9th template 1.

TABLE 3: Filter template 3.

| $\frac{\alpha * \alpha - \alpha}{2}$ | | $\frac{\alpha * \alpha - \alpha}{2}$ | | $\frac{\alpha * \alpha - \alpha}{2}$ |
|--------------------------------------|-----------|--------------------------------------|-----------|--------------------------------------|
| $\frac{\alpha * \alpha - \alpha}{2}$ | $-\alpha$ | $-\alpha$ | $-\alpha$ | $\frac{\alpha * \alpha - \alpha}{2}$ |
| $\frac{\alpha * \alpha - \alpha}{2}$ | $-\alpha$ | 8 | $-\alpha$ | $\frac{\alpha * \alpha - \alpha}{2}$ |
| $\frac{\alpha * \alpha - \alpha}{2}$ | $-\alpha$ | $-\alpha$ | $-\alpha$ | $\frac{\alpha * \alpha - \alpha}{2}$ |
| $\frac{\alpha * \alpha - \alpha}{2}$ | | $\frac{\alpha * \alpha - \alpha}{2}$ | | $\frac{\alpha * \alpha - \alpha}{2}$ |

light background. The digital camera's shooting angle is perpendicular to the palm thenar area and the palm is in the natural state of relaxation. Firstly, we carry out the palm print image segmentation to identify the thenar palm print region by using palm print positioning point segmentation method based on the segmentation of the thenar palm print image. Second, we use the three-edge detection method to process the thenar palm print image. The test results are shown in Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15. For comparison, we also use the traditional integer order differential operator Sobel, Roberts, Prewitt, LOG, and Laplace operator to construct filter templates to process the

thenar palm print image. The traditional edge detection is mainly to estimate the direction gradient of gray image by scale change detector of gradient operator or derivative based on these changes. We enhance the area in the image and then carry out the threshold operation on the gradient. If the gradient value is greater than a given threshold, then there is edge. Test effect is shown below. It can be seen from Figure 1 that different filter templates yield different palm print edge intensities.

3.1. Analysis of Fractional Order on the Processing Results.

The experiment analyses of filter template 1, filter template 2, and filter template 3 show that when the fractional order is 0.5, the palm print image is clear and with least noise. With the elevation of fractional order, palm print texture details become clear and rich gradually. As the filter size increases, the edge becomes clear, but not very obvious.

3.2. Comparison of Fractional Differential Algorithm with Other Edge Extract Operators.

The image in Figure 5 is the classical first-order Sobel operator, Prewitt operator, and the second-order Laplacian operator and fractional order (0.5 order) comparison diagram edge extraction operator. The



FIGURE 9: Image got by 0.8th template 1.



FIGURE 11: Image got by 0.6th template 1.

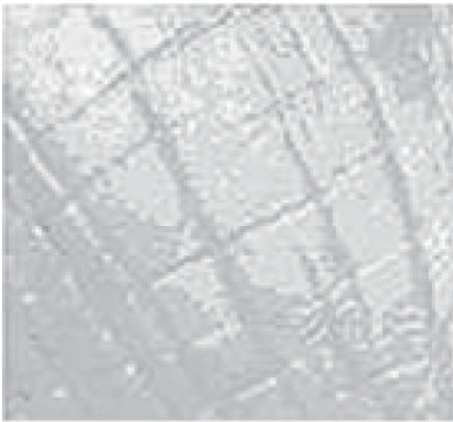


FIGURE 10: Image got by 0.7th template 1.

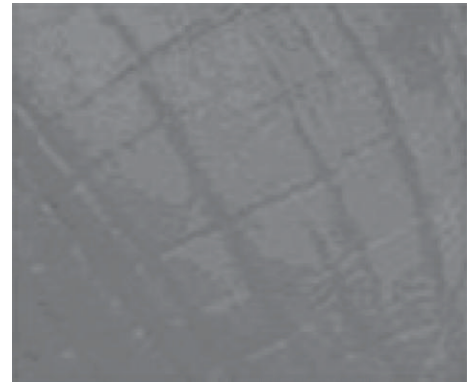


FIGURE 12: Image got by 0.1th template 2.

diagram shows that fractional templates detect edge detail and localization accurately.

3.3. The Peak Signal-to-Noise Ratio for Different Algorithms. We add random noise to palm print image and compare the peak signal-to-noise ratio (PSNR) for various algorithms to extract edge information. The peak signal-to-noise ratio (PSNR) formula is

$$\text{psnr}(f_0, f) = 10 \log_{10} \frac{N \times M \times 255^2}{\sum_{j=1}^M \sum_{i=1}^N [f(i, j) - f_0(i, j)]^2}. \quad (10)$$

The higher the peak signal-to-noise ratio (PSN) is, the better the effect to extract texture information is. We adopt the following evaluation method. We first do image edge extraction with adding random noise and without noise by using different methods. Then we analyze the difference between the two methods. In the above equation, $f(x, y)$ represents image edge extracted from the noise added image.

In Table 1, the peak signal-to-noise ratio of different order differential extractions based on edge information is listed where the stochastic noise with mean value 0 and variance 0.003 was added. And from the result we can see that the

TABLE 4: PSNR of edge information extracted by different order differential.

| Number of iter. | Method | | | | | |
|-----------------|--------|-------|------------------|-------|-------|-------|
| | Sob | Lap | Fractional order | | | |
| | | | 0.6 | 0.4 | 0.2 | 0.1 |
| 1 | 22.55 | 22.40 | 23.15 | 23.36 | 23.55 | 22.74 |
| 2 | 23.7 | 23.81 | 24.47 | 24.56 | 24.63 | 24.15 |

antinoise performance of differential algorithm is good and can be used to extract the edge of thenar palm print image.

In Table 4, we list the peak signal-to-noise ratio for different methods to extract the edge of image.

From Table 4, it can be seen that although the impact of the two added noises is different, the signal-to-noise ratio of the differential algorithm is higher than that of Sobel method and Laplace method. And the peak signal-to-noise ratio decreases with the increase of differential order. The order of peak signal-to-noise ratio and the maximum value appeared in the 0.2 order. When the order is less than 0.2, the peak signal-to-noise ratio began to decrease. If the differential order is very small, the fractional algorithm has no significant effect on the enhancement of palm print texture; meanwhile, the image texture information has an effect on extraction of the edge. We also notice that the noise has different effects on different parts of the image. As shown in Table 1, for the mean value 0 and variance 0.003 random noise, the effect on the

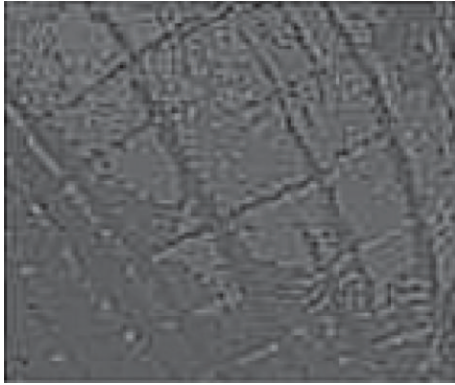


FIGURE 13: Image got by 0.9th template 3.

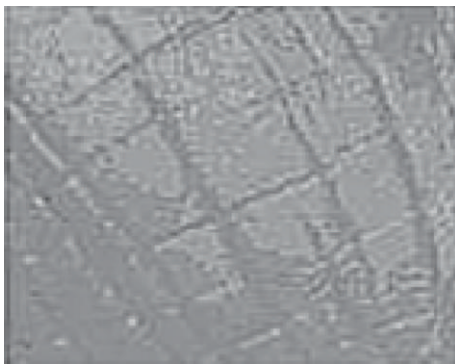


FIGURE 14: Image got by 0.8th template 3.

extraction of edge information is different. At the same time filter templates 1–3 are only an approximation of fractional order differential template, so we can only conclude that the optimal fractional order of differential template number is 0.2.

4. Conclusion

Simulation results show that the ability of fractional differential edge detection algorithm to extract the image edge information varies from different fractional orders. In general, the ability gradually increases with the increasing of order. The ability also increases with the template size but not in a positive proportional manner. Compared with the traditional first-order and second-order differential operators, the edge detection based on fractional differential method has higher signal-to-noise ratio. In this paper, the template itself is actually the approximate representation of a fractional differential method. So how to find the fractional order differential operators to achieve higher accurate edge extraction is a problem for further study.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.



FIGURE 15: Image got by 0.7th template 3.

Acknowledgment

This work is supported by the National Natural Science Foundation of China, no. 31271077.

References

- [1] D. Zhang, *Palm Print Authentication*, Kluwer Academic, New York, NY, USA, 2004.
- [2] E. Liu, A. K. Jain, and J. Tian, "A coarse to fine minutiae-based latent palm print matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, pp. 2307–2322, 2013.
- [3] J. Yao and W.-K. Cham, "3D modeling and rendering from multiple wide-baseline images by match propagation," *Signal Processing: Image Communication*, vol. 21, no. 6, pp. 506–518, 2006.
- [4] E. Liu, H. Zhao, J. Liang, L. Pang, H. Chen, and J. Tian, "Random local region descriptor (RLRD): a new method for fixed-length feature representation of fingerprint image and its application to template protection," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 236–243, 2012.
- [5] H. Jeong and C. I. Kim, "Adaptive determination of filter scales for edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 5, pp. 579–585, 1992.
- [6] T. Aydin, Y. Yemez, E. Anarim, and B. Sankur, "Multidirectional and multiscale edge detection via M-band wavelet transform," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1370–1377, 1996.
- [7] A. P. Paplinski, "Directional filtering in edge detection," *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 611–615, 1998.
- [8] C.-J. Sze, H.-Y. M. Liao, H.-L. Hung, K.-C. Fan, and J.-W. Hsieh, "Multiscale edge detection on range images via normal changes," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 1087–1092, 1998.
- [9] C. F. Olson, "Adaptive-scale filtering and feature detection using range data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 9, pp. 983–991, 2000.
- [10] D. Demigny, "On optimal linear filtering for edge detection," *IEEE Transactions on Image Processing*, vol. 11, no. 7, pp. 728–737, 2002.
- [11] S. Konishi, A. L. Yuille, J. M. Coughlan, and S. C. Zhu, "Statistical edge detection: learning and evaluating edge cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 57–74, 2003.

- [12] J. Wu, Z. Yin, and Y. Xiong, "The fast multilevel fuzzy edge detection of blurry images," *IEEE Signal Processing Letters*, vol. 14, no. 5, pp. 344–347, 2007.
- [13] P. Bao, L. Zhang, and X. Wu, "Canny edge detection enhancement by scale multiplication," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1485–1490, 2005.
- [14] B. Mathieu, P. Melchior, A. Oustaloup, and C. Ceyral, "Fractional differentiation for edge detection," *Signal Processing*, vol. 83, no. 11, pp. 2421–2432, 2003.
- [15] K. B. Oldham and J. Spanier, *The Fractional Calculus*, Academic Press, New York, NY, USA, 1974.
- [16] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo, *Theory and Applications of Fractional Differential Equations*, Elsevier, Amsterdam, The Netherlands, 2006.
- [17] S. W. Wheatcraft and M. M. Meerschaert, "Fractional conservation of mass," *Advances in Water Resources*, vol. 31, no. 10, pp. 1377–1381, 2008.
- [18] W. Chen, H.-G. Sun, X. Zhang, and D. Korošak, "Anomalous diffusion modeling by fractal and fractional derivatives," *Computers & Mathematics with Applications*, vol. 59, no. 5, pp. 1754–1758, 2010.
- [19] H.-G. Sun, W. Chen, and Y.-Q. Chen, "Variable-order fractional differential operators in anomalous diffusion modeling," *Physica A*, vol. 388, no. 21, pp. 4586–4592, 2009.
- [20] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives*, Gordon and Breach, Yverdon, Switzerland, 1993.

Research Article

A Joint Encryption and Reversible Data Hiding Scheme Based on Integer-DWT and Arnold Map Permutation

Shun Zhang, Tiegang Gao, and Guorui Sheng

College of Software, Nankai University, Wei Jin Road No. 94, Nankai District, Tianjin 300071, China

Correspondence should be addressed to Shun Zhang; shentengvip@gmail.com

Received 6 January 2014; Revised 26 March 2014; Accepted 26 March 2014; Published 9 April 2014

Academic Editor: Feng Gao

Copyright © 2014 Shun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A joint encryption and reversible data hiding (joint encryption-RDH) scheme is proposed in this paper. The cover image is transformed to the frequency domain with integer discrete wavelet transform (integer DWT) for the encryption and data hiding. Additional data is hidden into the permuted middle (LH, HL) and high (HH) frequency subbands of integer DWT coefficients with a histogram modification based method. A combination of permutations both in the frequency domain and in the spatial domain is imposed for the encryption. In the receiving end, the encrypted image with hidden data can be decrypted to the image with hidden data, which is similar to the original image without hidden data, by only using the encryption key; if someone has both the data hiding key and the encryption key, he can both extract the hidden data and reversibly recover the original image. Experimental results demonstrate that, compared with existing joint encryption-RDH schemes, the proposed scheme has gained larger embedding capacity, and the distribution of the encrypted image with data hidden has a random like behavior. It can also achieve the lossless restoration of the cover image.

1. Introduction

Compared with traditional watermarking and data hiding schemes, reversible data hiding schemes can be applied in a larger field of secure communication and watermarking due to its reversibility. Many reversible data hiding schemes have been proposed in recent years, which can be classified into three main catalogues: the first one is compression based scheme [1], the second one is difference expansion based scheme [2–5], and the third one is histogram modification based scheme [6–10]. Reversible data hiding based on compression makes use of the redundancy of cover images, so the characters of the cover images limit the capacity and quality of data hiding. Difference expansion based scheme was firstly proposed by Tian [5], which hid one-bit data by extending the difference between two neighbor pixels. Alattar [2–4] improved the hiding capacity by extending $n - 1$ pairs of neighbor pixels' differences to hide $n - 1$ bits data. However, the quality of the cover image drops quickly, while the hiding capacity increases. Schemes based on histogram modification cause less distortion to the cover image. However, the peak points of the histogram limit the hiding capacity [11]. There

are two measures to increase the hiding capacity in histogram modification based data hiding schemes: raising the peak points' height or increasing the number of peak points. Many schemes based on the two ways have been proposed. Lin et al. [6] proposed a multilevel embedding strategy to increase the number of peak points. Some schemes increased the height of the peak points through generating the histogram of the difference image. For example, Tsai et al. [8] constructed the difference image by a prediction model that makes full use of the similarity between neighbor pixels. Kim et al. [12] sampled the original image to construct the difference images. A predicted image based on the sampled images was constructed. Then the histograms of difference images between the predicted image and these sampled images were generated for data hiding.

As is well known, encryption is an old and efficient way in secure communication. If combined with encryption, reversible data hiding will achieve greater security. Besides, there are also scenarios that data hiding needs to be done in the encrypted domain or combined with the encryption, especially in the age of big data and cloud computing. A content owner does not trust the processing service provider, and

the ability to manipulate the encrypted data while keeping the plain content unrevealed is desired [13]. Suppose that there are sensitive images uploaded to the cloud storage in the encrypted form and some additional data needs to be hidden into these images to mark their ownership. However, the data hiding process has to be done in the encrypted domain because the data administrator does not have the right and the key to decrypt the image.

In the past few years, some schemes that combine encryption and data hiding have been proposed [13–18]. From the data hider's point of view, data can be hidden into the spatial domain, the encrypted domain [13, 14, 16–18], or both of the two domains [15]. Although high image quality after data hiding has been achieved in [15], the scheme is not reversible. Reversible data hiding schemes in encrypted images are proposed in [13, 14]. In [14], an improved measurement of smoothness is proposed to make full use of all the pixels in the image, and a side match scheme is proposed to further decrease the error rate of extracted bits, both of which have improved the embedding capacity of the basic data hiding scheme in the encrypted image proposed in [13]. In [16], a reversible data hiding scheme in encrypted images by reserving room before encryption is proposed. The self-embedding of LSB planes guarantees the reversibility of LSB substitution embedding. However, the embedding capacity is limited by the embedding capacity of the reversible data hiding scheme in the selected area. In [17], some pixels are selected and estimated before encryption, and additional data is embedded into the estimated errors with a histogram modification method. In the receiving end, one can either decrypt the image with hidden data first or extract the hidden data first. Scheme proposed in [18] separates the data extraction and the recovery of original image. The image is encrypted with the encryption key. Then the encrypted image is passed to the data hider, and additional data is embedded into the encrypted image with the data hiding key. In the receiving end, the hidden data can be extracted with only the data hiding key; and only similar (not reversible) image can be recovered with the encryption key; both the hidden data can be extracted and original image can be reversibly recovered with both keys.

Different from all the joint encryption and data hiding schemes mentioned above, a joint encryption and reversible data hiding scheme based on integer DWT and Arnold map permutation is proposed. Not in the spatial domain, the data hiding is imposed in the integer DWT domain, which is more secure compared with those schemes in the spatial domain. The cover image is firstly transformed to the frequency domain with discrete integer wavelet transform (integer DWT). Then coefficients of the four subbands are permuted with Arnold map transform, respectively, for the first time. After that, additional data is embedded into the permuted middle (LH, HL) and high (HH) frequency subbands through a histogram modification based method. Finally, inverse integer DWT is imposed to get the primary encrypted image with hidden data. Another Arnold permutation based on sampling, which is related to the permutation in the frequency domain, is imposed on the primary encrypted image with hidden data in the spatial domain. In the receiving

end, one can decrypt the image to get the image with hidden data, which is similar to the original image without hidden data, by only using the encryption key that includes permutation times of the twice permutations. If someone has both the encryption key and the data hiding key, he can both extract the hidden data and reversibly recover the original image. Note that the processing procedures in the sending end and in the receiving end described here are asymmetric, which can achieve many applications, such as scenarios mentioned above.

2. Preliminaries

2.1. Integer DWT. To achieve the reversible data hiding, reversible lifting integer DWT is applied. Integer DWT is implemented with the addition and subtraction of integers. Suppose that $I(x, y)$, $1 \leq x \leq M$, $1 \leq y \leq N$, is the pixel of the image size of $M \times N$; then 2D integer DWT is conducted as follows.

(A) Row Transformation

- (1) Let $f1 = I(2 * i - 1, :)$ and $f2 = I(2 * i, :)$, $i = 1, 2, \dots, M/2$, which are odd rows and even rows of I , respectively.
- (2) Acquire the high frequency coefficients by calculating the difference of the two: $h_r(i, :) = f1(i, :) - f2(i, :)$.
- (3) Acquire the low frequency coefficients by calculating the average of the two: $l_r(i, :) = f2(i, :) + \text{floor}(h(i, :))$.
- (4) Then coefficients after 1D transformation are $C_row = [l_r; h_r]$.

(B) Column Transformation

- (1) Let $f1 = I(:, 2 * i - 1)$ and $f2 = C_row(:, 2 * i)$, $i = 1, 2, \dots, N/2$, which are odd columns and even columns of C_row , respectively.
- (2) Acquire the high frequency coefficients by calculating the difference of the two: $h_c(:, i) = f1(:, i) - f2(:, i)$.
- (3) Acquire the low frequency coefficients by calculating the average of the two: $l_c(:, i) = f2(:, i) + \text{floor}(h(:, i))$.
- (4) Finally, the coefficients of 2D integer DWT are $C = [l_c \ h_c]$.

2.2. Reversible Data Hiding and Data Extraction Based on Histogram Expansion. Ni et al. [19] firstly proposed reversible data hiding based on histogram modification. It generates the histogram of an image; then a pair of peak point and zero point is found out in the histogram, and the histogram between peak point and zero point is shifted to the zero point side to produce the gap for data hiding. Very little distortion will be caused by such schemes, and Ni et al. [19] have pointed out that the peak signal-to-noise ratio (PSNR) between the original image and the image with hidden data is above 48. As mentioned in the beginning part of the paper, the drawback is the rare capacity of data hiding. A novel histogram modification based reversible data hiding scheme in integer DWT domain, which increases the capacity of data hiding greatly,

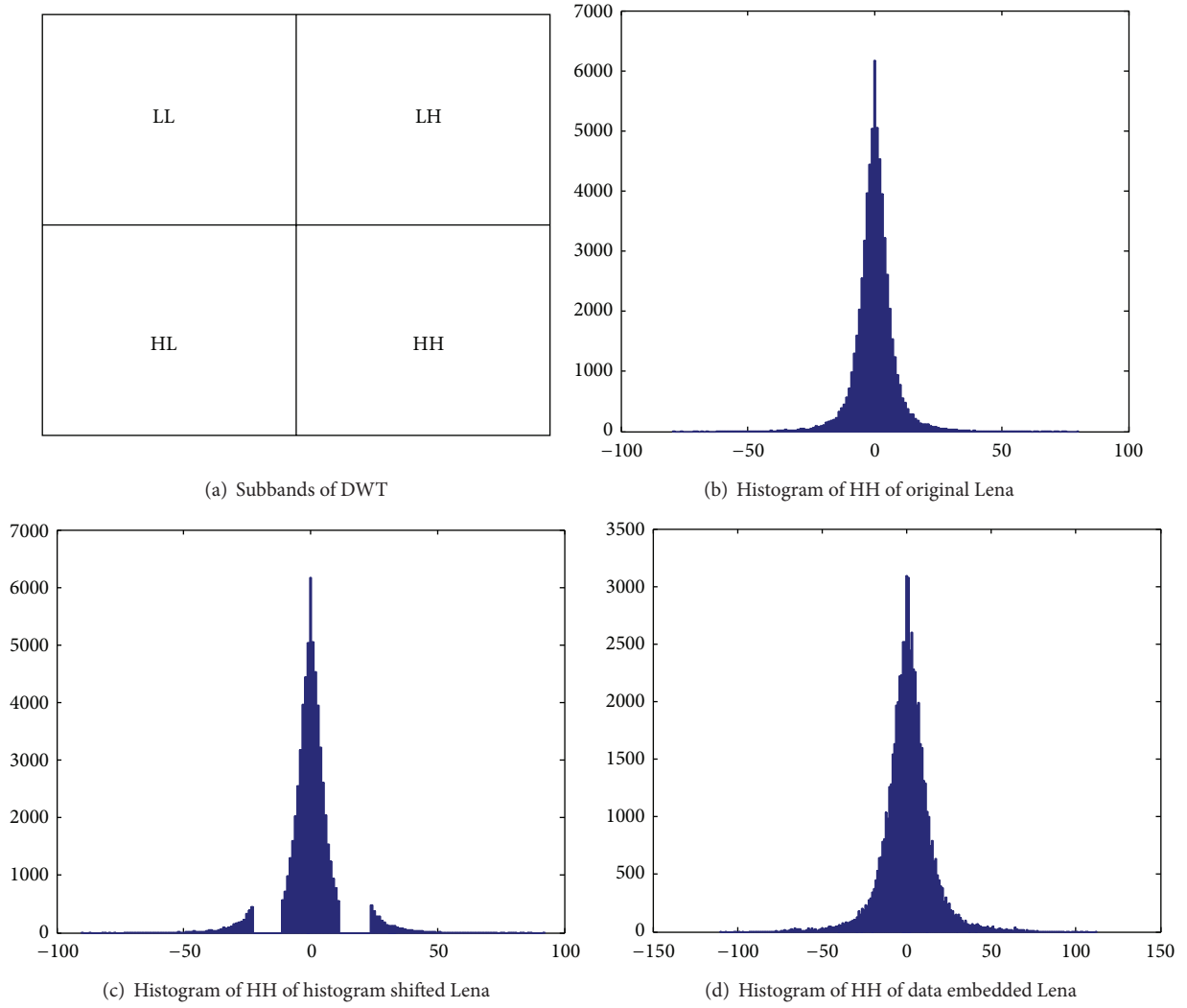


FIGURE 1: An example of histogram modification (HH subband of integer DWT of Lena).

is described here. Histograms of middle and high frequency subbands of integer DWT of images are Laplacian like distribution. Thus, they are suitable for histogram modification based data hiding method. Histograms are shifted to generate the gap for data hiding. A demo of histogram modification based data hiding method, which embeds data into the HH subbands of Lena image, is presented in Figure 1.

2.2.1. Reversible Data Embedding. The generated histogram of subband HH is depicted in Figure 1(b). Then the histogram is shifted to both sides by an embedding strength T (Figure 1(c)). At last, data is embedded by expanding histogram between T and $-T$, and the histogram after embedding is as Figure 1(d).

The histograms of LH, HL, and HH subbands are generated and data is embedded into the coefficients by histogram modification. For every coefficient C of LH, HL, and HH subbands, given an embedding strength parameter q ,

- (1) if $C \geq q$, then C is shifted to $C + q$;
- (2) else if $C \leq -q$, then C is shifted to $C - q + 1$;
- (3) else $C \leftarrow 2 \times C + B$, and B is the data to be embedded.

2.2.2. Data Extraction and Reversible Recovery of Matrix before Embedding. Generate the histograms of middle and high frequency subbands and shift these histograms to extract the hidden data, and the original coefficient matrices are reversibly recovered through the following steps. For every coefficient C of LH, HL, and HH subbands, given an embedding strength parameter q ,

- (1) if $C \geq 2 \times q$, then C is shifted to $C - q$;
- (2) else if $C \leq -2 \times q + 1$, then C is shifted to $C + q - 1$;
- (3) else $C \leftarrow \text{floor}(C/2)$, and data is extracted: $B = \text{mod}(C, 2)$.

Now, every coefficient C of subbands LH, HL, and HH is reversibly recovered and the extracted B is the data embedded before.

2.3. Arnold Permutation [20]. Russian mathematician Vladimir I. Arnold discovered Arnold's cat map using an image of cat. An image, not necessarily a cat, of course, can be transformed to a random noise like image by rearranging the position of original pixels. However, if iterated for moderate times (denoted by permutation periods as presented in Table 1), the original image will reappear. The permutation periods differ as the sizes of images differ. The permutation periods of images with different sizes of the traditional Arnold permutation are presented in Table 1.

Let $I(x, y)$ be the pixel of an image matrix with size $N \times N$; then $\begin{bmatrix} x \\ y \end{bmatrix}$ represents the position of the pixel. The Arnold transform Γ can be explained as $\Gamma \begin{bmatrix} x \\ y \end{bmatrix} \rightarrow \begin{bmatrix} x+y \\ x+2y \end{bmatrix} \bmod n$, where mod is the modulo-operation.

To better explain the theory, the transform can be decomposed into three elemental steps: in the x -direction: $\begin{bmatrix} x \\ y \end{bmatrix} \rightarrow \begin{bmatrix} x+y \\ y \end{bmatrix}$, in the y -direction: $\begin{bmatrix} x+y \\ y \end{bmatrix} \rightarrow \begin{bmatrix} x+y \\ x+2y \end{bmatrix}$, and in the modulo-operation: $\begin{bmatrix} x+y \\ x+2y \end{bmatrix} \rightarrow \begin{bmatrix} x+y \\ x+2y \end{bmatrix} \bmod n$.

2.4. Permutation in the Frequency Domain and Sample Permutation in the Spatial Domain. Through permutation in the frequency domain, nice encryption results will be achieved. A novel encryption scheme based on the cooperation of permutation in the frequency domain and sample permutation in the spatial domain is proposed to accommodate the joint encryption-RDH scheme in this paper. It is found out through experiments that the proposed permutation in the integer DWT domain can achieve the same results as the proposed sample permutation scheme in the spatial domain. Such features are applied in the design of joint encryption-RDH scheme. Suppose that there is an image matrix M with size $N \times N$, and the permutation in the integer DWT domain and sample permutation in the spatial domain are described, respectively, in the followings.

(A) Permutation in the Integer DWT Domain

- (1) Decompose the original image matrix M with integer DWT to obtain the four subbands (1 low frequency subband LL, 2 middle frequency subbands LH and HL, and 1 high frequency subband HH) as depicted in Figure 1(a).
- (2) Permute the four subbands after integer DWT with Arnold map permutation that is presented in Section 2.1, and the different permutation times are $P1_T$, $P2_T$, $P3_T$, and $P4_T$, respectively:

$$\begin{aligned} LL' &= \text{Arnold}(LL, P1_T), \\ LH' &= \text{Arnold}(LH, P2_T), \\ HL' &= \text{Arnold}(HL, P3_T), \\ HH' &= \text{Arnold}(HH, P4_T). \end{aligned} \quad (1)$$

- (3) Impose inverse integer DWT on the coefficients after permutation to get the encrypted matrix M' .

(B) Sample Permutation in the Spatial Domain

- (1) Sample matrix M into four submatrices:

$$\begin{aligned} \text{Sam1} &= M(1 : 2 : N - 1, 1 : 2 : N - 1), \\ \text{Sam2} &= M(2 : 2 : N, 1 : 2 : N - 1), \\ \text{Sam3} &= M(1 : 2 : N - 1, 2 : 2 : N), \\ \text{Sam4} &= M(2 : 2 : N - 1, 2 : 2 : N). \end{aligned} \quad (2)$$

- (2) Permute the four sampled submatrices with Arnold map with different permutation times $S1_T$, $S2_T$, $S3_T$, and $S4_T$:

$$\begin{aligned} \text{Sam1}' &= \text{Arnold}(\text{Sam1}, S1_T), \\ \text{Sam2}' &= \text{Arnold}(\text{Sam2}, S2_T), \\ \text{Sam3}' &= \text{Arnold}(\text{Sam3}, S3_T), \\ \text{Sam4}' &= \text{Arnold}(\text{Sam4}, S4_T). \end{aligned} \quad (3)$$

- (3) Compose the permuted sampled submatrix to get the encrypted matrix M'

$$\begin{aligned} M'(1 : 2 : N - 1, 1 : 2 : N - 1) &= \text{Sam1}', \\ M'(2 : 2 : N, 1 : 2 : N - 1) &= \text{Sam2}', \\ M'(1 : 2 : N - 1, 2 : 2 : N) &= \text{Sam3}', \\ M'(2 : 2 : N - 1, 2 : 2 : N) &= \text{Sam4}'. \end{aligned} \quad (4)$$

If the permutation times are equal to the corresponding permutation times in the two permutation schemes, which mean that $P1_T = S1_T$, $P2_T = S2_T$, $P3_T = S3_T$, and $P4_T = S4_T$, the encryption results are equivalent.

3. Proposed Scheme

Different from existing joint encryption-RDH schemes [13–18], which are based on the spatial domain, the proposed scheme is based on the integer DWT domain. The detailed joint encryption-RDH scheme is presented in this section. The scheme is composed of two parts. One part is data hiding and image encryption, as presented in Figure 2; the other part is data extraction and original image recovery, which is presented in Figure 3. Note that the two parts are not symmetric. The encryption is achieved with permutation before data hiding and after data hiding in Figure 2, while the data extraction is after decryption in Figure 3. The asymmetric design can be applied in such a scenario. When someone only has the encryption key, he can decrypt the image and get the decrypted image with hidden information, which is very similar to the original image. The decrypted image can be utilized

TABLE 1: Arnold permutation periods of images with different sizes.

| Image size | 512×512 | 256×256 | 128×128 | 64×64 | 32×32 | 16×16 | 8×8 | 4×4 |
|------------|------------------|------------------|------------------|----------------|----------------|----------------|--------------|--------------|
| Period | 384 | 192 | 96 | 48 | 24 | 12 | 6 | 3 |

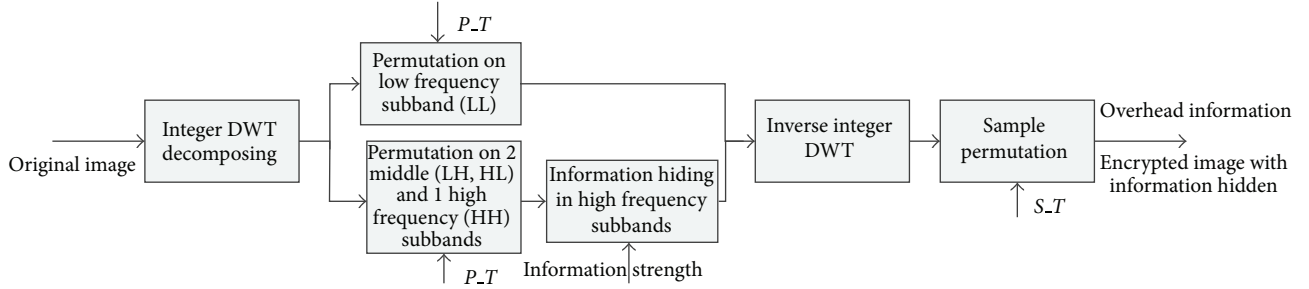


FIGURE 2: Data hiding and image encryption.

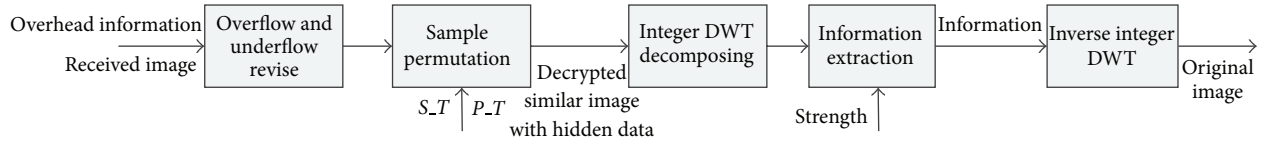


FIGURE 3: Data extraction and image recovery.

in a variety of applications. However, he cannot get rid of the hidden data, which may work as the watermark for the copyright or authentication.

In the sending end, the data hiding and image encryption process are achieved alternately. Original image I is firstly decomposed with the integer DWT proposed in Section 2.1. Then, the Arnold permutation is imposed on the four subbands for the first time encryption. Data is embedded into the permuted middle and high frequency subbands with a reversible data hiding scheme based on histogram modification. After that inverse integer DWT is imposed to acquire the primary permuted image with hidden data I' . Finally, a sample-permutation scheme is imposed on I' to get the final encrypted image with hidden data I'' . Because the reversible data hiding is based on histogram modification, overflow/underflow is hard to avoid. Therefore, a location map for recording the positions and values of the underflow and overflow pixels in the spatial domain is constructed. The location map is compressed and encrypted as the key for data extraction. The flow chart is presented in Figure 2.

In the receiving end, there are two cases. One is simple decryption, and the other one is data extraction and original image recovery. In the former case, the image with hidden data that is similar to the original image I is decrypted. In the latter case, the hidden data is extracted and the original image I is reversibly recovered. The received encrypted image with hidden data I'' is firstly revised according to the location map. Then it is decrypted into the image with hidden data. Note that, the encryption based on the two permutations in the sending end can be decrypted by one permutation based on the sample-permutation method. The permutation based on sample permutation in the spatial domain can achieve the

same results as the permutation in the integer DWT domain as presented in Section 2.4. Besides, the data hiding scheme based on histogram modification in the integer DWT domain can be implemented either before or after the permutation in the integer DWT domain. Both of the features guarantee the asymmetric decryption and data extraction. The total permutation times are calculated according to the Arnold permutation periods of image with different sizes (Table 1), the permutation time $P.T$ in the integer DWT domain permutation, and the permutation time $S.T$ of the sample permutation in the spatial domain. Through the delicate design of the two permutations, decryption can be done without integer DWT and inverse integer DWT. Although additional data is hidden in the permuted integer DWT domain, the proposed histogram modification based data hiding scheme in the integer DWT domain guarantees the integrity of the hidden data and the reversibility of the original cover image. The flow chart is presented in Figure 3.

3.1. Data Hiding and Image Encryption

- (1) Decompose the original image I (with size $N \times N$) with integer DWT (proposed in Section 2.1) to obtain the four subbands (one low frequency subband LL, two middle frequency subbands LH and HL, and one high frequency subband HH) as depicted in Figure 1(a).
- (2) Permute the four subbands (LL, LH, HL, and HH) with Arnold map permutation (proposed in

Section 2.3) synchronously. Note that, the permutation times are the same, denoted by P_T :

$$\begin{aligned} LL &= \text{Arnold}(LL, P_T), & LH &= \text{Arnold}(LH, P_T), \\ HL &= \text{Arnold}(HL, P_T), & HH &= \text{Arnold}(HH, P_T). \end{aligned} \quad (5)$$

- (3) Embed the preprocessed data into the middle and high frequency subbands (LH, HL, and HH) with histogram modification based method (proposed in Section 2.2.1).
- (4) Impose inverse integer DWT on the coefficients after permutation and data hiding to get the primary permuted image with hidden data, denoted by I' .
- (5) Sample I' to four subimages:

$$\begin{aligned} \text{Sam1} &= I' (1 : 2 : N - 1, 1 : 2 : N - 1), \\ \text{Sam2} &= I' (2 : 2 : N, 1 : 2 : N - 1), \\ \text{Sam3} &= I' (1 : 2 : N - 1, 2 : 2 : N), \\ \text{Sam4} &= I' (2 : 2 : N - 1, 2 : 2 : N). \end{aligned} \quad (6)$$

- (6) Permute the four sampled submatrices with Arnold map with different permutation times $S1_T$, $S2_T$, $S3_T$, and $S4_T$:

$$\begin{aligned} \text{Sam1}' &= \text{Arnold}(\text{Sam1}, S1_T), \\ \text{Sam2}' &= \text{Arnold}(\text{Sam2}, S2_T), \\ \text{Sam3}' &= \text{Arnold}(\text{Sam3}, S3_T), \\ \text{Sam4}' &= \text{Arnold}(\text{Sam4}, S4_T). \end{aligned} \quad (7)$$

- (7) Compose the permuted sampled submatrix to get the postprocessed matrix I'' :

$$\begin{aligned} I'' (1 : 2 : N - 1, 1 : 2 : N - 1) &= \text{Sam1}', \\ I'' (2 : 2 : N, 1 : 2 : N - 1) &= \text{Sam2}', \\ I'' (1 : 2 : N - 1, 2 : 2 : N) &= \text{Sam3}', \\ I'' (2 : 2 : N - 1, 2 : 2 : N) &= \text{Sam4}'. \end{aligned} \quad (8)$$

- (8) Construct the location map L of overflow and underflow pixels according to I'' , and for those few overflow/underflow pixels, change their values with random integer value in the range of (0,255).

These permutation times are encoded and encrypted as the encryption key. The location map and embedding strength parameters are compressed and encrypted as the data hiding key.

3.2. Data Extraction and Original Image Recovery

- (1) Revise the received image matrix according to location map to get image I'' .
- (2) Sample image I'' with size $N \times N$ into four submatrices:

$$\begin{aligned} \text{Sam1}_r &= I'' (1 : 2 : N - 1, 1 : 2 : N - 1), \\ \text{Sam2}_r &= I'' (2 : 2 : N, 1 : 2 : N - 1), \\ \text{Sam3}_r &= I'' (1 : 2 : N - 1, 2 : 2 : N), \\ \text{Sam4}_r &= I'' (2 : 2 : N - 1, 2 : 2 : N). \end{aligned} \quad (9)$$

- (3) Permute the four sampled submatrices with Arnold map with different permutation times $S1_T'$, $S2_T'$, $S3_T'$, and $S4_T'$, respectively:

$$\begin{aligned} \text{Sam1}_{r'} &= \text{Arnold}(\text{Sam1}_r, S1_T'), \\ \text{Sam2}_{r'} &= \text{Arnold}(\text{Sam2}_r, S2_T'), \\ \text{Sam3}_{r'} &= \text{Arnold}(\text{Sam3}_r, S3_T'), \\ \text{Sam4}_{r'} &= \text{Arnold}(\text{Sam4}_r, S4_T'), \end{aligned} \quad (10)$$

where $S1_T' = T - P_T - S1_T$, $S2_T' = T - P_T - S2_T$, $S3_T' = T - P_T - S3_T$, and $S4_T' = T - P_T - S4_T$. T is the permutation period of the sample images, and for the sample images with size 256×256 , $T = 192$, just as presented in Table 1.

- (4) Compose the permuted sampled submatrix to get the similar image with hidden data M_s :

$$\begin{aligned} M_s (1 : 2 : N - 1, 1 : 2 : N - 1) &= \text{Sam1}_{r'}, \\ M_s (2 : 2 : N, 1 : 2 : N - 1) &= \text{Sam2}_{r'}, \\ M_s (1 : 2 : N - 1, 2 : 2 : N) &= \text{Sam3}_{r'}, \\ M_s (2 : 2 : N - 1, 2 : 2 : N) &= \text{Sam4}_{r'}. \end{aligned} \quad (11)$$

Until now, the decryption process has been completed, and the similar image with hidden data is M_s . The hidden data can be extracted and the original image I can be reversibly recovered through the following steps.

- (5) Impose integer DWT on image M_s to get the four subbands LL, LH, HL, and HH.
- (6) Generate the histograms of the middle (LH, HL) and high (HH) frequency subbands and shift these histograms to extract the hidden data and reversibly recover the original subbands. The detailed steps are depicted in Section 2.2.2.

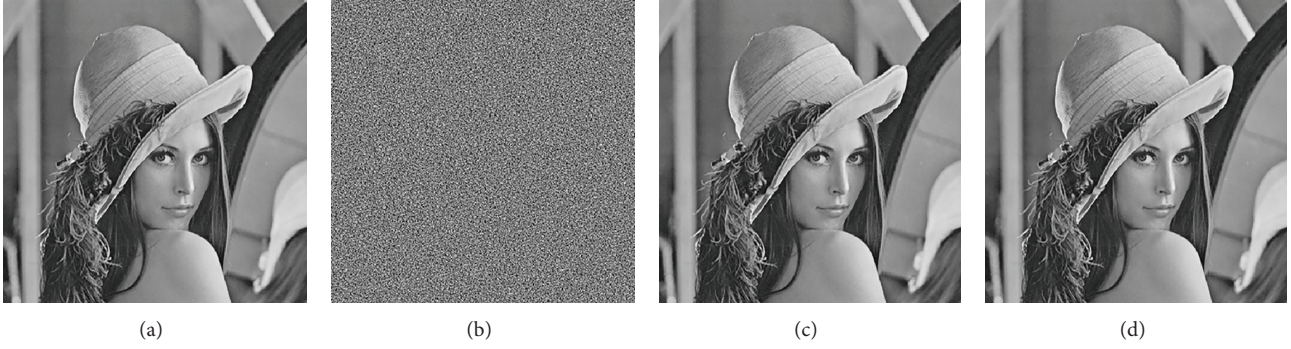


FIGURE 4: Images before and after disposing.

- (7) Impose inverse integer DWT with the coefficients of the subbands after histogram shifting to recover the original image.

4. Experimental Results and Analysis

To testify the efficiency and validity of the proposed scheme, images (with size 512×512) from Miscellaneous gray level images [21] and USC-SIPI image database [22] are selected for the experiments. Random binary bits are embedded into these images as the hidden data. All of these experiments are performed on the MATLAB 2012a platform running on a personal computer with CPU of AMD Phenom (tm) IIX4 810 Processor at 2.6 GHz, memory of 4 GB, and operating system of Windows 7 x64 Ultimate Edition.

In Figure 4, standard image “Lena” is adopted to demonstrate the feasibility of the proposed scheme. The subfigure (a) is the original Lena, (b) is encrypted image with embedding rate 0.0827 bpp, (c) is decrypted image with data embedded (PSNR = 50.7279), and (d) is the reversibly recovered image.

The hiding capacity with different embedding strength parameters, the corresponding PSNRs after data hiding, and the overhead data needed to dispose for the reversible recovery of the original image are presented in Tables 2 and 3.

As is seen in the tables, the embedding strength parameter q is 1, 2, 4, 8, 16, and 32, respectively. The embedding rates (ER) increase as the embedding strength parameters increase. In Table 2, images from USC-SIPI image database are tested. In Table 3, images from Miscellaneous gray level images are tested. It is easily seen that the overhead data for reversible recovery of the original image is rare and even zero for most of the test images. However, it is necessary especially when multilevel embedding is utilized. If the location map is transferred as a part of the payload, the pure embedding rates (PER) that exclude the overhead are also given in the table.

As can be seen in Tables 2 and 3, the embedding rates increase as the embedding strength parameter q increases. However, more overhead information is generated in accompany with the increase of embedding rate and the embedding strength parameter q . More distortion will be caused by the

greater amount of data hiding. Different images have different sensitivity to the embedding strength parameter q . Smooth images, such as “Airplane,” “Lena,” and “Boat,” are less sensitive to the parameter q than those complex images, such as “Baboon” and “Peppers.” That is because the histogram shifting based data hiding scheme imposed in the integer DWT domain depends largely on the similarity of adjacent pixels in the images.

Given the fix embedding rate, the plots between PSNR and embedding rate with different embedding strength parameters after decryption are demonstrated in Figure 5. The test images are selected from Miscellaneous gray level images database. The embedding strength of subfigure (a) is $q = 32$, (b) $q = 16$, (c) $q = 8$, (d) $q = 4$, (e) $q = 2$, and (f) $q = 1$, respectively.

The security of the proposed scheme is testified. As is known, there are similarities between adjacent pixels in natural images. One of the important things for the encryption of image is to destroy the correlation between two adjacent pixels. It can be calculated by the following formulas:

$$\begin{aligned}
 E(x) &= \frac{1}{N} \sum_{i=1}^N x_i, \\
 D(x) &= \frac{1}{N} \sum_{i=1}^N (x_i - E(x))^2, \\
 \text{cov}(x, y) &= \frac{1}{N} \sum_{i=1}^N E(x - E(x))(y - E(y)), \\
 r_{xy} &= \frac{\text{cov}(x, y)}{\sqrt{D(x)}\sqrt{D(y)}}.
 \end{aligned} \tag{12}$$

We randomly select 4096 pairs of two adjacent horizontal pixels, two adjacent vertical pixels, and two adjacent diagonally pixels in “Lena” image, respectively, for the demonstration. Figure 6 presents the correlation of adjacent pixels of image “Lena” before encryption and after encryption. The detail coefficients r_{xy} of selected images from Miscellaneous gray level images are presented in Table 4.

Obviously, the similarities have been thoroughly destroyed after encryption. Through the delicate design of the

TABLE 2: Embedding rate and PSNR of different images (USC-SIPI).

| Images | | $q = 1$ | $q = 2$ | $q = 4$ | $q = 8$ | $q = 16$ | $q = 32$ |
|----------|----------|---------|---------|---------|---------|----------|----------|
| Airplane | ER | 0.1206 | 0.3101 | 0.5044 | 0.6311 | 0.6971 | 0.7317 |
| | PSNR | 50.6962 | 46.8158 | 42.4294 | 38.5350 | 35.0636 | 32.1368 |
| | Overhead | 0 | 0 | 0 | 0 | 0 | 0 |
| | PER | 0.1206 | 0.3101 | 0.5044 | 0.6311 | 0.6971 | 0.7317 |
| Baboon | ER | 0.0251 | 0.0756 | 0.1707 | 0.3210 | 0.4960 | 0.6421 |
| | PSNR | 50.5884 | 46.0507 | 40.1912 | 34.7103 | 30.0639 | 26.4296 |
| | Overhead | 14 | 26 | 53 | 118 | 231 | 405 |
| | PER | 0.0250 | 0.0755 | 0.1705 | 0.3205 | 0.4951 | 0.6406 |
| Barbara | ER | 0.0656 | 0.1882 | 0.3602 | 0.5124 | 0.6113 | 0.6887 |
| | PSNR | 50.5979 | 46.3715 | 41.2511 | 36.5779 | 32.3365 | 28.7201 |
| | Overhead | 0 | 0 | 0 | 0 | 0 | 23 |
| | PER | 0.0656 | 0.1882 | 0.3602 | 0.5124 | 0.6113 | 0.6886 |
| Boat | ER | 0.0545 | 0.1577 | 0.3223 | 0.5162 | 0.6665 | 0.7287 |
| | PSNR | 50.5664 | 46.3085 | 41.0439 | 36.5185 | 33.1225 | 30.7374 |
| | Overhead | 3 | 10 | 19 | 38 | 109 | 296 |
| | PER | 0.0545 | 0.1576 | 0.3223 | 0.5161 | 0.6661 | 0.7276 |
| Lena | ER | 0.0827 | 0.2241 | 0.4310 | 0.6119 | 0.7032 | 0.7391 |
| | PSNR | 50.7257 | 46.5336 | 41.7441 | 37.8559 | 31.8851 | 32.7387 |
| | Overhead | 0 | 0 | 0 | 0 | 0 | 5 |
| | PER | 0.0827 | 0.2241 | 0.4310 | 0.6119 | 0.7032 | 0.7390 |
| Peppers | ER | 0.0632 | 0.1844 | 0.3743 | 0.5716 | 0.6961 | 0.7385 |
| | PSNR | 50.6754 | 46.4399 | 41.4909 | 37.5625 | 34.7859 | 32.8502 |
| | Overhead | 1 | 5 | 27 | 79 | 213 | 526 |
| | PER | 0.6632 | 0.1843 | 0.3472 | 0.5713 | 0.6953 | 0.7365 |

TABLE 3: Embedding rate and PSNR of different images (Miscellaneous gray level images).

| Images | | $q = 1$ | $q = 2$ | $q = 4$ | $q = 8$ | $q = 16$ | $q = 32$ |
|----------|----------|---------|---------|---------|---------|----------|----------|
| Airplane | ER | 0.1227 | 0.3110 | 0.5071 | 0.6342 | 0.6988 | 0.7325 |
| | PSNR | 50.6757 | 46.8180 | 42.4507 | 38.5947 | 35.1570 | 32.2588 |
| | Overhead | 0 | 0 | 0 | 0 | 0 | 0 |
| | PER | 0.1227 | 0.3110 | 0.5071 | 0.6342 | 0.6988 | 0.7325 |
| Baboon | ER | 0.0252 | 0.0754 | 0.1702 | 0.3211 | 0.4961 | 0.6421 |
| | PSNR | 50.5954 | 46.0404 | 40.1840 | 34.7072 | 30.0631 | 26.4306 |
| | Overhead | 19 | 36 | 65 | 138 | 237 | 433 |
| | PER | 0.0251 | 0.0753 | 0.1700 | 0.3206 | 0.4953 | 0.6405 |
| Barbara | ER | 0.0604 | 0.1704 | 0.3350 | 0.4932 | 0.6047 | 0.6852 |
| | PSNR | 50.6255 | 46.3325 | 41.0880 | 36.3009 | 32.0515 | 28.5263 |
| | Overhead | 0 | 0 | 0 | 0 | 19 | 219 |
| | PER | 0.0604 | 0.1704 | 0.3350 | 0.4932 | 0.6046 | 0.6484 |
| Boat | ER | 0.0857 | 0.2367 | 0.4296 | 0.5819 | 0.6791 | 0.7295 |
| | PSNR | 50.6239 | 46.5614 | 41.8009 | 37.5187 | 33.8471 | 31.0650 |
| | Overhead | 1 | 1 | 1 | 1 | 2 | 38 |
| | PER | 0.0857 | 0.2367 | 0.4296 | 0.5819 | 0.6791 | 0.7294 |
| Lena | ER | 0.0837 | 0.2239 | 0.4304 | 0.6117 | 0.7031 | 0.7390 |
| | PSNR | 50.7225 | 46.5535 | 41.7346 | 37.8432 | 34.8781 | 32.7376 |
| | Overhead | 0 | 0 | 0 | 0 | 0 | 5 |
| | PER | 0.0837 | 0.2239 | 0.4304 | 0.6117 | 0.7031 | 0.7390 |
| Peppers | ER | 0.0673 | 0.1924 | 0.3838 | 0.5833 | 0.6995 | 0.7379 |
| | PSNR | 50.6647 | 46.4457 | 41.5086 | 37.5847 | 34.8350 | 32.8801 |
| | Overhead | 30 | 92 | 183 | 545 | 1119 | 1826 |
| | PER | 0.0672 | 0.1921 | 0.3810 | 0.5812 | 0.6952 | 0.7371 |

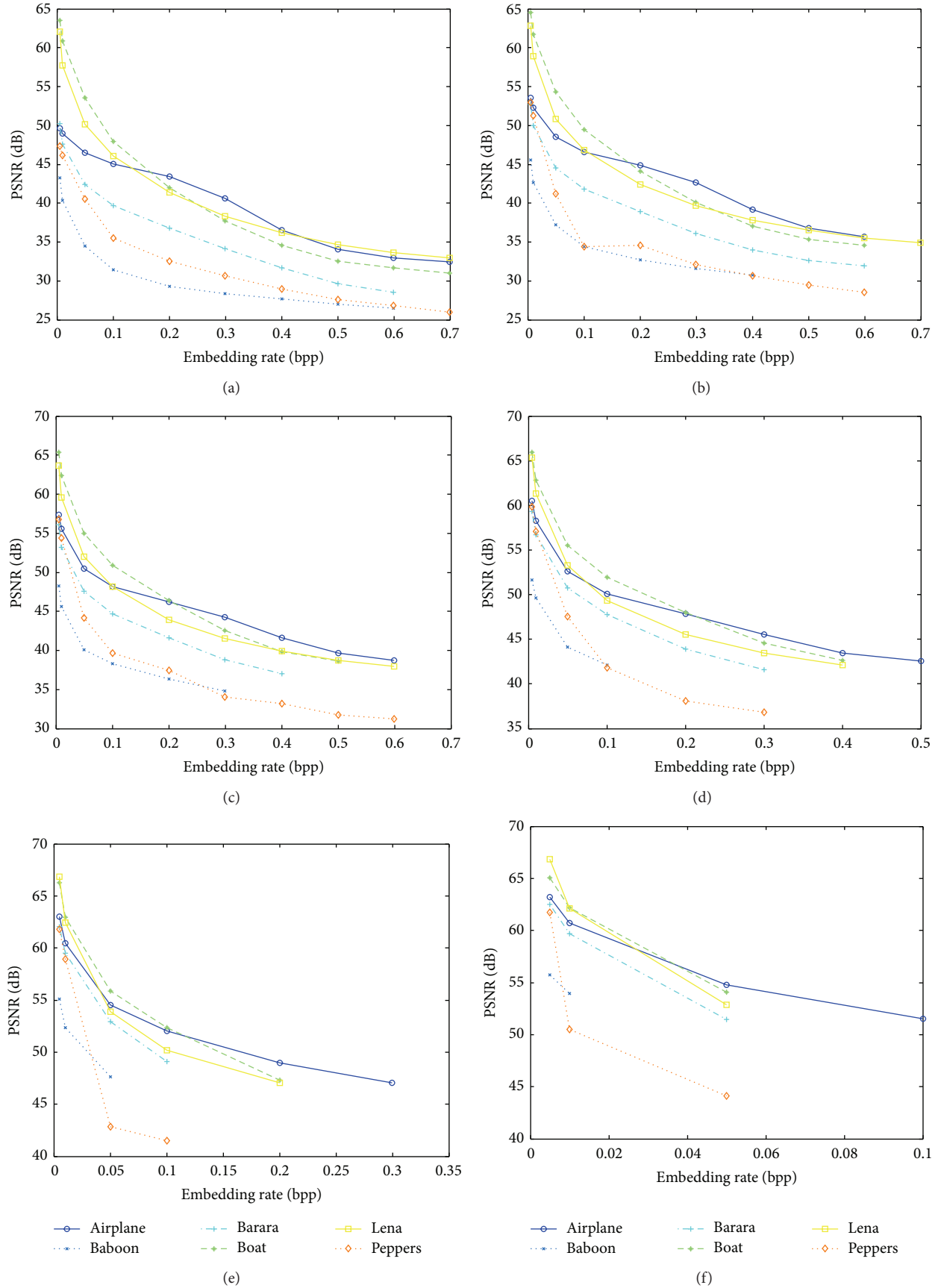


FIGURE 5: Embedding rate and PSNR of different images with different embedding strength parameter.

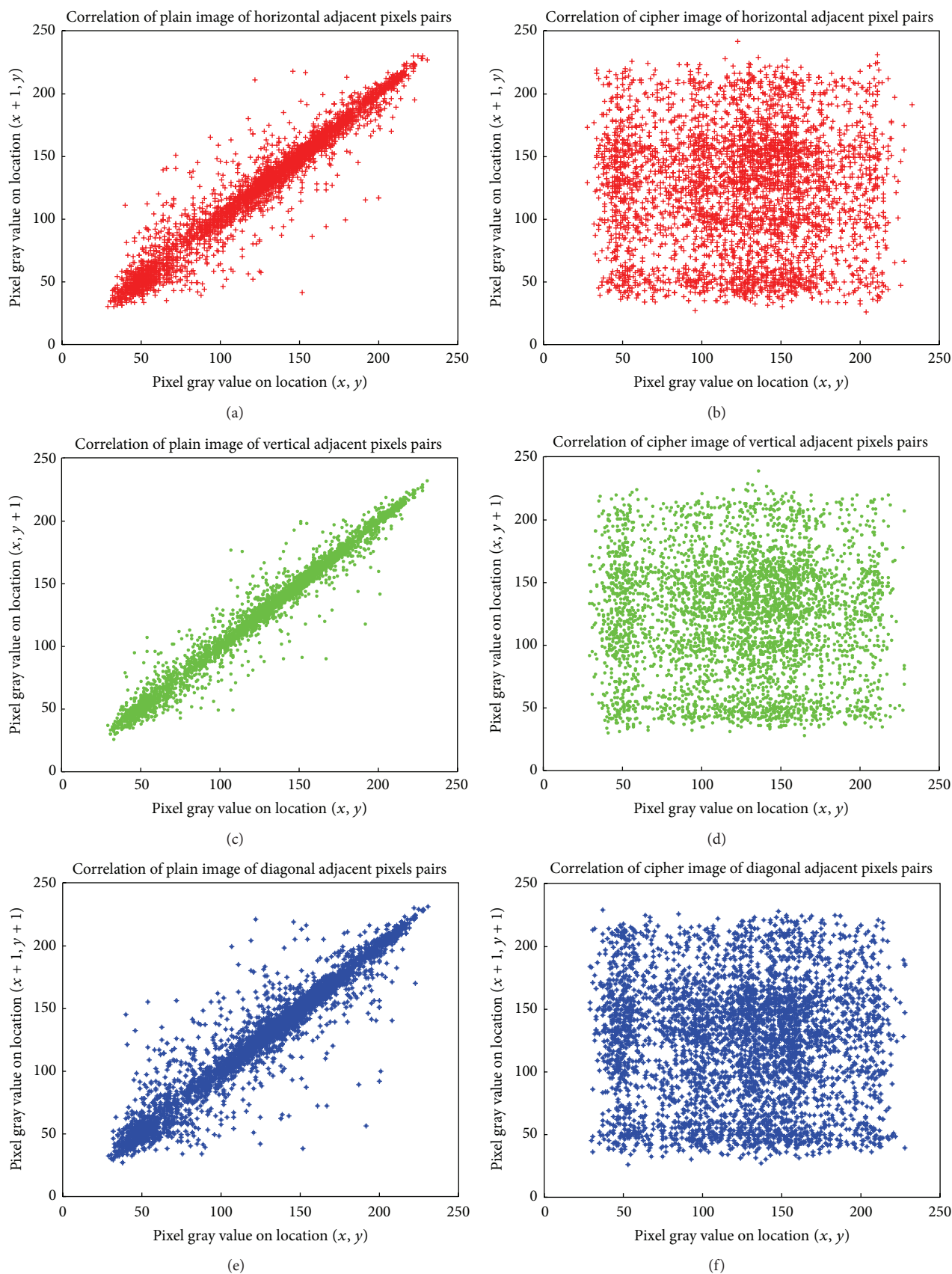


FIGURE 6: Correlations of two adjacent pixels in the plain image and in the cipher image of Lena.

TABLE 4: Coefficients of different images.

| Coefficients | Plain image | | | Cypher image | | |
|--------------|-------------|----------|----------|--------------|----------|----------|
| | Horizontal | Vertical | Diagonal | Horizontal | Vertical | Diagonal |
| Airplane | 0.9728 | 0.9648 | 0.9416 | -0.0092 | 0.0008 | -0.0054 |
| Baboon | 0.8407 | 0.7500 | 0.7160 | 0.0437 | 0.0101 | 0.0162 |
| Barbara | 0.9076 | 0.9648 | 0.8898 | -0.0142 | -0.0078 | -0.0222 |
| Boat | 0.9531 | 0.9827 | 0.9405 | 0.0320 | -0.0088 | -0.0090 |
| Lena | 0.9711 | 0.9851 | 0.9598 | -0.0022 | -0.0204 | 0.0215 |
| Peppers | 0.9779 | 0.9777 | 0.9665 | -0.0400 | 0.00474 | -0.0183 |

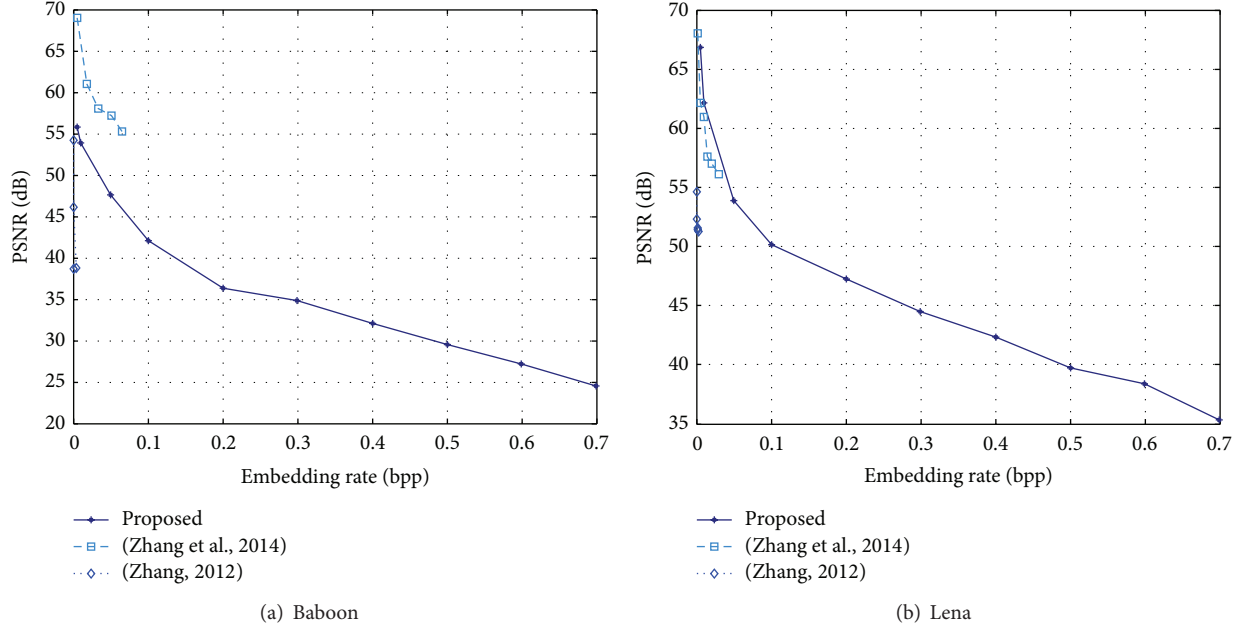


FIGURE 7: Comparisons of embedding rate and PSNR with existing schemes.

permutation in both transformed domain and spatial domain, the simple Arnold permutation can achieve nice encryption results. Besides, the encryption scheme proposed in this paper is efficient and timesaving due to the permutation only scheme. The stream cipher based encryption [13–15, 18] is more time-consuming because the encryptions are achieved by the bitwise exclusive OR operation or even the RC4 and AES encryption. Simulation results show that, for the images with size 512×512 , the average time for image encryption and data hiding is 4.3012 s and the average time for decryption and data extraction is 4.0013 s using the proposed scheme. If the same amount of data is embedded in the images with size 512×512 , the average encryption and data hiding time is more than 8.0332 s, and the average decryption and data extraction time is more than 7.8231 s for the encryption scheme with bitwise exclusive OR operation with hyperchaotic system.

The joint-RDH scheme proposed in [15] is applied in the medical images. It is not reversible. The joint-RDH scheme proposed in [14] increased the embedding capacity of the scheme proposed in [13]. However, their embedding capacity is rather low when the reversibility is achieved due to the design of the data hiding. At least a 8×8 block is needed for

TABLE 5: Embedding rate comparison with existing schemes.

| ER (bpp) | Baboon | Lena | Lake | Man | Splash |
|----------------|--------|--------|--------|--------|--------|
| Reference [14] | 0.0013 | 0.0069 | 0.0025 | 0.0024 | 0.0156 |
| Reference [13] | 0.0010 | 0.0039 | 0.0025 | 0.0025 | 0.0039 |
| Proposed | 0.6405 | 0.7390 | 0.6381 | 0.6659 | 0.7123 |

embedding one-bit information in their experiments. Therefore, the embedding rate is no more than $1/(8 \times 8)$ according to their experiments. Detailed comparisons of the embedding rate are presented in Table 5. Comparisons of the plot between PSNR and embedding rate with scheme proposed in [17] and in [18] are presented in Figure 7.

Obviously, the proposed scheme has been achieved better performances compared with exiting schemes. The reason why the PSNRs are higher at the same embedding rate is that, in the proposed scheme, data is hidden in the transformed domain through the difference histogram modification method. Such reversible data hiding schemes can achieve larger embedding capacity while keeping low distortion to the cover image. The encryption is achieved through the corporation of permutation in the integer DWT domain and in the

spatial domain. Moreover, the permutation in the integer DWT domain will not affect the data hiding. Due to the design of the existing joint encryption-RDH schemes [13, 14, 17, 18], a group of pixels is operated only for one bit data hiding. Their embedding rates are rather low as can be seen in Table 5 and Figure 7. The proposed scheme can provide a much larger embedding capacity.

5. Conclusion

A joint encryption-RDH scheme based on integer DWT and Arnold permutation is proposed. Data is hidden in the integer DWT domain with histogram modification based method, which guarantees the high embedding capacity and safety of data hiding. Although data is embedded in the DWT domain, reversible recovery of original images has been achieved through the integer transform. Different from those traditional encryption schemes such as bitwise XOR with random streams, AES, RC4, and so forth, the encryption scheme designed in this paper is based on Arnold permutation and thus is less time consuming and more efficient. Besides, permutation will not change the value of matrix, and thus data embedded will not be lost during the decryption process. Sufficient experiments demonstrate the efficiency and validity of the proposed scheme. Adaptive embedding can be adopted for better results. Multilevel integer DWT can be adopted for an even higher embedding capacity.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

Acknowledgment

The work described in this paper was supported by the Key program of National Science Fund of Tianjin, China (Grant no. 11JCZDJC16000).

References

- [1] J. Fridrich, M. Goljan, and R. Du, "Invertible authentication," in *Security and Watermarking of Multimedia Contents III*, pp. 197–208, January 2001.
- [2] A. M. Alattar, "Reversible watermark using difference expansion of triplets," in *Proceedings of International Conference on Image Processing (ICIP '03)*, vol. 1, pp. 501–504, September 2003.
- [3] A. M. Alattar, "Reversible watermark using difference expansion of quads," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 3, pp. 377–380, May 2004.
- [4] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1147–1156, 2004.
- [5] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 890–896, 2003.
- [6] C.-C. Lin, W.-L. Tai, and C.-C. Chang, "Multilevel reversible data hiding based on histogram modification of difference images," *Pattern Recognition*, vol. 41, no. 12, pp. 3582–3591, 2008.
- [7] W.-L. Tai, C.-M. Yeh, and C.-C. Chang, "Reversible data hiding based on histogram modification of pixel differences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 6, pp. 906–910, 2009.
- [8] P. Tsai, Y.-C. Hu, and H.-L. Yeh, "Reversible image hiding scheme using predictive coding and histogram shifting," *Signal Processing*, vol. 89, no. 6, pp. 1129–1143, 2009.
- [9] C.-H. Yang and M.-H. Tsai, "Improving histogram-based reversible data hiding by interleaving predictions," *IET Image Processing*, vol. 4, no. 4, pp. 223–234, 2010.
- [10] Z. Zhao, H. Luo, Z.-M. Lu, and J.-S. Pan, "Reversible data hiding based on multilevel histogram modification and sequential recovery," *AEU-International Journal of Electronics and Communications*, vol. 65, no. 10, pp. 814–826, 2011.
- [11] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 912–915, May 2003.
- [12] K.-S. Kim, M.-J. Lee, H.-Y. Lee, and H.-K. Lee, "Reversible data hiding exploiting spatial correlation between sub-sampled images," *Pattern Recognition*, vol. 42, no. 11, pp. 3083–3096, 2009.
- [13] X. Zhang, "Reversible data hiding in encrypted image," *IEEE Signal Processing Letters*, vol. 18, no. 4, pp. 255–258, 2011.
- [14] W. Hong, T.-S. Chen, and H.-Y. Wu, "An improved reversible data hiding in encrypted images using side match," *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 199–202, 2012.
- [15] D. Bouslimi, G. Coatrieux, and C. Roux, "A joint encryption/watermarking algorithm for verifying the reliability of medical images: application to echographic images," *Computer Methods and Programs in Biomedicine*, vol. 106, no. 1, pp. 47–54, 2012.
- [16] K. Ma, W. Zhang, X. Zhao, N. Yu, and F. Li, "Reversible data hiding in encrypted images by reserving room before encryption," *IEEE Transactions on Data Forensics and Security*, vol. 8, pp. 553–562, 2013.
- [17] W. Zhang, K. Ma, and N. Yu, "Reversibility improved data hiding in encrypted images," *Signal Processing*, vol. 94, pp. 118–127, 2014.
- [18] X. Zhang, "Separable reversible data hiding in encrypted image," *IEEE Transactions on Data Forensics and Security*, vol. 7, pp. 826–832, 2012.
- [19] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354–361, 2006.
- [20] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, Mathematical Physics Monograph Series, Addison-Wesley, 1968.
- [21] "Miscellaneous gray level images," <http://decsai.ugr.es/cvg/dbimagenes/g512.php>.
- [22] USC-SIPI Image Database, <http://sipi.usc.edu/database/database.php?volume=textures>.

Research Article

A Novel Algorithm of Quantum Random Walk in Server Traffic Control and Task Scheduling

Dong Yumin¹ and Xiao Shufen²

¹ Network Center, Qingdao Technological University, Qingdao 266033, China

² College of Automobile and Transportation, Qingdao Technological University, Qingdao 266033, China

Correspondence should be addressed to Dong Yumin; dym1188@163.com

Received 18 January 2014; Accepted 19 February 2014; Published 7 April 2014

Academic Editor: Feng Gao

Copyright © 2014 D. Yumin and X. Shufen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A quantum random walk optimization model and algorithm in network cluster server traffic control and task scheduling is proposed. In order to solve the problem of server load balancing, we research and discuss the distribution theory of energy field in quantum mechanics and apply it to data clustering. We introduce the method of random walk and illuminate what the quantum random walk is. Here, we mainly research the standard model of one-dimensional quantum random walk. For the data clustering problem of high dimensional space, we can decompose one m -dimensional quantum random walk into m one-dimensional quantum random walk. In the end of the paper, we compare the quantum random walk optimization method with GA (genetic algorithm), ACO (ant colony optimization), and SAA (simulated annealing algorithm). In the same time, we prove its validity and rationality by the experiment of analog and simulation.

1. Introduction

The server cluster technology may be connecting multiple independent servers, and, in the same time, it must provide services as a whole by a cluster. In the server cluster, how to solve the problem of server traffic control and task scheduling is very important.

In order to reduce the access time, optimize the overall performance and achieve parallel program in a high efficiency; the task request must be allocated to each on the server. So, load balancing mechanism is the core of cluster technologies.

In the literature [1], it expands the analogies employed on the development of quantum evolutionary algorithms by putting forward quantum-inspired Hadamard walks, called QHW. In order to solve combinatorial optimization problems, a quantum evolutionary algorithm, abbreviated HQEA, is proposed. From the results of the experiments carried out on the knapsack problem, HQEA performs noticeably better than a conventional genetic algorithm, in terms of convergence speed and accuracy. The literature [2] explores how a spectral technique suggested through coined quantum

walks can be used to differentiate between graphs which are cospectral as for standard matrix representations. This algorithm runs in polynomial time; it can differentiate many graphs for there is no subexponential time algorithm which is proven to be able to differentiate between them.

By the literature [3], they propose a quantum algorithm to evaluate formulas by an extended gate set, including two- and three-bit binary gates. This algorithm is more optimal on read, once formulas for that each gate's inputs are balanced in a certain sense. It describes a very compact triaxial instrument in the literature [4]. The triaxial instrument is based on a rhombic dodecahedral geometry that can accommodate three nonplanar ring light paths with orthogonal sensing axes. Component count can be substantially reduced by a discharge of layout to use a single cathode and two anodes running all three axes in balanced plasma currents. Two Monte Carlo-based approaches to assess parameter uncertainty with complex hydrologic models are considered in the literature [5]. The importance sampling has been carried out in the generalized likelihood uncertainty estimation framework by Beven and Binley. The metropolis algorithm is different from importance sampling which uses a random

walk that adapts to the real probability distribution describing parameter uncertainty.

Because existing search protocols for unstructured peer-to-peer systems to create huge burden on communications or cause long response time and result in unreliable performance. In the literature [6], in order to discover service providers, it reports that an important function of a peer-to-peer system is a distributed message relaying. They present an incentive mechanism which not only relieves the free-riding problem but also accomplishes good system efficiency in message relaying for peer discovery. The passed along message propagation process is promised rewards in the mechanism.

In the literature [7], it analyzes the discrete-time quantum walk by separating the quantum evolution equation into Markovian and its interference terms. Because of this separation, it is possible to show analytically which quadratic increase in the variation of the position of quantum walker with time is a direct aftermath of the coherence of the quantum evolution. As expected, the variation is shown to increase linearly with time, if the evolution is decoherent, as in the classical case. Moreover, it shows that the system has an evolving operator analogous to which of a resonant quantum kicked rotor. At the same time, the rotator can be described by evolution of the quantum walker.

Quantum random walks on a graph, which is analogous to classical stochastic walk, form the basis for many of the recent quantum algorithms that promise to obviously outperform existing classical random walk algorithms. A number of studies have been done on the many applications of quantum random walk to some important computing problems. There are two kinds of quantum random walk algorithms: continuous-time and discrete-time. It is reported that a quantum arithmetic is defined by a sequence of the operations that runs on an actual model of quantum computation in the literature [8]. It proposes quantum circuit designs for both kinds of random walk algorithms which operate on various graphs. It considers two important problems to which random walk arithmetic are applicable: the triangle finding problem and binary tree problem. Because of it a few research works that are related to quantum random walk circuit design on graphs exist; the circuit designs they present here are the first of their kind. At the same time, they also provide an estimate of the quantum cost of the circuits of quantum systems. And it is based on the number of execution cycles and quantum operations.

In the literature [9], the natural random walk causing Brownian motion occurs to be always biased in a very delicate way: emphasizing some possibilities by only approximative maximal uncertainty principle. It introduces a new method of stochastic model, and they use the merely maximizing entropy choice of transition probabilities.

Berry and Wang show numerically that a discrete time quantum random walk of two irrelevant particles is able to differentiate some nonisomorphic powerfully regular graphs from the same household in the literature [10]. They analytically show how it is possible for the walks to differentiate such graphs, while the continuous time quantum walks of two irrelevant particles cannot.

It is reported that the quantum walks are quantum mechanical theory analog of random walks in the literature [11]. By traversing the edges of a graph, a quantum “walker” progresses between initial and final states. They present a hybrid model for general quantum computing in which a quantum walker gets discrete steps of continuous evolution.

Effective server traffic control can extend the “capacity” of the server, and the task scheduling can improve system throughput. In early research methods of it, such as Min-Min algorithm, Max-Min algorithm, genetic algorithm (GA), round robin (RR), simulated annealing algorithm (SAA), dynamic feedback algorithm (DFA), and ant colony optimization algorithm (ACO). These arithmetics have some improvements in different degree on the task scheduling.

But these algorithms have this or that problem, such as local premature problem and divergence problem.

In order to overcome the instability above the algorithms, the quantum random walk algorithm is proposed, and it is proved better than above GA, ACO, and SAA by simulation experiments.

2. Quantum Random Walk

2.1. Random Walk. Random walk is a mathematical method to study the formation of trajectory by a random sequence of continuous; it is not only a means to study mathematics but also a basic tool in the natural sciences. Any stage of the random walk behavior is not limited to previous history of migration; the process is also called Markov process. Random walk can be simply described as follows.

Suppose in a straight line, there is a moving particle; it is at the origin to move left or right one unit of distance; the probability is p and $q = 1 - p$, respectively, each time the particle in accordance with the probability to move a unit distance to the left or right. Here we assume that the probability of the particle is equal to the left or right; that is, $p = q = 1/2$; random variable can be used to represent the probability σ_i ; its value is as follows:

$$\sigma_i = \begin{cases} -1, & p = \frac{1}{2} \\ 1, & q = \frac{1}{2}. \end{cases} \quad (1)$$

If the particles every moment in a straight line position constitute an independent identical distribution of random variables sequence, denoted $\sigma_1, \sigma_2, \dots, \sigma_t, \dots$ is a sequence of independent and identically distributed variables on meet $EX_1 = 0$, $\text{Var}(\sigma_1) = 1$. $S_t = \sum_{i=1}^t \sigma_i$ is its first n terms and $S = (S_t)_{t \in T}$ is called a random walk. After the particles move n steps, the probability of it being found in position m is

$$p_{0,m}^n = \left(\frac{n+m}{2} \right) q^{(n+m)/2} p^{(n-m)/2}. \quad (2)$$

Among them, $(n+m)/2$ take only integer, $(n+m)/2 \in 0, 1, \dots, n$; other cases were 0.

Although the classical random walk has a broad application, but compared with the quantum random walk, it feels much ashamed of its inferiority.

Quantum random walk is a quantum computing model proposed in recent years; scholars have also become increasingly interested in research.

2.2. Quantum Random Walk. For discrete quantum random walk, the system added an extra degree of freedom; some literature defines it as chirality that can build an adaptation to global local unitary process. This walk is also called quantum Hadamard walk. The only possible remained unchanged in the global process of unitary transformation is the only mobile operator between adjacent lattice points to the left or the right.

Significantly different between quantum Hadamard random walk and classical quantum is the interference, the diffusion rate of quantum walk square magnitude faster than classical square. Due to the existence of quantum superposition states, in quantum random walk, position of the particle from the probability distribution may be seen; particles may be in several locations simultaneously with different probability. Quantum random walk process is accomplished by a unitary matrix transformation [12].

2.2.1. One-Dimensional Quantum Random Walk. The random walk model sets up corresponding quantum algorithms and does quantum information processing. People commonly used coined quantum walks; it corresponds to Hilbert space which can be expressed as follows:

$$H = H_C \otimes H_V, \quad (3)$$

where $H_V = \text{span}\{|v_i\rangle\}_{i=1}^{|V|}$ is the random walk of grid space; it corresponds to a classic case of d -degree regular graphs $G(V, E)$, $H_C = C^d$, which is a coin flip operator space (coin space). The total unitary evolution matrix U is by the two independent parts, namely, flipping a coin and conditional replacement

$$U = S \cdot (C \otimes I). \quad (4)$$

The first step of the quantum walk is to perform a rotation operation C in coin space, equivalent to the classical random walk in a coin toss, through this operation to get a coin superposition state [13]. Then, the replacement operator S makes the particles by a coin to decide an edge vertex adjacent to move to the next. Starting from the initial state $|\psi(0)\rangle$, repeat the walk after n steps and obtain the probability distribution of each vertex as follows:

$$P(v, n) = {}_v\langle v | N r_c [|\psi(n)\rangle \langle \psi(n)|] | v \rangle_v. \quad (5)$$

Quantum random walks of a variety of ways common are one-dimensional linear walk, ring walk, hypercube walk, and so forth. For one-dimensional linear walk, $H_v = \text{span}\{|x\rangle : x \in Z\}$, $H_c = \text{span}\{|R\rangle, |L\rangle\}$, replacement operator applied to the base is expressed as

$$S|R, x\rangle = |R, x+1\rangle, S|L, x\rangle = |L, x-1\rangle. \quad (6)$$

Starting from the initial state $|\Phi_{in}\rangle = |\downarrow\rangle \otimes |0\rangle$, continuous action $U = S \cdot (C \otimes I)$; after each step, the distribution of every point is as follows:

$$\begin{aligned} |\Phi_{in}\rangle &\xrightarrow{U} \frac{1}{\sqrt{2}} (|\uparrow\rangle \otimes |1\rangle - |\downarrow\rangle \otimes |-1\rangle) \\ &\xrightarrow{U} \frac{1}{2} [|\uparrow\rangle \otimes |2\rangle - (|\uparrow\rangle - |\downarrow\rangle) \otimes |0\rangle + |\downarrow\rangle \otimes |-2\rangle] \\ &\xrightarrow{U} \frac{1}{2\sqrt{2}} [|\uparrow\rangle \otimes |3\rangle + |\downarrow\rangle \otimes |1\rangle + (|\uparrow\rangle - 2|\downarrow\rangle) \\ &\quad \otimes |-1\rangle - |\downarrow\rangle \otimes |-3\rangle]. \end{aligned} \quad (7)$$

Not only is the distribution situation different with the classical random walk, but one-dimensional linear walk is also higher and faster than the classical random walk in the diffusion velocity.

2.2.2. M -Dimensional Hypercube Quantum Random Walk. For the M -dimensional hypercube quantum random walk, it has 2^m vertices each vertex can be marked by an m binary string,

$H_v = \text{span}\{|\vec{x}\rangle : x \in [0, 2^n]\}$. Each vertex is m degrees; therefore, coin space $H_c = C^c$, using $|d\rangle$ to mark the coin space basis vectors, $d \in [0, n]$; it indicates the direction of the next step; each $|d\rangle$ corresponds to a M -dimensional vector $|\vec{e}_d\rangle$,

$$|\vec{e}_d\rangle = \left| \begin{matrix} 0 \cdots 0 1 0 \cdots 0 \\ (d-1) \quad (n-d) \end{matrix} \right\rangle. \quad (8)$$

$|\vec{e}_d\rangle$, the number of d bit is 1 and that of the other bits is 0. The replacement operator walk on the hypercube is applied to the base vectors and can be expressed as

$$S|d, \vec{x}\rangle = |d, \vec{x} \oplus \vec{e}_d\rangle. \quad (9)$$

It represents two points marked by M quantum bits; only when they have only one bit is not the same when they are connected directly via a side (e.g., 001101 and 011101 are communicating). There are many ways to select hypercube walking coin flip operators; however, the following form is usually taken to maintain certain symmetry:

$$C_{\alpha, \beta} = \begin{pmatrix} \alpha & \beta & \beta & \cdots & \beta & \beta \\ \beta & \alpha & \beta & \cdots & \beta & \beta \\ & & & \ddots & & \\ \beta & \beta & \beta & \cdots & \alpha & \beta \\ \beta & \beta & \beta & \cdots & \beta & \alpha \end{pmatrix}. \quad (10)$$

The operator of such forms has the characteristics that in all directions are a permutation invariant; it retains hypercube replacement invariance. Such a form is a commonly used Grover diffusion operator selection

$$G = -I_n + 2|s^c\rangle \langle s^c|. \quad (11)$$

Among them, $|s^c\rangle = 1/\sqrt{m} \sum_{d=1}^m |d\rangle$ are equally weighted superposition states in all directions. Grover operator is one

of the permutation invariant operators, which is the farthest with unit transformation I ; It will effectively mix all of the any given initial state into the superposition of them. The total random walk evolution operator can be expressed as

$$U = S \cdot (G \otimes I). \quad (12)$$

Define the Hamming distance, starting from any point to another point of the minimum number of edges experience (i.e., the required number of steps), with $d_h = 1$. Marking each vertex string the number of "1" is called Hamming weights, for example, Hamming weight 010 to 1. Starting from $00 \cdots 0$, any Hamming weighing the same point total can reach at the same number of steps. When the coins are formed to the symmetry type such as (10), it has the same probability starting from this point to reaching the point with the same weight. This allows us to put all Hamming weight of the same point "accumulation" to a point, thereby reducing the symmetry of the random walk; the walk on the hypercube is becoming walk in a straight line. It is noteworthy that this walk difference on the straight is not unbiased; it differs from the previously discussed one-dimensional linear walk. The number of total vertices after walking on the variable linear is $m + 1$; m is the hypercube dimension.

By putting a hypercube walk into walk on straight line, many problems can be resolved to simplify and get results. Moor and Russell found that when $T = \pi \cdot m/4$, random walk is a balanced distribution. Kempe through the research on the hitting time found hypercube quantum random walk to reach vertical angles of time relative to the classic case which is an exponential acceleration; this shows that the quantum random walk has the potential to make quantum algorithm acceleration.

3. The Model of Server Traffic Control and Task Scheduling

In cluster services, the task scheduling can be described as follows: N tasks need to be allocated to m nodes (these nodes are the servers) with different handling capacity; the goal is finding an optimization schedule to minimize the total completion time. The system model is shown as follows.

We suppose there are m nodes (or servers) and n tasks. Every task should be assigned to only one node. We use $P = \{p_1, p_2, \dots, p_m\}$ denoting the nodes (or servers), in this paper, where p_i denotes one of the nodes (or servers); $L = \{l_1, l_2, \dots, l_m\}$ expresses the current load, where l_i expresses the current load of node p_i . For instance, $l_i = 0$ means that the node (server) p_i has a current load of 0; in other words, the node is idle. Here n tasks are expressed by $X = \{x_1, x_2, \dots, x_n\}$, where x_j is one of the tasks. A $m \times n$ matrix is built between servers and tasks: W_{mn} , where W_{ij} is one of the elements. So, there are two states as follows:

$$W_{ij} = \begin{cases} 1 & \text{Task } x_i \text{ is assigned on node } p_j \\ 0 & \text{Task } x_i \text{ is not assigned on node } p_j \end{cases} \quad (13)$$

where, $i \in \{1, 2, \dots, m\}$, $j \in \{1, 2, \dots, n\}$.

We use t_{ij} to express the time of processing on one task; in other words, the time of task x_j processed on node p_i . The processing time is denoted as follows:

$$T_{ij} = \begin{cases} t_{ij} & \text{Task } x_i \text{ is processed on node } p_i \\ 0 & \text{Task } x_i \text{ is not processed on node } p_i \end{cases} \quad (14)$$

where, $i \in \{1, 2, \dots, m\}$, $j \in \{1, 2, \dots, n\}$.

It is not difficult to see that T_{ij} is also an $m \times n$ matrix.

Here, we define the optimal state occurring with these conditions: (a) the total system has a relative short time of processing; (b) the throughput of system is relatively larger in unite time. We can describe this state using the following equations:

$$Y_{\max} = \sum_i^m \omega(x_i, l_i, q_i) \quad (15)$$

$$\omega(x_i, l_i, q_i) = c_1 \left(\left(\sum_{i=1}^m \sum_{j=1}^n t_{ij} + \sum_{i=1}^m q_i t_i \right) - c_2 \sum_{i=1}^m l_i \right)^2,$$

where $X = \{x_1, x_2, \dots, x_n\}$ is the new task, $L = \{l_1, l_2, \dots, l_m\}$ is the current total load at the node, q_i is the length of ready queue at node p_i , t_i is the average processing time at node p_i , c_1, c_2 are constant, $\omega(x_i, l_i, q_i)$ is a function which can show the ability of node processing. The system is on the optimal running state, when the capacity of processing tasks (or task scheduling) reaches the maximal matching at one node.

4. The Method of Task Scheduling Based on Quantum Random Walk (QRW) Clustering Algorithm

In the paper, we mainly research the standard model of one-dimensional quantum random walk. For the data clustering problem of high dimensional space, we can decompose one m -dimensional quantum random walk into m one-dimensional quantum random walk.

4.1. Clustering Algorithm Based on One-Dimensional Quantum Random Walk, Referred to as Quantum Random Walk Clustering Algorithm (QRWA)

Step 1. Assume an unlabeled data set $X = \{X_0^1, X_0^2, \dots, X_0^n\}$, where each data point with m features.

Step 2. Each data point in the data set can be considered as a particle that transfers in the space according to the probability.

Step 3. Establish clustering algorithm based on the one-dimensional quantum random walk.

The clustering algorithm uses a distributed control strategy, that is, each data point of the data set only affected by its neighbor within the neighborhood. The neighbor of data points available k -nearest neighbor method or the method of default scope of R to determine and use $\Gamma_t(i)$ indicates the set of neighbors of a data point X_t^i in t time.

Step 4. Calculate the probability for each data point transfer to all neighbors in the neighborhood $p_t(i, j), j \in \Gamma_t(i)$; the formula is as follows:

$$p_t(i, j) = \begin{cases} \frac{a_t(i, j)}{\sum_{j \in \Gamma_t(i)} a_t(i, j)} & \text{if } j \in \Gamma_t(i) \\ 0 & \text{otherwise} \end{cases}$$

$$a_t(i, j) = \left(\left(\frac{\text{Deg}_t(j)}{\sum_{j \in \Gamma_t(i)} \text{Deg}_t(j)} \right) \times \left(\frac{\text{Deg}_0(j)}{\sum_{j \in \Gamma_t(i)} \text{Deg}_0(j)} \right) \right) \times ((d(X_t^i, X_t^j)) \times (d(X_0^i, X_0^j)))^{-1}. \quad (16)$$

Among them, $\text{Deg}_t(\cdot)$ and $\text{Deg}_0(\cdot)$, respectively, represent the degree of current and initial data points; similarly, $d(X_t^i, X_t^j)$ and $d(X_0^i, X_0^j)$, respectively, represent the current and initial data point distances between X^i and X^j .

Step 5. Find the maximum transition probability $p_t(i, h)$ and the neighbors of greatest probability of metastasis $X_t^h, h \in \Gamma_t(i), h \neq i$.

4.2. The Server Traffic Control Clustering Method of Quantum Random Walk. As previously mentioned, m -dimensional quantum random walk is decomposed into m one-dimensional quantum random walk. For each dimension, data points X_t^i can only move a step left or right, l_L or l_R . Therefore, the maximum transition probability $p_t(i, h)$ is mapped to the interval $\rho = f(p_t(i, h)) \in [0.5, 1]$, so, the probability of transfer in the opposite direction is $1 - \rho$. When $\rho = 0.5$, and probability of metastasis in both directions is equal, available aforementioned Hadamard transforms H as a coin matrix. However, in normal conditions, $\rho \neq 1 - \rho$, therefore, the coin matrix C is used in the algorithm is

$$C = \begin{pmatrix} \sqrt{\rho} & \sqrt{1-\rho} \\ \sqrt{1-\rho} & -\sqrt{\rho} \end{pmatrix}, \quad (17)$$

$$\rho = f(p_t(i, h)) = f\left(\max_{j \in \Gamma_t(i)} (p_t(i, j))\right).$$

It is easy to verify that the matrix C is unitary matrix meeting the reversibility requirements of quantum mechanics.

Since the quantum random walk clustering algorithm will use r consecutive transformation $U = S \cdot (C \otimes I)$ for initial state, thus, every time it changes left and right transfer step l_L and l_R to the original $1/2$, that is, l_L/r and l_R/r . Then, conditional operator S will press type structure:

$$S = |\uparrow\rangle\langle\uparrow| \otimes \sum_b \left| \frac{b+l_R}{r} \right\rangle \langle b+l|\downarrow\rangle\langle\downarrow|$$

$$\otimes \sum_b \left| \frac{b-l_L}{r} \right\rangle \langle b|. \quad (18)$$

As is known, in quantum mechanics, each one of superposition states can be seen as a position of particle and indicates the probability of finding the particle at this location. If repeatedly used U transforms for initial state, then the resulting superposition state $|\psi\rangle$ will contain more items; this increases possible appearing position of the particle, and this is not present in the classical random walk. It is these possible positions that increase the searching range of solution space and provide an opportunity for better results. To calculate the probability of multiple locations of particles and their appearance in the corresponding locations, a unitary operation is sufficient because of the quantum parallelism, but in the classical world, you need multiple operations to complete; it also reflects an aspect of quantum computing to accelerate the classical computing.

If the initial state of the particle is $|\psi_0\rangle = |\uparrow\rangle \otimes |0\rangle$, then, after applying $r = 2$ times transform U , get superposition state $|\psi\rangle$ which is

$$|\psi_0\rangle \xrightarrow{U} \sqrt{\rho} |\uparrow\rangle \otimes \left| \frac{l_R}{r} \right\rangle + \sqrt{1-\rho} |\downarrow\rangle \otimes \left| -\frac{l_L}{r} \right\rangle$$

$$\xrightarrow{U} \rho |\uparrow\rangle \otimes |l_R\rangle + \sqrt{\rho(1-\rho)} |\downarrow\rangle \otimes \left| \frac{(l_R-l_L)}{r} \right\rangle$$

$$+ (1-\rho) |\uparrow\rangle \otimes \left| \frac{(l_R-l_L)}{r} \right\rangle$$

$$- \sqrt{\rho(1-\rho)} |\downarrow\rangle \otimes |-l_L\rangle = |\psi\rangle. \quad (19)$$

From (19) particles can be found not only with probability ρ^2 and $\rho(1-\rho)$ at the same time appear in the probability of l_R and l_L appearing in another new position $(l_R-l_L)/r$ could be $(1-\rho)$. At this point, if projection measurement of the superposition state $|\psi\rangle$, it will collapse to one of these three positions according to the probability; then, the component of particle in the j dimension is updated with the following formula:

$$X_{t+1}^i(j) = \begin{cases} X_t^i(j) + l_R & \text{After the measurement, if the position} \\ & \text{of the particle at } l_R \\ X_t^i(j) + \frac{(l_R-l_L)}{r} & \text{After the measurement, if the position} \\ & \text{of the particle at } \frac{(l_R-l_L)}{r} \\ (l_R-l_L) - l_L & \text{After the measurement, if the position} \\ & \text{of the particle at } -l_L \end{cases}$$

$$I : l_R = l_L = (X_t^h(j) - X_t^i(j)), \quad j \in \{1, 2, \dots, m\}.$$

$$II : \begin{cases} l_R = \rho \times (X_t^h(j) - X_t^i(j)), & j \in \{1, 2, \dots, m\} \\ l_L = (1-\rho) \times (X_t^h(j) - X_t^i(j)), & j \in \{1, 2, \dots, m\}. \end{cases} \quad (20)$$

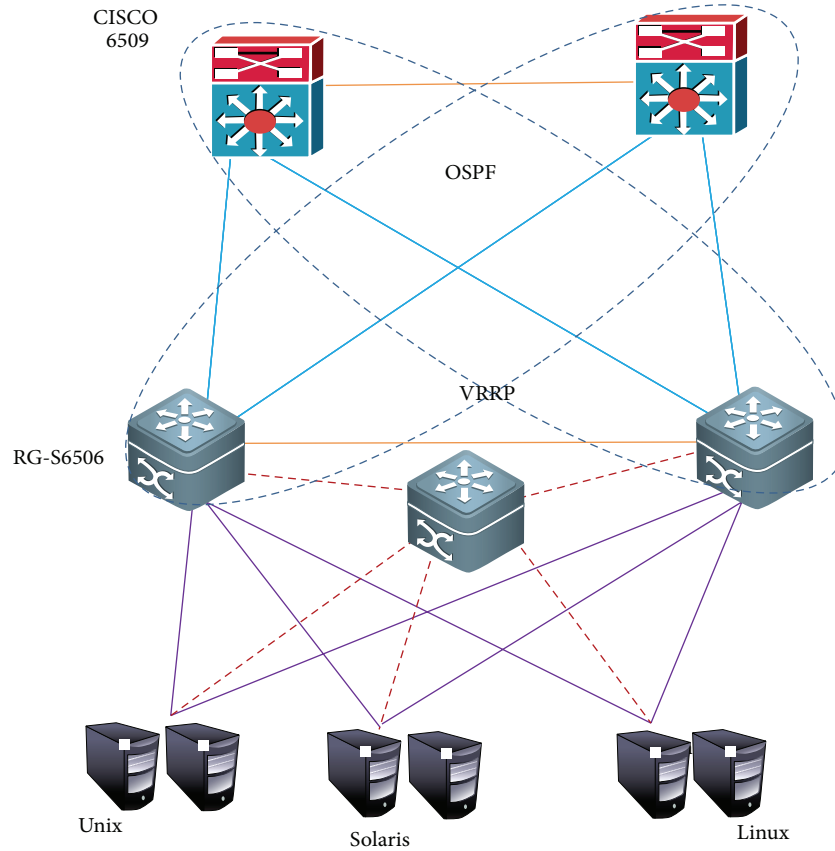


FIGURE 1: The topological structure of the network servers.

As data points are random walk in the space, its position and its nearest neighbor are constantly changing with time. Therefore, in the process of walking, the distance of the data points and the degree of it need to be recalculated. Repeat the entire process above until the sum of moving length of all the particles is less than some preset threshold ϵ . At this time, some separating section of the natural emergence in the space can be observed; each section corresponds a separate cluster.

5. Analog and Simulation Experiments

5.1. The Experimental Environment. In order to compare quantum random walk clustering algorithm (QRWA), genetic algorithm (GA), ant colony optimization (ACO), and simulated annealing algorithm (SAA), we select six servers as nodes. In the experiments, we select the number of task from 0 to 2500 (or 3000). We compare the results of these schemes by Matlab. The correlation parameters of selected servers for experiments are in Table 1.

In Table 1, OS represents operation system; NA represents network adapter; MS represents memory size; SM represents specifications and models.

The topological structure of the network servers is as shown in Figure 1.

5.2. Results. Figure 2 shows the system flow control rate of QRWA is better than GA, SAA, and ACO. And the more the

TABLE 1: Parameters of selected servers.

| SM | Model CPU | MS | NA | OS |
|----------------------------|---------------------------|--------|------------|---------|
| SUN SPARC Enterprise T5120 | UltraSPARC T2 1.2 G | 32.0 G | 2 * 1000 M | Solaris |
| IBM System p5520Q | POWER5+ 2.66 G | 32.0 G | 4 * 1000 M | Linux |
| HP rx4640 | intel Itanium 2. 1.5 G | 32.0 G | 2 * 1000 M | Unix |

task quantity is, the closer the flow control rate is. The task quantity is from 0 to 2500.

Figure 3 shows the server traffic of GA, SAA, ACO, and QRWA. That is to say, the QRWA is bigger than GA, SAA, and ACO.

The Figure 5 shows that the throughput of QRWA is bigger than ACO, GA, and SAA.

The Figure 6 shows that the throughput of QRWA is bigger than ACO, GA, and SAA.

Figures 4, 5, and 6 show the performances of QRWA are better than GA, SAA, and ACO.

From the results, it is clear that quantum random walk algorithm (QRWA) is better in server traffic control and task scheduling than genetic algorithm (GA), simulated

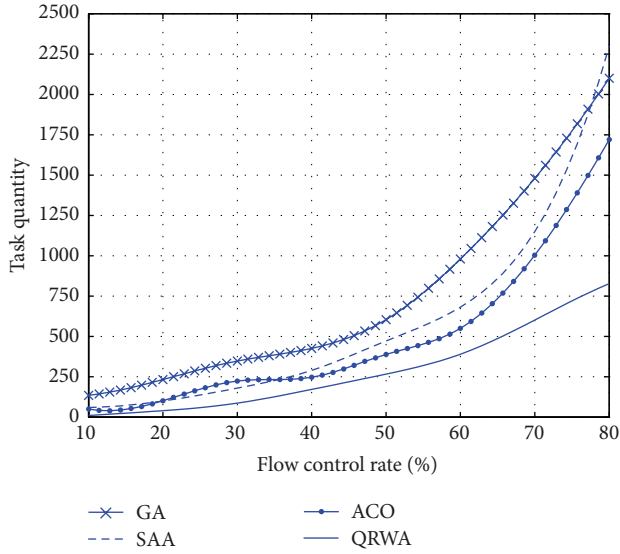


FIGURE 2: The flow control rate of SUN SPARC Enterprise T5120 in GA, SAA, ACO, and QRWA.

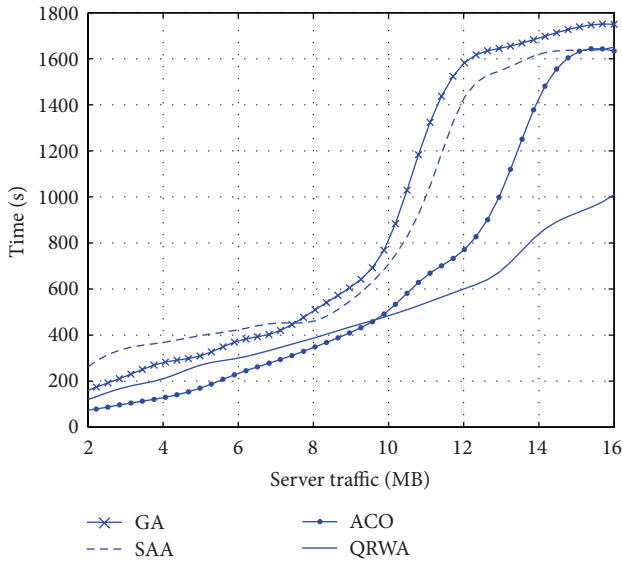


FIGURE 3: The server traffic of SUN SPARC Enterprise T5120 in GA, SAA, ACO, and QRWA.

annealing algorithm (SAA), and ant colony optimization (ACO). QRWA is more effective in task scheduling.

6. Conclusions

The paper gives a quantum random walk model and algorithm on server traffic control and task scheduling. We mainly research the standard model of one-dimensional quantum random walk. For the data clustering problem of high dimensional space, we can decompose one m -dimensional quantum random walk into m one-dimensional quantum random walk.

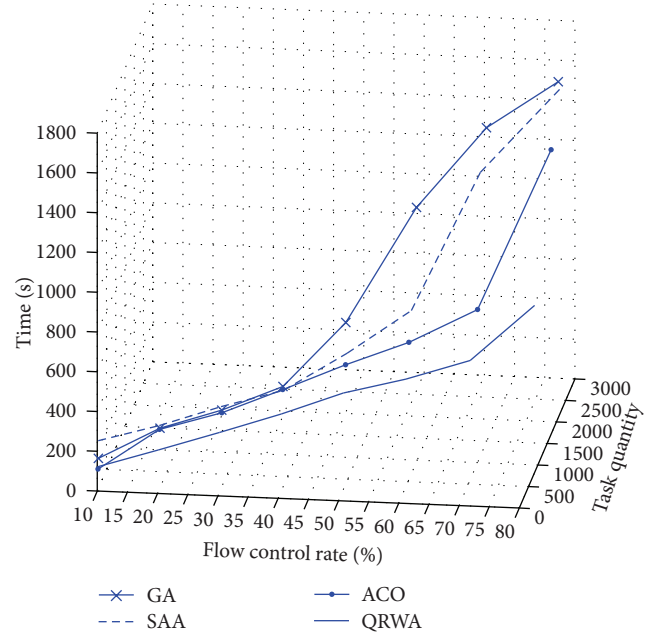


FIGURE 4: In SUN SPARC Enterprise T5120, the relationship of the flow control rate, task quantity and time. And the QRWA is better than GA, SAA, and ACO.

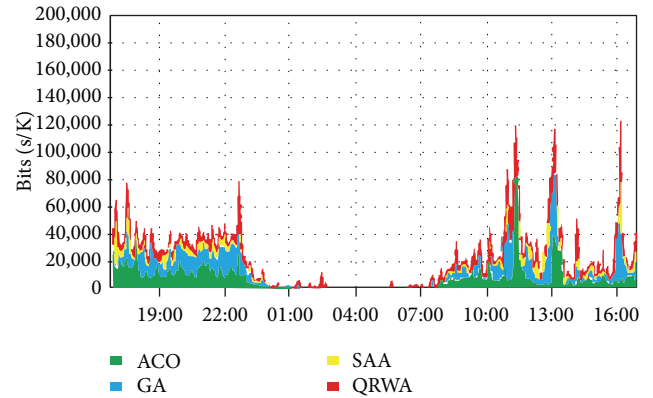


FIGURE 5: The throughput of HP rx4640 in ACO, GA, SAA, and QRWA during one day.

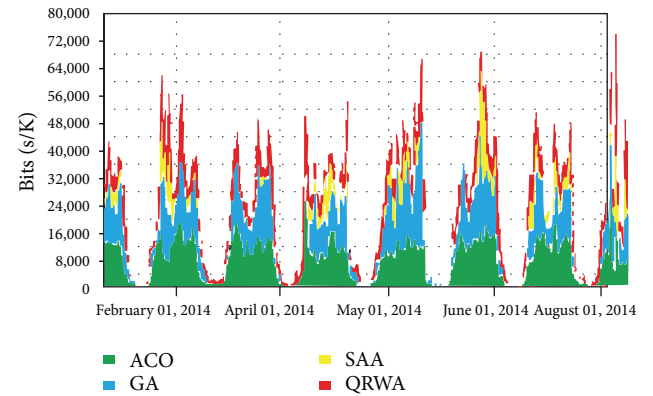


FIGURE 6: The throughput of IBM System p5520Q in ACO, GA, SAA, and QRWA during one week.

The simulation results demonstrate the effectiveness and superiority of QRWA.

The model and algorithm increases the throughput and efficiency of the system, and it had some merits than traditional model and arithmetics.

We will research the two directions in the future. The first one is the effects of noise on the scheme and model; the second one is the method of how to apply in the field of intelligence.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (61173056).

References

- [1] S. Koppaka and A. R. Hota, "Superior exploration-exploitation balance with quantum-inspired Hadamard walks," in *Proceedings of the 12th Annual Genetic and Evolutionary Computation Conference (GECCO '10)*, pp. 2093–2094, Companion Publication, July 2010.
- [2] D. Emms, S. Severini, R. C. Wilson, and E. R. Hancock, "Coined quantum walks lift the cospectrality of graphs and trees," *Pattern Recognition*, vol. 42, no. 9, pp. 1988–2002, 2009.
- [3] B. W. Reichardt and R. Špalek, "Span-Program-based quantum algorithm for evaluating formulas," in *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC '08)*, pp. 103–112, May 2008.
- [4] G. J. Martin, S. C. Gillespie, and C. H. Volk, "Litton 11 cm triaxial zero-lock gyro," in *Proceedings of the IEEE Position Location and Navigation Symposium (PLANS '96)*, pp. 45–55, April 1996.
- [5] G. Kuczera and E. Parent, "Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm," *Journal of Hydrology*, vol. 211, no. 1-4, pp. 69–85, 1998.
- [6] C. Li, B. Yu, and K. Sycara, "An incentive mechanism for message relaying in unstructured peer-to-peer systems," *Electronic Commerce Research and Applications*, vol. 8, no. 6, pp. 315–326, 2009.
- [7] A. Romanelli, A. C. S. Schifino, R. Siri, G. Abal, A. Auyuanet, and R. Donangelo, "Quantum random walk on the line as a Markovian process," *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 3-4, pp. 395–405, 2004.
- [8] A. Chakrabarti, C. Lin, and N. K. Jha, "Design of quantum circuits for random walk algorithms," in *Proceedings of the IEEE Computer Society Annual Symposium on VLSI (ISVLSI '12)*, pp. 135–140, 2012.
- [9] J. Duda, "From maximal entropy random walk to quantum thermodynamics," *Journal of Physics: Conference Series*, vol. 361, no. 1, Article ID 012039, 2012.
- [10] K. Rudinger, J. K. Gamble, E. Bach, M. Friesen, R. Joynt, and S. N. Coppersmith, "Comparing algorithms for graph isomorphism using discrete-and continuous-time quantum random walks," *Journal of Computational and Theoretical Nanoscience*, vol. 10, no. 7, pp. 1653–1661, 2013.
- [11] M. S. Underwood and D. L. Feder, "Universal quantum computation by discontinuous quantum walk," *Physical Review A: Atomic, Molecular, and Optical Physics*, vol. 82, no. 4, Article ID 042304, 2010.
- [12] L. Jun, *Investigations on Quantum Random Walk Search Algorithm*, 2006.
- [13] N. Shenvi, J. Kempe, and K. B. Whaley, "Quantum random-walk search algorithm," *Physical Review A: Atomic, Molecular, and Optical Physics*, vol. 67, no. 5, Article ID 052307, 2003.

Research Article

A New Method of Image Denoising for Underground Coal Mine Based on the Visual Characteristics

Gang Hua¹ and Daihong Jiang^{1,2}

¹ School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221008, China

² School of Information and Electronic Engineering, Xuzhou Institute of Technology, Xuzhou, Jiangsu 221008, China

Correspondence should be addressed to Gang Hua; ghua3323@163.com

Received 15 January 2014; Accepted 12 March 2014; Published 6 April 2014

Academic Editor: Feng Gao

Copyright © 2014 G. Hua and D. Jiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Affected by special underground circumstances of coal mine, the image clarity of most images captured in the mine is not very high, and a large amount of image noise is mingled with the images, which brings further downhole images processing many difficulties. Traditional image denoising method easily leads to blurred images, and the denoising effect is not very satisfactory. Aimed at the image characteristics of low image illumination and large amount of noise and based on the characteristics of color detail blindness and simultaneous contrast of human visual perception, this paper proposes a new method for image denoising based on visual characteristics. The method uses CIELab uniform color space to dynamically and adaptively decide the filter weights, thereby reducing the damage to the image contour edges and other details, so that the denoised image can have a higher clarity. Experimental results show that this method has a brilliant denoising effect and can significantly improve the subjective and objective picture quality of downhole images.

1. Introduction

The application environment in the coal industry is always special, and downhole images are always mingled with large amount of image noise interfered by complex underground environment, mechanical vibration, and dust noise. This brings many difficulties for the subsequent processing of the image. Therefore, the denoising process on images captured and transmitted from downhole becomes very necessary so as to provide better image quality and better followup processing.

There are many ways in terms of denoising, such as the mean denoising, median denoising, and wavelet denoising. A representative study is from Narendra, who raised the row-column decomposition median filtering algorithm [1]; moreover, He et al. have put forward a multimedian filter algorithm [2] which can effectively remove the image impulse noise; additionally, Darsow and Olsen have proposed three denoising methods based on phases of wavelet: local phase variance threshold, edge tracking, and scale phase fluctuation threshold method [3]. In order to overcome the weakness that wavelet transform can only carry out point singularity

detection, Minh N. Do and Martin Vetterli proposed contourlet transform in 2002 [4]. With the gradual deepening of the various branches of mathematics in the theory and applications, great progress in image denoising technology has been achieved in terms of mathematical morphology, partial differential equations, genetic algorithms, information theory, and so forth, producing a number of new denoising algorithms [5–8], including denoising algorithm based on mathematical morphology [9–12], denoising algorithm based on fuzzy theory [13, 14], denoising algorithm based on genetic algorithms [15], neural network-based denoising algorithm [16], and denoising algorithm based on information entropy.

Although, with the maturity and improvement of the various theories, image denoising methods have gained a lot of progress; these methods have their respective advantages, disadvantages, and application areas. For example, the mean denoising is suitable for removing grain noise in images, but the images always easily become blur because this method is too average; median denoising is good for removing impulse noise in the image, but the denoising effect is not very ideal when the noise area inside the window is too large; Wiener filter is suitable for removing the white

noise in the image; however, the calculation amount is too large; wavelet denoising has a wide range of algorithms and superior performances, but the realization is very difficult and complex. Moreover, the effect of the above algorithms will not be satisfactory when an image is simultaneously interfered by a variety of noise; in other words, while the image smoothens, the quality declines. Besides, denoising research on specific areas with complex environment is also very rare. The aforementioned are the current research status of image denoising; therefore, with the combination of noise characteristics of the coal mine, finding a method that can preserve the image detail and textural features while at the same time reducing image noise has become the research goal of this paper.

2. CIELab Color Space

CIELab color space is defined by the International Commission on Illumination (CIE) in 1976, and it is currently one of the most uniform color spaces with a set data of L , a , and b representing one color, and one Lab values group formed one corresponding relationship with one color. In the set, L indicates the luminance value; a and b are the chromaticity coordinates. Value a indicates the color change direction of red-green, $+a$ indicates the change towards the red direction, and $-a$ shows the change in the direction of green. b shows the change in the yellow-blue direction, $+b$ shows the change in the direction of yellow, and $-b$ shows the change in the blue direction. As shown in Figure 1, a represents axis of red and green and b represents axis of yellow and blue. Their values range from 0 to 10. $a = b = 0$ means colorlessness, and L represents scale factor from black to white.

3. Interconversion of RGB and CIELab Color Space

The interconversion of RGB and CIELab color space needs to convert RGB to CIEXYZ color space first and then to CIELab color space [5].

The conversion formula of RGB color space to XYZ color space is

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{pmatrix} 0.4303 & 0.3416 & 0.1784 \\ 0.2219 & 0.7068 & 0.0713 \\ 0.0202 & 0.1296 & 0.9393 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (1)$$

Then, convert XYZ color space to LAB color space, and the conversion formula is

$$\begin{aligned} a &= 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right], \\ b &= 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right], \\ f(x) &= \begin{cases} x^{1/3} & \frac{Y}{Y_n} > 0.008856, \\ 7.787 \times x + \frac{16}{116} & \frac{Y}{Y_n} \leq 0.008856, \end{cases} \end{aligned} \quad (2)$$

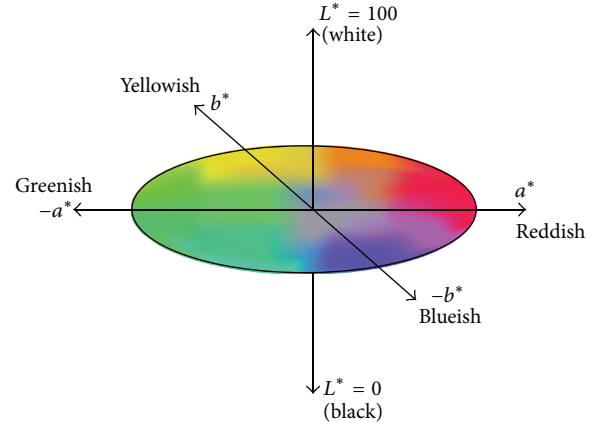


FIGURE 1: CIELab color space.

where $L = f(x)$, $X_n = 95.04$, $Y_n = 100.00$, and $Z_n = 108.89$ are white tristimulus values of D_{65} , the CIE standard illuminant; X , Y , and Z are coordinate value of CIEXYZ space.

The computational formula of aberration ΔE_{ab}^* between two colors in CIEXYZ space is as follows:

$$\Delta E_{ab} = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}, \quad (3)$$

where

$$\begin{aligned} \Delta L^* &= L_1^* - L_2^*, \\ \Delta a^* &= a_1^* - a_2^*, \\ \Delta b^* &= b_1^* - b_2^*. \end{aligned} \quad (4)$$

4. New Adaptive Image Denoising Method Based on Visual Characteristic

The classic image denoising methods are done in the RGB color space, while RGB color space is a nonuniform color space, and it does not take into account important information such as the image brightness and chroma. Due to the fact that images captured from downhole are affected by low light or uneven illumination, it is very hard to reach satisfactory denoising effect by adopting the classic image denoising method. In order to improve the image denoising performance, a more uniform CIELab color space is needed. In this color space, human visual sensitivity to the color differences of different wavelengths is not the same; the relevant data are shown in Table 1. Aimed at the special downhole circumstances and based on CIELab uniform color space, the characteristics of color detail blindness, and simultaneous contrast of human visual perception, this paper presents a new image denoising method to improve the subjective and objective image quality.

4.1. Algorithm Thought. In uniform color space, the chromatism value of two human-eye distinguishable colors is equal; that is, when the chromatism is smaller than a certain

TABLE 1: Value of chromatism and sensitivity level to the color differences.

| Value of chromatism | Sensitivity level to the color differences |
|---------------------|---|
| 0.0~0.50 | (Tiny chromatic aberration) trivial feeling |
| 0.5~1.51 | (Small chromatic aberration) slight feeling |
| 1.5~3 | (Lesser chromatic aberration) noticeable feeling |
| 3~6 | (Larger chromatic aberration) appreciable feeling |
| Above 6 | (Large chromatic aberration) strong feeling |

threshold, human eyes consider them as the same color; but when the chromatism is greater than a threshold, human eye will be able to distinguish the two different colors. In CIELab color space, the value of this threshold is generally 3. In view of this, the paper will divide image noise into two types: flat region noise and nonflat region noise. The algorithm carries out different processing self-adaptively according to the different region of the pixel.

In a region, let x_{ij} represent the polluted pixel value of point (i, j) , and then let the CIELab median of the pixel in the area with x_{ij} being the center point and neighborhood being $(2N + 1) \times (2N + 1)$ be

$$y_{ij} = \text{median} = \{x_{i-N, j-N}, \dots, x_{ij}, \dots, x_{i+N, j+N}\}. \quad (5)$$

Let the CIELab chromatic aberration of current pixel point and center point be $d = |x^{ij} - y^{ij}|$; preset a threshold t , and compare it with d :

$$\alpha_{ij} = \begin{cases} 1 & d > t, \\ 0 & \text{else.} \end{cases} \quad (6)$$

If α_{ij} of all pixels within the neighborhood of $(2N + 1) \times (2N + 1)$ are 1, then, it is called flat region noise; the color of flat region noise point can be affected by the rest of pixels within the same neighborhood, and convolution denoising can be conducted by using traditional Gaussian filter template. If there are pixels within the neighborhood of $(2N + 1) \times (2N + 1)$ whose α_{ij} are 0; then, this area is nonflat region. Noise in nonflat region is not entirely affected by other pixels in the neighborhood and is only related to pixels in the same area; that is, the noise is only related to pixels whose CIELab chromatic aberration is less than the relevant threshold; if it is greater than the threshold, then it is considered that the contribution of the pixel color value to the center point remains conforming to Gaussian distribution; otherwise, the pixel does not have any contribution to the center point. Let the convolution weights be 0. Thus, the pixel color will be more consistent within the same region, achieving the purpose of preserving image detail.

4.2. Gaussian Filter. Gaussian filter [7] is the linear smoothing filter which determines weights according to the shape of the Gaussian function. One-dimensional Gaussian function is

$$g(x) = e^{-x^2/2\sigma^2}, \quad (7)$$

wherein σ determines the width of the Gaussian filter. The greater σ becomes, the wider the band of the Gaussian filter will be and the better the smoothness will be. Moreover, a compromise can be obtained by adjusting the smoothness degree parameter σ , when the image features are too vague and when excessive unwelcome break variables are caused by noise and texture in smooth images. For image processing, two-dimensional Gaussian function is commonly used as smoothing filter. The function expression is

$$g[i, j] = e^{-(i^2+j^2)/2\sigma^2}. \quad (8)$$

The convolution and denoising formula of the input image $f[i, j]$ via Gaussian filter is

$$\begin{aligned} g[i, j] * f[i, j] &= \sum_{k=0}^{m-1} \sum_{l=0}^{n-1} g[k, l] f[i-k, j-l] \\ &= \sum_{k=0}^{m-1} \sum_{l=0}^{n-1} e^{-(k^2+l^2)/2\sigma^2} f[i-k, j-l] \\ &= \sum_{k=0}^{m-1} e^{-k^2/2\sigma^2} \left\{ \sum_{l=0}^{n-1} e^{-l^2/2\sigma^2} f[i-k, j-l] \right\}. \end{aligned} \quad (9)$$

In order to reduce the time complexity of Gaussian filter convolution calculation, the two-dimensional Gaussian function can be converted to the combination of two one-dimensional convolution templates [8]. First, convolve the input image $f[i, j]$ and the Gaussian template in the horizontal direction; set a temporary array; store the result in the temporary array; then convolve the image and the Gaussian template in the horizontal direction and transpose the result; the final smooth image can be obtained.

Figure 2 is a schematic diagram of the separability of the Gaussian function convolution. This method is completed through the combination of two horizontal convolution templates. First, convolve the input image $f[i, j]$ and the Gaussian function in the horizontal direction, and then transpose and store the result in the temporary array; after that, take the temporary array as the input, and carry out the convolution with the same Gaussian function to realize the purpose of replacing the vertical convolution with the horizontal convolution. Transpose the output information after the second convolution, and then the final smoothed output image can be obtained. Since the separability of the Gaussian function is high, Gaussian filter can be effectively achieved. Two-dimensional Gaussian function can be carried out in two steps. First, convolve the image with the one-dimensional Gaussian function; then, convolve the result with the same one-dimensional Gaussian function perpendicular to the first result. Therefore, the calculation amount of two-dimensional Gaussian filter presents a linear growth along with the width of the filter template rather than a square growth.

4.3. Algorithm Flow. Define the size of denoising template as $\delta \times \delta$ and the distinguishable color threshold as T ; traverse any pixel (r, g, b) in the image; the specific realization steps of this algorithm are as follows.

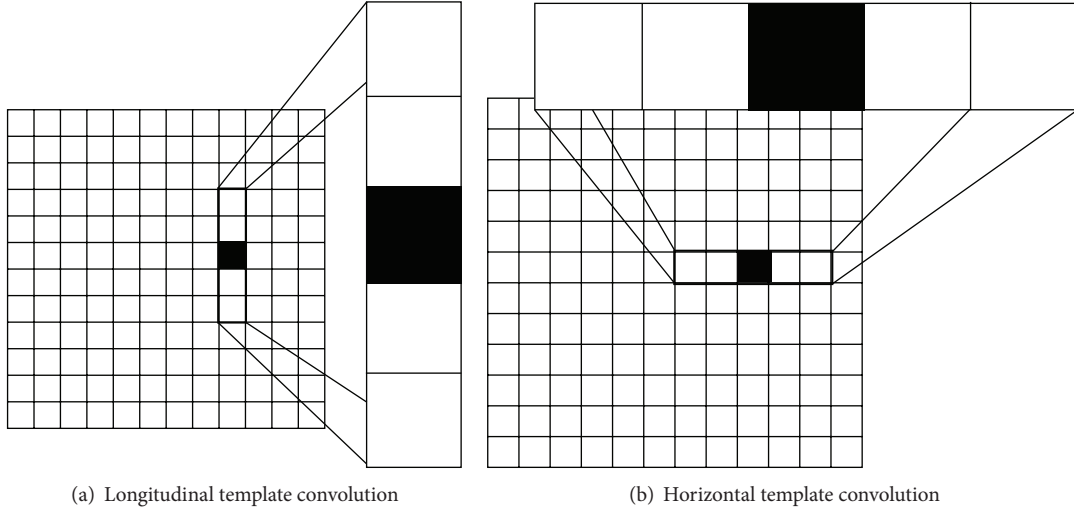


FIGURE 2: Schematic diagram of the separability of Gaussian function convolution.

- (1) Input the pending image f and initialize the Gaussian convolution template whose size is $\delta \times \delta$.
- (2) Use the chromatic aberration computational formula of CIELab space color to calculate the value of chromatism between the center pixel and pixels within neighborhood δ ; collect number n which stands for the value of chromatism that is greater than the threshold T .
- (3) If $n = \delta \times \delta - 1$, then it indicates that except the center pixel all the values of chromatism between any two pixels in the region are greater than the threshold T . This is flat region noise; convolution and denoising can be reached via traditional Gaussian filter template. Turn to Step (4).
- (4) If $n \neq \delta \times \delta - 1$, $n \neq 0$, and if the value of chromatism between center pixel and pixel (i, j) in the region is greater than threshold T , then this is nonflat region noise. Set the value of pixel (i, j) in Gaussian convolution template to be 0. Turn to Step (5).
- (5) If none of the above is satisfied, no operation is required; keep the original value.
- (6) Process the next pixel and turn to Step (2).
- (7) Judge whether all the pixels are processed; if yes, end the algorithm; otherwise, turn to Step (6) and continue the processing.

The algorithm flow is as shown in Figure 3.

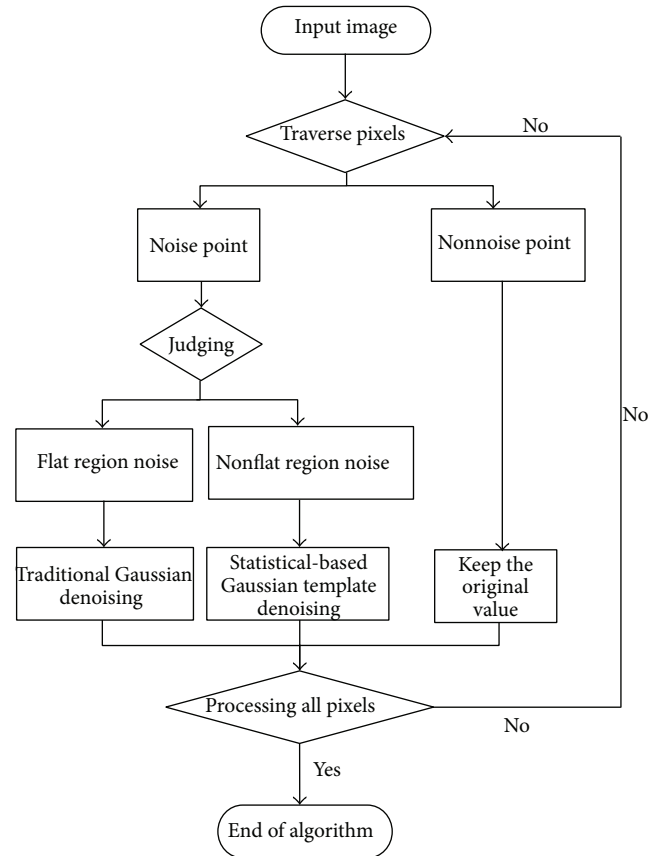


FIGURE 3: Algorithm flow.

5. Experimental Performance and Comparative Analysis

A noisy downhole image was selected and denoising processing was carried out by, respectively, using methods of mean denoising, median denoising, hybrid denoising, traditional

Gaussian denoising, and the method proposed in this paper. Image RMSE (root mean square error) and PSNR (peak signal to noise ratio) were calculated, so as to evaluate the performance of different algorithms.



FIGURE 4: Effect contrast of different denoising methods.

RMSE and PSNR are defined as follows [17, 18]:

$$\text{RMSE} = \frac{1}{MN} \sum_{j=1}^N \sum_{i=1}^M |f(x, y) - g(x, y)|^2, \quad (10)$$

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2 \times M \times N}{\sum_{j=1}^N \sum_{i=1}^M (f(i, j) - g(i, j))^2} \right),$$

wherein g indicates source image, f stands for images which goes through scales addressing first and then same multiple

processing second with the use of relevant algorithm, and M and N represent the length and width of the image. RMSE reflects the approximate extent of the scaled image to the source image; the smaller the RMSE is, the more approximate the scaled image is to the source image; PSNR reflects the image magnification effect. The higher the PSNR is, the clearer the scaled image will be.

The experimental hardware environment is Pentium 4 CPU (2.80 GHz), memory capacity 1.5 GB, and resolution 1024×768 ; the software environment is Microsoft Windows XP Professional SP3 operating systems. The effect contrast of

TABLE 2: Evaluation index of denoising effect of different denoising algorithms.

| Denoising method | RMSE | PSNR |
|--------------------------------|-----------|----------|
| Mean denoising | 307.14593 | 21.03142 |
| Median denoising | 259.65736 | 32.73148 |
| Hybrid denoising | 242.47657 | 33.45765 |
| Traditional Gaussian denoising | 210.32459 | 34.87167 |
| Paper-proposed denoising | 190.10174 | 36.91472 |

different denoising methods is as shown in Figure 4, and the comparison of denoising experimental data is as shown in Table 2.

It can be seen from Table 2 that RMSE of mean denoising is the maximum, and that of the PSNR is minimum, which means the denoising effect of this method is the most unsatisfactory. RMSE of median and hybrid denoising method is greatly lower than that of mean denoising, and PSNR has also been improved. Compared with traditional Gaussian denoising and other methods, the algorithm proposed by this paper has further improved the objective quality of the image. Moreover, it can also be seen from Figure 3 that the contours of the image processed by mean denoising are vague, and the image also contains a lot of noise; what is worse, there is a serious loss of image detail information; the performance of median and hybrid denoising is a little bit better, but the processed image still contains residual noise; traditional Gaussian denoising method did not manage to eliminate the obvious noise within the visual range, and the effect is not very ideal; the algorithm proposed in this paper successfully improved the peak signal to noise ratio and at the same time managed to reserve image detail features and texture changing features well; more importantly, the processed image has higher resolution and better visual effects.

6. Conclusions

Denoising processing on underground images has been conducted in this paper based on CIELab color space and CIELab chromatism aberration computational formula. The paper has proposed that weight should be determined by visual chromatism aberration, which makes up the insufficiency of letting space coordinates distance decide weight in RGB space; in the meantime, the characteristic that CIELab color space is more uniform in visual perception is used, taking fully into account the luminance and chrominance information of the processed image, making the denoised images and human vision maintain better correlation. The experimental results have demonstrated the effectiveness of the algorithm from aspects of the subjective quality and the objective quality. This method is very helpful to further processing and application of downhole images.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by advancing projects of the industrialization of scientific research achievements in Universities of Jiangsu Province (Grant JHB2012-36) and the science and technology fund of the Ministry of Housing and Urban-Rural Development of China (Grant 2013-K2-5).

References

- [1] P. M. Narendra, "A separable median filter for image noise smoothing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3, no. 1, pp. 20–29, 1981.
- [2] J.-M. He, Q. Li, and Y. Ming, "Realization and performance test of improved median filter," *Computer Systems & Applications*, vol. 18, no. 8, pp. 172–174, 2009.
- [3] W. F. Darsow and E. T. Olsen, "Characterization of idempotent 2-copulas," *Note di Matematica*, vol. 30, no. 1, pp. 147–177, 2010.
- [4] J. J. Li, H. Fan, and Y. Wang, "Image denoising algorithm based on dyadic contourlet transform," *Applied Mechanics and Materials*, vol. 40–41, pp. 591–597, 2011.
- [5] G. Hernandez-Gomez, R. E. Sanchez-Yanez, V. Ayala-Ramirez, and F. E. Correa-Tome, "Natural image segmentation using the CIELab space," in *Proceedings of the 19th International Conference on Electronics Communications and Computers (CONIELECOMP '09)*, pp. 107–110, February 2009.
- [6] G. Cui, M. R. Luo, B. Rigg, G. Roesler, and K. Witt, "Uniform colour spaces based on the DIN99 colour-difference formula," *Color Research and Application*, vol. 27, no. 4, pp. 282–290, 2002.
- [7] D.-H. Shin, R.-H. Park, S. Yang, and J.-H. Jung, "Block-based noise estimation using adaptive Gaussian filtering," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 218–226, 2005.
- [8] S. Gao, C. Li, and D. Bi, "Image enhancement algorithm based on NF-ICM," *Chinese Optics Letters*, vol. 8, no. 5, pp. 474–477, 2010.
- [9] A. Saffor, A. R. Ramli, and K.-H. Ng, "A comparative study of image compression between JPEG and wavelet," *Malaysian Journal of Computer Science*, vol. 14, no. 1, pp. 39–45, 2001.
- [10] R. R. Coifman and D. L. Donoho, "Translation-invariant denoising," in *Wavelet in Statistics*, vol. 103 of *Lecture Notes in Statistics*, pp. 125–150, Springer, New York, NY, USA, 1994.
- [11] A. G. Bruce and H.-Y. Gao, "WaveShrink with firm shrinkage," *Statistica Sinica*, vol. 7, no. 4, pp. 855–874, 1997.
- [12] M. Jansen, M. Malfait, and A. Bultheel, "Generalized cross validation for wavelet thresholding," *Signal Processing*, vol. 56, no. 1, pp. 33–44, 1997.
- [13] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744–2756, 2002.
- [14] C. S. Lee and Y. H. Kuo, *Fuzzy Techniques in Image Processing*, vol. 52, Springer, New York, NY, USA, 2000.
- [15] C. S. Lee, Y. H. Kuo, and P. T. Yu, "Weighted fuzzy mean filters for image processing," *Fuzzy Sets and Systems*, vol. 89, pp. 15–180, 1997.

- [16] S. H. Wang, L. N. Wu, Y. D. Zhang et al., "Ant colony algorithm used for bankruptcy prediction," in *Proceedings of the 2nd International Symposium on Information Science and Engineering (ISISE '09)*, pp. 137–139, Shanghai, China, 2009.
- [17] D. S. Levine and M. Aparicio IV, Eds., *Neural Networks for Knowledge Representation and Inference*, Psychology Press, 2013.
- [18] S. Yuying and L. Jingjing, "A semi-implicit image denoising algorithm in matrix form," *Journal of Xuzhou Institute of Technology (Natural Sciences Edition)*, vol. 27, no. 4, 2012.

Research Article

A Multiresolution Image Completion Algorithm for Compressing Digital Color Images

R. Gomathi¹ and A. Vincent Antony Kumar²

¹ Department of ECE, Anna University, Dindigul Campus, Dindigul, Tamil Nadu 624 622, India

² Department of Information Technology, PSNA College of Engineering and Technology, Dindigul, Tamil Nadu 624 622, India

Correspondence should be addressed to R. Gomathi; mathi_r2k2@yahoo.co.in

Received 28 November 2013; Revised 31 January 2014; Accepted 3 February 2014; Published 25 March 2014

Academic Editor: Feng Gao

Copyright © 2014 R. Gomathi and A. Vincent Antony Kumar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper introduces a new framework for image coding that uses image inpainting method. In the proposed algorithm, the input image is subjected to image analysis to remove some of the portions purposefully. At the same time, edges are extracted from the input image and they are passed to the decoder in the compressed manner. The edges which are transmitted to decoder act as assistant information and they help inpainting process fill the missing regions at the decoder. Textural synthesis and a new shearlet inpainting scheme based on the theory of p -Laplacian operator are proposed for image restoration at the decoder. Shearlets have been mathematically proven to represent distributed discontinuities such as edges better than traditional wavelets and are a suitable tool for edge characterization. This novel shearlet p -Laplacian inpainting model can effectively reduce the staircase effect in Total Variation (TV) inpainting model whereas it can still keep edges as well as TV model. In the proposed scheme, neural network is employed to enhance the value of compression ratio for image coding. Test results are compared with JPEG 2000 and H.264 Intracoding algorithms. The results show that the proposed algorithm works well.

1. Introduction

Image inpainting [1–5] is a method for recovering regions in images whose pixels are distorted or removed in some way. Inpainting methods are commonly based on Partial Differential Equations (PDEs) and Total Variation (TV) models. In PDEs technique [6], pixel values around the region to be inpainted are considered to be the boundary condition for a boundary value problem. Then, a proper equation for interpolating in that area will be solved. Image inpainting has a variety of applications such as text and object removal, denoising, superresolution, digital zooming, filling-in, and compression. Galić et al. [7] use an inpainting technique directly for compression whereas, in previous studies, image inpainting is treated only as a preprocessing step to increase other existing image compression standards. Bugeau et al. [8] offered a working algorithm for image inpainting trying to approximate the global minimum of an energy functional that combines the three fundamental concepts of self-similarity, coherence, and propagation. In

this method, when the image does not have enough patches to copy from, either because the mask is too spread and the patch size is large or because the mask is placed on a singular location on the image, then the results are poor and although the presence of a geometry term seems to help, it is clearly not enough. The authors did not solve these problems and they did not optimize the search in the patch space.

Chan et al. [6] have proposed TV wavelet inpainting models. The main benefit of TV model is that it can keep the edges very well. But, the method has the drawback called staircase effect. To overcome this defect, we analyze the physical characteristics of TV model [9] and p -Laplacian operator [10] in local coordinates. At the same time, the traditional wavelets [11] are not very effective in dealing with multidimensional signals containing distributed discontinuities such as edges. To overcome this limitation, one has to use basis elements with much higher directional sensitivity and of various shapes to be able to capture the intrinsic geometrical features of multidimensional phenomena.

In this paper, a new discrete multiscale representation [12] called the Discrete Shearlet Transform (DST) is introduced to perform the inpainting based on p -Laplacian operator in wavelet domain. This approach, which is based on the shearlet transform, combines the power of multiscale methods with a unique ability to capture the geometry of multidimensional data and is optimally efficient in representing images containing edges. In the proposed algorithm, the correlated portions are identified and they are removed at encoder and filled by image inpainting at the decoder. A p -Laplacian based inpainting method which employs a DST is presented which can effectively reduce the staircase effect in TV model and can be used to achieve less computing time. Gradient descent back propagation [13] with adaptive learning rate is proposed for compression.

This paper is arranged as follows. First, in Section 2, the statement of the problem is described. In Section 3, the framework of proposed coding scheme is discussed. Specifically, Section 3 shows the ANNs based compression in shearlet domain and the shearlet image inpainting based on p -Laplacian operator. In Section 4, the experimental results are presented. Section 5 describes conclusion.

2. Statement of the Problem

A standard image model [6] is defined as

$$z(x) = f(x) + n(x), \quad (1)$$

where $f(x)$ is original noise free model; $n(x)$ is Gaussian white noise.

The standard wavelet transform of $z(x)$ is given by

$$z(\alpha, x) = \sum_{j,k} \alpha_{j,k} \psi_{j,k}(x), \quad j \in Z, k \in Z^2, \quad (2)$$

where $\alpha = \alpha_{j,k}$ and is wavelet coefficients; ψ is mother wavelet function.

Damages in the wavelet domain cause loss of wavelet coefficients of $z(x)$ on the index region I , $\{\alpha_{j,k}\}$'s with $j, k \in I$ represent those wavelet components missing or damaged. The task of inpainting is to restore the missing coefficients in a proper manner, so that the image will have as much information being restored as possible.

For inpainting when we used the traditional wavelets [14–16], they do not deal with multidimensional discontinuities such as edges. Recently, a theory for multidimensional data called Multiscale Geometric Analysis (MGA) has been developed. Many new MGA tools have been proposed such as ridgelet, curvelet, bandlet, and contourlet, which provide higher directional sensitivity than wavelets. Shearlets [17–19], a new approach proposed in this paper, not only possess all the above properties but also are equipped with a rich mathematical structure similar to wavelets, which are associated with a multiresolution analysis. Shearlets form a tight frame at various scales and directions and are optimally sparse in representing image with edges.

The discrete shearlet transform of $z(x)$ is given by

$$z(\alpha, x) = \langle z, \psi_{j,l,k} \rangle \quad j, l \in Z, k \in Z^2 \quad (3)$$

and the corresponding image inpainting model in shearlet domain is

$$\min \varsigma_{j,l,k} T(f, z) = \int |\nabla_x f(\varsigma, x)| dx + \sum_{j,l,k} \frac{\lambda}{2} (\varsigma_{j,l,k} - \alpha_{j,l,k})^2, \quad (4)$$

where,

$$f(\varsigma, x) = \sum_{j,l,k} \varsigma_{j,l,k} \psi_{j,l,k}(x); \quad \varsigma_{j,l,k}(j, l, k) \in I, \\ j, l \in Z, k \in Z^2, \quad (5)$$

$\lambda = 0$ if $j, l, k \in I$, otherwise it is assigned with

positive constant.

When this model is solved, the use of TV norm can retain sharp edges while reducing noise and other oscillations. But the corresponding Euler-Lagrange equation is not trivial to compute since it is highly nonlinear and ill-posed in strong sense. Furthermore, this model has the drawback that is called stair case effect. To overcome these deficiencies, a p -Laplacian operator [20] is introduced in this new shearlet inpainting model and is

$$\min \varsigma_{j,l,k} T(f, z) = \frac{1}{p} \int |\nabla_x f(\varsigma, x)|^p dx \\ + \sum_{j,l,k} \frac{\lambda}{2} (\varsigma_{j,l,k} - \alpha_{j,l,k})^2, \quad 1 < p < 2, \quad (6)$$

where p can be adaptively selected based on the local gradient features of images. That is, away from edges, p will be approached to 2 to overcome the staircase effect; on the contrary, p will be approached to 1 to preserve edges. So this new model can effectively reduce the staircase effect in TV model whereas it can still retain the sharp edges as TV model.

The Euler-Lagrange equation of the above inpainting model is

$$-\nabla \cdot (|\Delta_x f(\varsigma, x)|^{p-2} \cdot \nabla_x f(\varsigma, x)) + \lambda_{j,l,k} (\varsigma_{j,l,k} - \alpha_{j,l,k}) = 0. \quad (7)$$

The gradient descent flow of (7) is

$$(\varsigma_{j,l,k})_t = \nabla \cdot (|\Delta_x f(\varsigma, x)|^{p-2} \cdot \nabla_x f(\varsigma, x)) \\ - \lambda_{j,l,k} (\varsigma_{j,l,k} - \alpha_{j,l,k}). \quad (8)$$

The above equation is solved by the simple explicit finite difference algorithm. To simplify the formulation, we introduce the standard finite difference notations, such as the forward differences:

$$D_1^+ f_{k,l} = f_{k+1,l} - f_{k,l}, \quad D_2^+ f_{k,l} = f_{k,l+1} - f_{k,l} \quad (9)$$

and the backward differences:

$$D_1^- f_{k,l} = f_{k,l} - f_{k-1,l}, \quad D_2^- f_{k,l} = f_{k,l} - f_{k,l-1}. \quad (10)$$

- (1) Divide the original image in frequency domain into 8×8 pixel blocks and reshape each one into 64×1 column vector.
- (2) Arrange the column vectors into a matrix of 64×1024 .
- (3) Let the target matrix equal to the matrix in step 2.
- (4) Choose a Gradient Descent Back Propagation learning algorithm to start training.
- (5) Simulate the network with the input matrix and the target matrix.
- (6) Obtain the output matrices of the hidden layer and the output layer.
- (7) Post-process them to obtain the compressed image, and the reconstructed image, respectively.

ALGORITHM 1: Image compression and decompression using ANNs.

We note that it is important to evaluate the nonlinear term, which we denote as

$$SH_{\text{curv}} = \nabla \cdot \left(\left| \nabla_x \hat{f} \right|^{p-2} \nabla_x \hat{f}(\zeta, x) \right) \quad (11)$$

in (8). However, the p -Laplace operator is defined in the pixel domain. In this paper, we calculate it straightforwardly by transforming the shearlet domain to the pixel domain to compute the p -Laplace operator and then transform back to the shearlet domain. That is, we calculate the following:

$$f = \sum_{j,l \in \mathbb{Z}, k \in \mathbb{Z}^2} \langle \zeta, \psi_{j,l,k} \rangle \psi_{j,l,k}. \quad (12)$$

For all (i, j) , we compute the following:

$$\begin{aligned} \text{curv}_{i,j} = & D_1^- \frac{D_1^+ f_{i,j}}{\left(|D_1^+ f_{i,j}|^2 + |D_2^+ f_{i,j}|^2 + \varepsilon \right)^{(2-p)/2}} \\ & + D_2^- \frac{D_2^+ f_{i,j}}{\left(|D_1^+ f_{i,j}|^2 + |D_2^+ f_{i,j}|^2 + \varepsilon \right)^{(2-p)/2}}, \end{aligned} \quad (13)$$

where ε is a small positive number which is used to prevent the numerical blow-up when

$$|D_1^+ f_{i,j}|^2 + |D_2^+ f_{i,j}|^2 = 0. \quad (14)$$

Then we compute the curvature projection on the wavelet basis by

$$SH_{\text{curv}}(j, l, k) = \langle \text{curv}, \psi_{j,l,k} \rangle. \quad (15)$$

3. Framework of Proposed Scheme

The method proposed in this section is based on removing redundancy at the encoder and restoring the removed information using an inpainting method at the decoder. In this algorithm, redundancy removal is performed through detecting texture regions with similar statistical characteristics and dividing the image into homogeneous regions. The overall system with encoder and decoder diagrams is depicted in Figures 1 and 2. In the following subsections, encoder and decoder blocks are discussed separately.

3.1. Design of Encoder

3.1.1. Image Analysis. The input image is subjected to image analysis by extracting edges from the images in order to identify both the structural and textural regions. An input image is divided into 8×8 blocks. Then they are identified as textural or structural blocks based on their distance from edges. The block is identified as a structural one when it contains more numbers of pixels having very small distance from edges. The remaining blocks are called textural blocks.

The important blocks from both textural regions and structural regions are selected by using different algorithms [21]. The remaining blocks are easily removed during encoding. In this way, the blocks are identified and removed. After the removal of specified blocks, in the original image, the regions are filled with the corresponding DC value or filled with pure red, or green, or blue component, in the case of color images.

3.1.2. ANN Based Image Compression. Gradient descent back propagation algorithm [13] is the widely used algorithm in Artificial Neural Networks (ANNs). The feed-forward neural network architecture is capable of approximating most problems with high accuracy and generalization ability. This algorithm mainly focuses on the error correction learning rule. Error propagation consists of two passes through the different layer of the network: a forward pass and a backward pass. In the forward pass, the input vector is applied to the sensory nodes of the network and its effect propagates through the network layer by layer.

Finally, a set of outputs is produced as the actual response of the network. During the forward pass, the synaptic weights of the networks are all fixed. During the back pass, the synaptic weights are all adjusted in accordance with an error correction rule. The actual response of the network is subtracted from the desired response to produce an error signal. This error signal is then propagated through the network against the direction of synaptic conditions. The synaptic weights are adjusted to make the actual response of the network move closer to the desired response. *Procedure*

for Image Compression. For experiment, feed-forward neural network with three layers is selected. Input layer, hidden layer with 16 neurons, and output layer with 64 neurons are introduced in that network. Back propagation algorithm is used for training process. For training the network, 256×256 Lena image is selected (see Algorithm 1).

- (1) Decompose f_a^{j-1} into a low pass image f_a^j and a high pass image f_d^j by applying Laplacian Pyramid Scheme.
- (2) Use pseudo polar grid to compute $P \cdot f_d^j$.
- (3) Band pass filter the output matrix $P \cdot f_d^j$
- (4) Re-assemble the Cartesian sampled values
- (5) Take the 2D IFFT.

ALGORITHM 2: Construction of discrete shearlet transform.

- (1) Obtain a noisy image as input

$$z(x) = f(x) + n(x)$$
- (2) Smooth the noisy image using

$$\frac{\partial z}{\partial x} = \text{div}(g(\|\nabla z\|) \cdot \nabla z)$$
- (3) Analyze the smoothed image by extracting edges through

$$|\nabla z(x, y)| = \sqrt{(\partial_x z(x, y))^2 + (\partial_y z(x, y))^2}$$

$$\theta = \arctan\left(\frac{\partial_y z(x, y)}{\partial_x z(x, y)}\right)$$
- (4) Identify the structural and textural regions based on the extracted edges.
- (5) Skip the regions that are not selected as necessary information instead fill the DC value and create a DC filled image
- (6) Perform ANN Compression and Decompression using

$$\text{comp}^{\text{in}} = \sum_{i=1}^N V_{ij} z_i, \quad 1 \leq j \leq K$$

$$z_j^{\text{in}} = \sum_{i=1}^K W_{ij} \text{comp}^{\text{in}}, \quad 1 \leq i \leq N,$$

where V_{ij} & W_{ij} are weights of compressor and decompressor respectively.
- (7) Start the inpainting process with initial guess $\varsigma_{j,l,k}^{\text{new}} = \alpha_{j,l,k} \chi_{j,l,k}$.
Set $\varsigma_{j,l,k}^{\text{old}} = 0$ and the initial error $E = \|\varsigma^{\text{new}} - \varsigma^{\text{old}}\|_2$
Where $\chi_{j,l,k} = \begin{cases} 1 & (j, l, k) \in I \\ 0 & (j, l, k) \notin I \end{cases}$ and I is the inpainting region.
- (8) While $i \leq N$ or $E \leq \delta$, do
 - (a) Set $\varsigma^{\text{old}} = \varsigma^{\text{new}}$
 - (b) Calculate SH_{curv} using (12)–(15)
 - (c) For all j, l, k update

$$\varsigma_{j,l,k}^{\text{new}} = \varsigma_{j,l,k}^{\text{old}} + \frac{\Delta_t}{\Delta_x} (\text{SH}_{\text{curv}} - \lambda_{j,l,k} (\varsigma_{j,l,k} - \alpha_{j,l,k})),$$

where Δ_t is the time step size & Δ_x is the space grid size
 - (d) Compute error $E = \|\varsigma^{\text{new}} - \varsigma^{\text{old}}\|_2$
 - (e) End the while loop

ALGORITHM 3: Pseudocode of proposed framework with PDEs inpainting algorithm based on p Laplacian operator.

3.2. Design of Decoder. After performing the ANN based compression at encoder, the decompression process is carried out at decoder. Final reconstructed image is obtained through p -Laplacian inpainting based on DST and texture synthesis. The theory of shearlets is discussed in addition to the two most important modules, namely, shearlet domain p -Laplacian inpainting and texture synthesis.

3.2.1. Continuous Shearlet Transform. The continuous shearlet transform [17, 18] is a nonisotropic version of the continuous wavelet transform with a superior directional sensitivity. For $n = 2$,

$$\text{SH}_{\psi} f(a, s, t) = \langle f, \psi_{a,s,t} \rangle. \quad (16)$$

Each analyzing element $\psi_{a,s,t}$ is termed shearlet. The frequency tiling of shearlets is shown in Figure 3.

3.2.2. Discrete Shearlet Transform. By sampling the continuous shearlet transform, we can get a discrete transform which is shown in Figure 4 (see Algorithm 2).

3.2.3. Shearlet Domain p -Laplacian Inpainting. From the pure inpainting perspective, the inpainting problem may be stated as follows. Let I be the original image, which is composed by a source area, denoted by ϕ , whose pixel values are known, and a target area, denoted by Ω , representing the damaged region to be repaired or a region to be filled

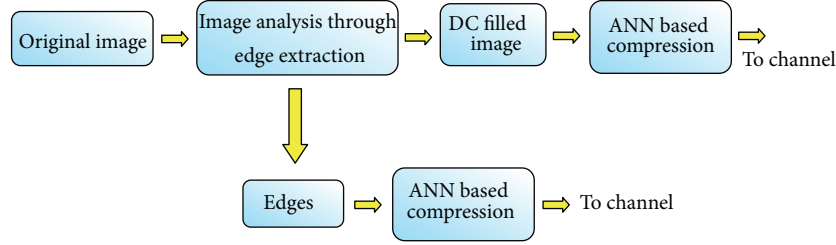


FIGURE 1: Block diagram of proposed image compression scheme—encoder.

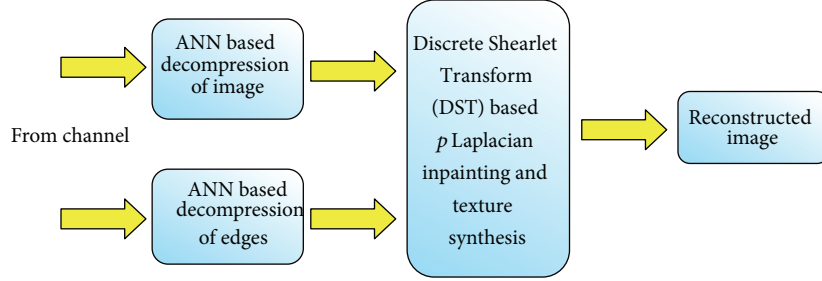


FIGURE 2: Block diagram of proposed image compression scheme—decoder.

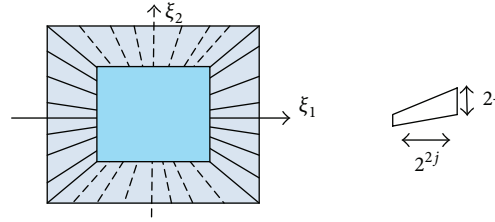
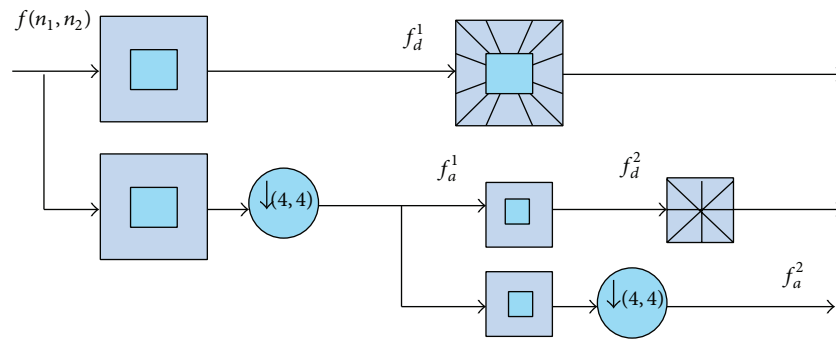
FIGURE 3: (a) The tiling of the frequency plane. The tiling of horizontal cone is illustrated in solid line; the tiling of vertical cone is in dashed line. (b) Frequency support of shearlet satisfies parabolic scaling (j -decomposition level). The Figure shows only the support for $\xi_1 > 0$; the other half of the support, for $\xi_1 < 0$, is symmetrical.

FIGURE 4: Succession of Laplacian pyramid and directional filtering.

in, and this means being inpainted. As shown in Figure 5, these are nonoverlapping areas, that is, $I = \phi \cup \Omega$ and $\partial\Omega$ for the boundary between the source and target regions. The simplified architecture for image coding with inpainting is shown in Figure 5 and the pseudocode of proposed work is described as follows: the pseudocode of proposed framework with PDEs inpainting algorithm based on p -Laplacian operator (see Algorithm 3).

3.2.4. Texture Synthesis. The missing texture blocks are filled in with the texture from its surrounding [2]. Let the region to be filled be denoted by Ω . The lost block will now be filled, pixel by pixel, in a raster fashion. Let I_t be a representative template touching the left of a pixel $p(i, j) \in \Omega$. We proceed to find an estimate of I_t from the available neighborhood, such that a given distance $d(I_t, \hat{I}_t)$ is minimized. As per [2], d is a normalized Sum of Squared Differences (SSD) metric.

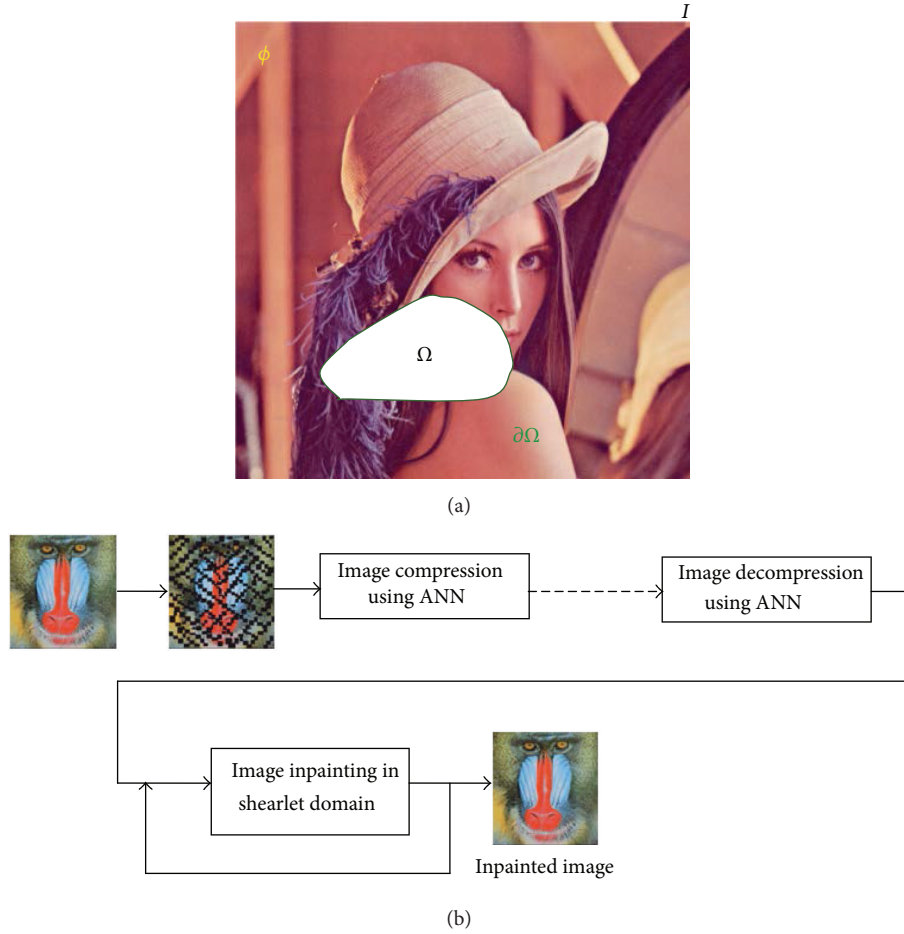


FIGURE 5: (a) Illustration of the inpainting problem. (b) The simplified architecture for image coding with inpainting.

Once such an estimate of I_t is found, we choose the pixel to the immediate right of estimate of I_t as our candidate for $p(i, j) \in \Omega$. For stochastic textures, the algorithm selects at random one of the pixels neighboring estimate of I_t . The template I_t can be a simple seed block of 3×3 pixels. Then, of all possible 3×3 blocks in the 8 neighbourhoods, the one with the minimum normalized SSD is found and a pixel to its right is copied into the current pixel in the lost block. This algorithm is considerably fast when using the improvements in [21–25].

4. Performance Evaluation

We illustrate the performance of the proposed algorithm for image compression with image inpainting in shearlet domain using ANNs and compare it with the image inpainting method proposed by Liu et al. [3]. The codes are written in MATLAB 2008a.

4.1. Test Conditions and Parameters. The proposed algorithm is tested with color images from USC-SIPI and Kodak image database. In all testes, we use shearlet base p -Laplacian inpainting and we set the parameter values $p = 1.2$, $\lambda = 0$ for noiseless images and $\lambda = 0.08$ for noisy images.

4.2. Overall Performance. Figure 6 shows test image Lena and corresponding results of proposed system. In this test, edges are extracted from the input image and shown in Figure 6(b) and the image with removed blocks (25% removal) is shown in Figure 6(c).

Based on the preserved blocks, the shearlet p -Laplacian inpainting gives results in Figure 6(d). When comparing the restored image in Figure 6(e) with JPEG 2000, proposed scheme saves 36.15% of bits with QP = 75. The comparison of standard images shows up to 50% bits saving which is achieved by the proposed scheme compared to JPEG 2000 and up to 31.61% bit saving compared to edge based image inpainting method proposed by Liu et al. [3]. The bit-saving results are shown in Table 1.

Figure 7 shows the reconstructed images by the proposed scheme with standard H.264/AVC Intracoding. The bit-rate saving is also noticeable, shown in Table 2, but not as much as the comparison with JPEG 2000. The proposed scheme can acquire 48.07% bit-rate saving compared to the state-of-the-art H.264 Intracoding with QP = 24 for an image Kodim19. Bit-rate saving of 33.01% is achieved for an image Kodim11 by the proposed scheme when compared with edge based image inpainting method.

Figure 8 shows the comparison results with JPEG 2000. The first row in Figure 8 shows results of the proposed



FIGURE 6: Comparison with JPEG 2000 with QP = 75. (a) Input image. (b) Edges. (c) Image with 25% of portions removed. (d) Reconstructed image after inpainting and texture synthesis. (e) Reconstructed image by JPEG 2000.

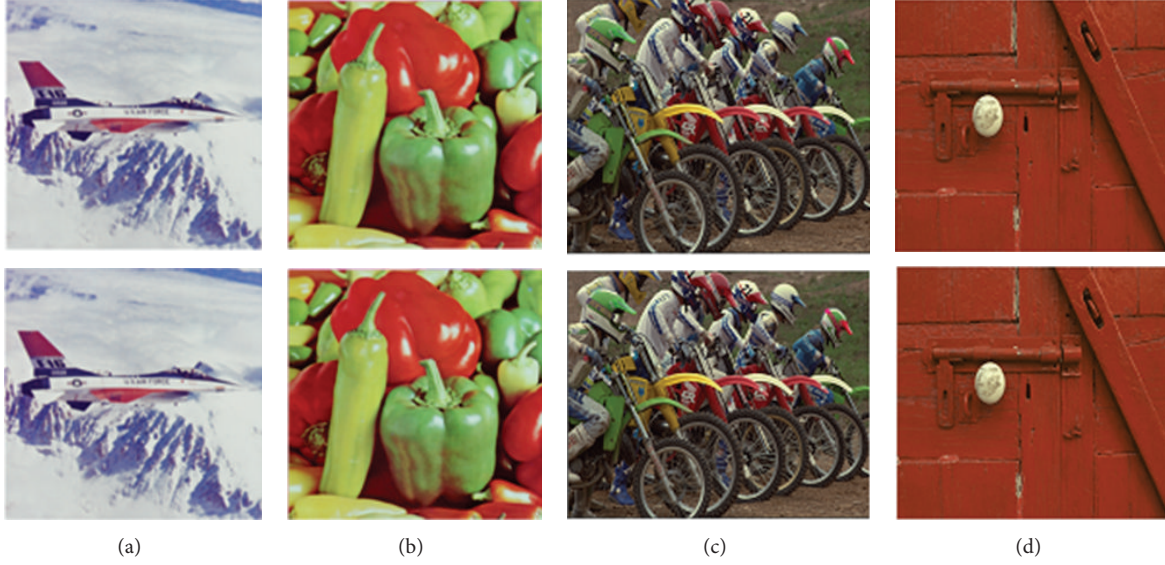


FIGURE 7: Comparison with H.264/AVC with QP = 24. (a) Jet (25% removal); (b) peppers (32.5% removal); (c) kodim05 (23% removal); (d) kodim02 (51% removal). The top row shows the reconstructed images by proposed scheme and the bottom row shows the reconstructed images by H.264/AVC Intracoding.

TABLE 1: Bit-rate saving of proposed scheme compared to JPEG 2000 and edge based image inpainting method (QP = 75).

| Original image | Block removal (%) | PSNR (dB) | Bit-rate (bpp) | | | Bit-rate saving (%) | |
|----------------|-------------------|-----------|----------------|-----------------------------|-----------------|-------------------------|---|
| | | | JPEG 2000 | Edge based image inpainting | Proposed scheme | Compared with JPEG 2000 | Compared with edge based Image inpainting |
| Jet | 25.0 | 45.67 | 1.156 | 0.919 | 0.780 | 32.53 | 15.13 |
| Lena | 25.0 | 37.23 | 1.112 | 0.888 | 0.710 | 36.15 | 20.05 |
| Peppers | 32.5 | 39.87 | 1.217 | 0.965 | 0.812 | 33.28 | 15.85 |
| Kodim02 | 51.0 | 35.78 | 1.058 | 0.709 | 0.529 | 50.00 | 25.39 |
| Kodim03 | 42.8 | 36.45 | 0.895 | 0.608 | 0.485 | 45.81 | 20.23 |
| Kodim07 | 32.8 | 39.34 | 1.079 | 0.802 | 0.622 | 42.35 | 22.44 |
| Kodim11 | 35.0 | 38.00 | 1.368 | 1.047 | 0.716 | 47.66 | 31.61 |
| Kodim19 | 26.7 | 41.34 | 1.276 | 0.915 | 0.680 | 46.71 | 25.68 |
| Kodim20 | 54.2 | 33.11 | 0.897 | 0.638 | 0.522 | 41.81 | 18.18 |
| Kodim23 | 53.0 | 34.12 | 0.821 | 0.567 | 0.492 | 40.07 | 13.23 |

TABLE 2: Bit rate savings of proposed scheme compared to H.264/AVC intra and edge based image inpainting method (QP = 24).

| Original image | Block removal (%) | PSNR (dB) | Bit-rate (bpp) | | | Bit-rate saving (%) | |
|----------------|-------------------|-----------|----------------|-----------------------------|-----------------|---------------------|---|
| | | | H.264 | Edge based image inpainting | Proposed scheme | Compared with H.264 | Compared with edge based image inpainting |
| Jet | 25.0 | 42.56 | 0.985 | 0.880 | 0.727 | 26.16 | 17.35 |
| Lena | 40.0 | 35.67 | 0.993 | 0.869 | 0.772 | 22.26 | 11.16 |
| Peppers | 32.5 | 36.45 | 1.311 | 1.080 | 0.818 | 37.61 | 24.27 |
| Mandrill | 43.0 | 36.48 | 0.880 | 0.783 | 0.691 | 21.47 | 11.74 |
| Kodim02 | 51.0 | 32.11 | 0.948 | 0.701 | 0.510 | 46.20 | 27.25 |
| Kodim03 | 42.8 | 34.23 | 0.710 | 0.562 | 0.495 | 30.28 | 11.92 |
| Kodim07 | 32.8 | 35.00 | 0.876 | 0.751 | 0.671 | 23.38 | 10.63 |
| Kodim11 | 35.0 | 36.78 | 1.354 | 1.098 | 0.736 | 45.68 | 33.01 |
| Kodim13 | 42.0 | 37.02 | 0.900 | 0.872 | 0.714 | 20.66 | 18.12 |
| Kodim19 | 26.7 | 39.36 | 1.246 | 0.956 | 0.647 | 48.07 | 32.32 |
| Kodim20 | 54.2 | 30.00 | 0.823 | 0.636 | 0.473 | 42.49 | 25.58 |

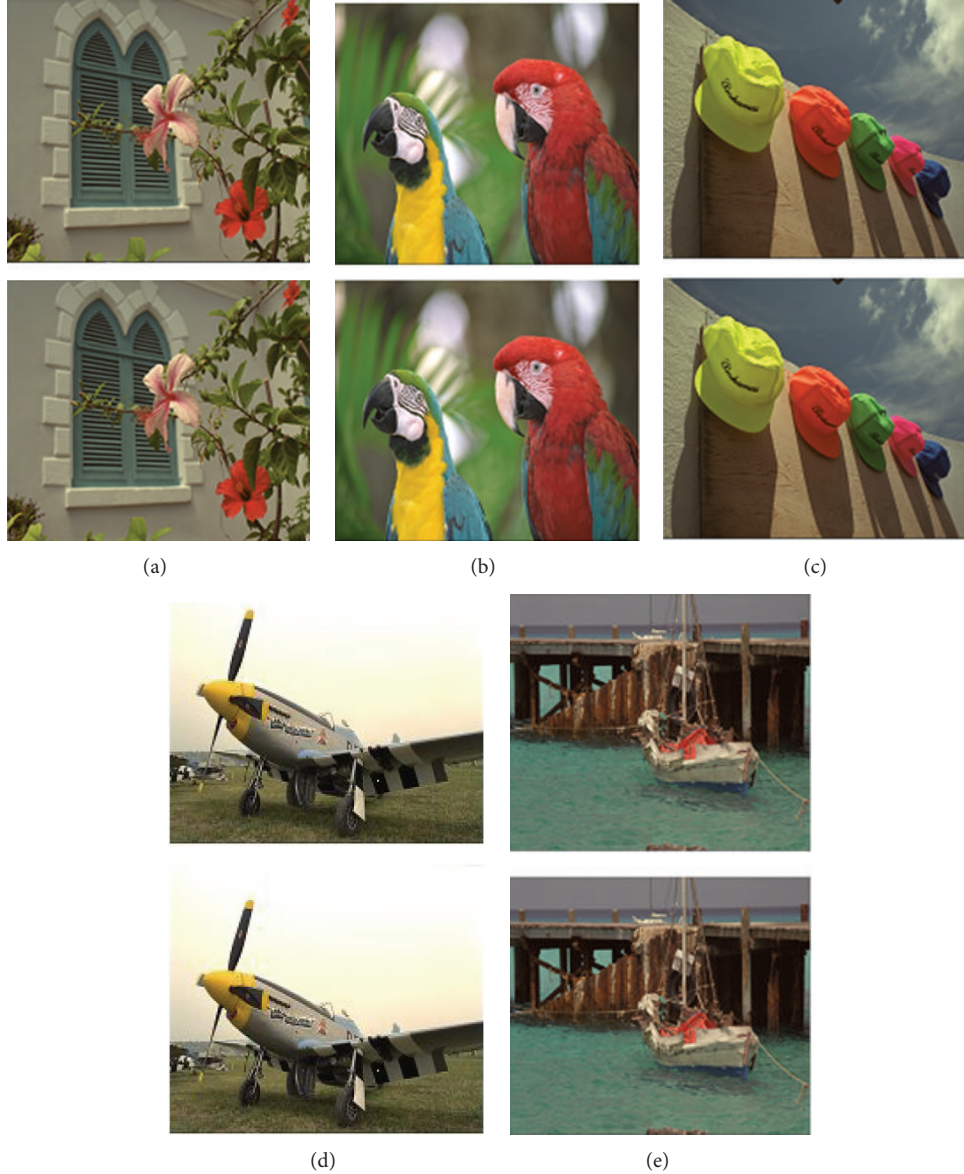


FIGURE 8: Comparison with JPEG 2000 with QP = 75. (a) Kodim07 (32.8% removal); (b) kodim23 (53% removal); (c) kodim03 (42.8% removal); (d) kodim20 (54.2% removal); (e) kodim11 (35% removal). The first row shows the restored images by proposed algorithm and the second row shows the restored images by JPEG 2000.

method and the second row presents JPEG 2000 results. The bits saving of proposed system is indicated in Table 1. The proposed scheme averagely saves 43.54% bits with QP = 75 for the five images shown in Figure 8. Peak Signal-to-Noise Ratio (PSNR) is also used to measure the quality of the restored images. It is defined as follows:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\max_I^2}{\text{MSE}} \right), \quad (17)$$

where \max_I is the maximum possible pixel value of the image.

Figure 9 gives objective quality comparisons between the proposed scheme, JPEG 2000, and H.264/AVC Intracoding. It can be observed that the objective quality of the proposed scheme outperforms the two standard compression schemes

(JPEG 2000 and H.264/AVC Intra) and edge based inpainting algorithm at both low and high bit rates. The proposed algorithm is adapted for subjective quality, and PSNR is not a good measure to evaluate the subjective quality, especially for the resulting images of p -Laplacian inpainting in shearlet domain.

The reconstructed images with same visual quality regardless of large PSNR difference are shown in Figure 10. Inpainting with edges produces the results which are difficult to differentiate compared with H.264/AVC Intra. But the inpainting method reduces the bit rate because much less number of bits is required for coding edges. The proposed algorithm averagely saves 21.07% bit rate for images having high percentage of texture regions (mandrill and kodim13) when QP = 24 and it is shown in Table 2.

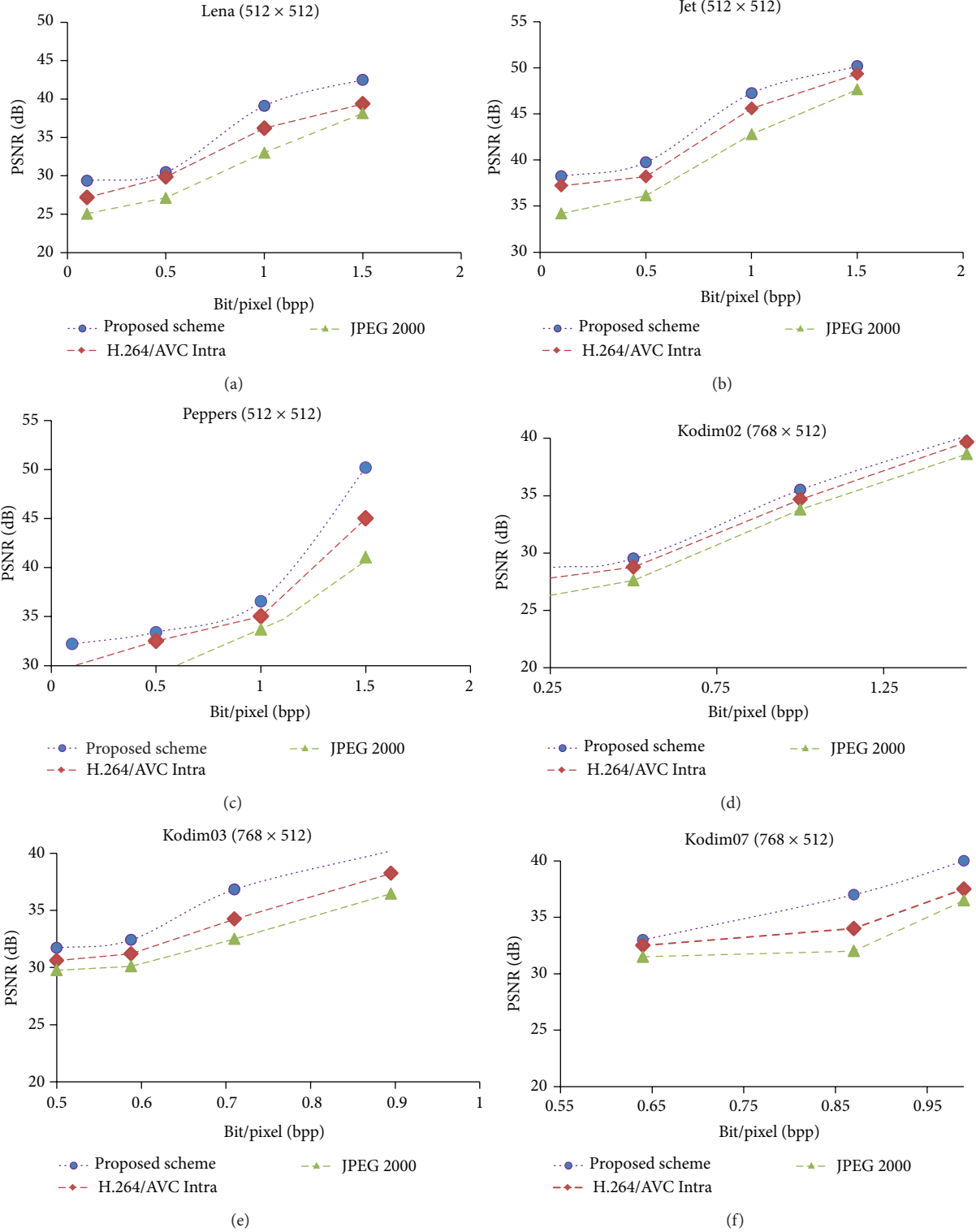


FIGURE 9: Objective quality comparison between proposed scheme, H.264/AVC Intra, and JPEG 2000 on some typical color images.

The recent no-reference quality assessment method is adopted in this paper that is proposed in [26] to evaluate the proposed scheme in comparison with JPEG 2000 and H.264/AVC Intra. Note that in [26], blocking artifacts are detected within a compressed image. From Figure 11, it can

be observed that the reconstructed images by the proposed scheme contain less blocking than those by JPEG 2000 and H.264/AVC Intra at similar compression ratios. The result shows the better bit-rate reduction when compared with previous methods as far as visual quality is concerned.

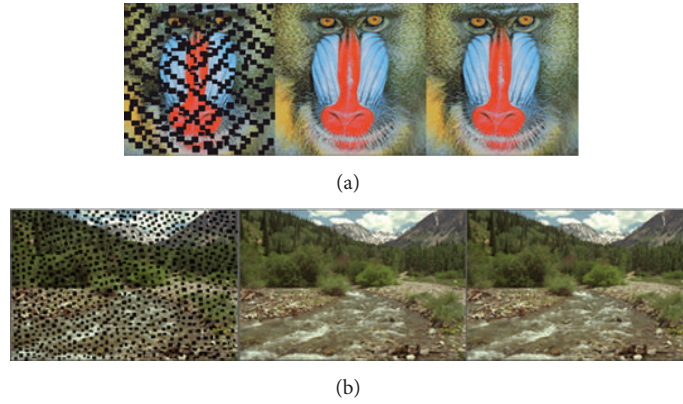


FIGURE 10: Subjective quality comparison between the proposed scheme and H.264/AVC Intra. QP is 24 for high quality. From top to bottom: mandrill (43% removal) and kodim13 (42% removal). From left to right: incomplete image with black blocks; reconstructed image by the proposed scheme; reconstructed image by H.264/AVC Intra. Please note that the proposed scheme reconstructs both highly textured images with 21.07% bit-rate saving. Please observe that in kodim13 (the first row) the mandrill eye can also be reconstructed on both sides.

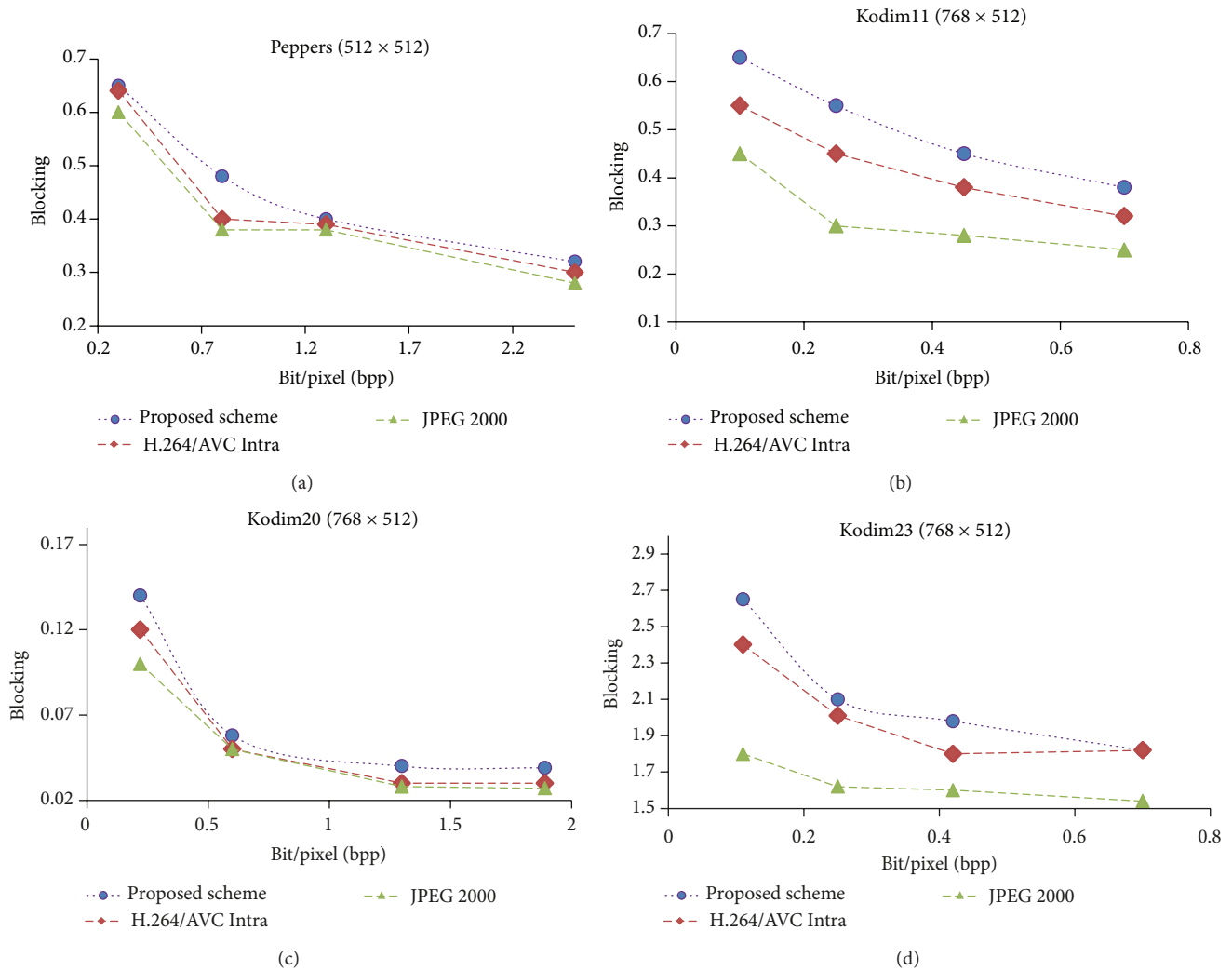


FIGURE 11: Quality assessment (by measuring the blocking artifacts by the method in [26]) results that compare the proposed scheme with JPEG 2000 and H.264/AVC Intra on some typical color images.

4.3. Computational Complexity. For the experiment, Intel 2 GHz CPU is used. The shearlet domain p -Laplacian inpainting with the help of assistant information module is realized in MATLAB 2008a. According to our empirical results, the shearlet domain p -Laplacian inpainting process at the decoder needs averagely 3–4 iterations to converge. We found that the proposed scheme needs 1–3 s to decode a 768×512 image at 40% removal. The time complexity of inpainting is in general proportional to the size of removed regions.

The proposed algorithm is simpler because of the presence of region removal and assistant information generation based on Discrete Shearlet Transform (DST).

5. Conclusion

In this paper, we develop an ANNs based image compression framework that adopts shearlet domain inpainting technique. In this proposed algorithm, the correlated regions are identified and removed automatically at the encoder. Then they are restored at the decoder by using inpainting scheme. The key techniques used for coding are gradient descent back propagation algorithm with adaptive learning rate and p -Laplacian image inpainting in shearlet domain. Experimental results show that the proposed scheme produces good results. The proposed scheme produces up to 50% and 48.07% bits saving when compared with JPEG 2000 and H.264/AVC Intra, respectively.

Edge extraction can be flexible and adaptable for compression. Finding the regions that can be eliminated is considered to be an open problem and it seems that solving this problem will lead to increase in compression ratios and output quality.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] V. Bastani, M. S. Helfroush, and K. Kasiri, "Image compression based on spatial redundancy removal and image inpainting," *Journal of Zhejiang University: Science C*, vol. 11, no. 2, pp. 92–100, 2010.
- [2] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, pp. 1033–1038, September 1999.
- [3] D. Liu, X. Sun, F. Wu, S. Li, and Y.-Q. Zhang, "Image compression with edge-based inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 10, pp. 1273–1286, 2007.
- [4] S. D. Rane, G. Sapiro, and M. Bertalmio, "Structure and texture filling-in of missing image blocks in wireless transmission and compression applications," *IEEE Transactions on Image Processing*, vol. 12, no. 3, pp. 296–303, 2003.
- [5] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th ACM International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*, pp. 417–424, July 2000.
- [6] T. F. Chan, J. Shen, and H.-M. Zhou, "Total variation wavelet inpainting," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, pp. 107–125, 2006.
- [7] I. Galić, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel, "Image compression with anisotropic diffusion," *Journal of Mathematical Imaging and Vision*, vol. 31, no. 2-3, pp. 255–269, 2008.
- [8] A. Bugeau, M. Bertalmio, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2634–2645, 2010.
- [9] M. Lysaker and X.-C. Tai, "Iterative image restoration combining total variation minimization and a second-order functional," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 5–18, 2006.
- [10] H.-Y. Zhang, Q.-C. Peng, and Y.-D. Wu, "Wavelet inpainting based on p -Laplace operator," *Acta Automatica Sinica*, vol. 33, no. 5, pp. 546–549, 2007.
- [11] U. A. Ignácio and C. R. Jung, "Block-based image inpainting in the wavelet domain," *The Visual Computer*, vol. 23, no. 9–11, pp. 733–741, 2007.
- [12] G. R. Easley, D. Labate, and F. Colonna, "Shearlet-based total variation diffusion for denoising," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 260–268, 2009.
- [13] A. Shaik, C. R. K. Reddy, and S. A. Ali, "Empirical analysis of image compression through wave transforms and NN," *International Journal of Computer Science and Information Technologies*, vol. 2, pp. 924–931, 2011.
- [14] R. H. Chan, Y.-W. Wen, and A. M. Yip, "A fast optimization transfer algorithm for image inpainting in wavelet domains," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1467–1476, 2009.
- [15] M. S. Joshi, R. R. Manthalkar, and Y. V. Joshi, "Image compression using curvelet, ridgelet and wavelet transform: a comparative study," *ICGST International Journal on Graphics, Vision and Image Processing (GVIP)*, vol. 8, pp. 25–31, 2008.
- [16] L. Granai, F. Moschetti, and P. Vanderheynt, "Ridgelet transform applied to motion compensated images," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 381–384, April 2003.
- [17] G. Kutyniok and D. Labate, "Resolution of the wavefront set using continuous shearlets," *Transactions of the American Mathematical Society*, vol. 361, no. 5, pp. 2719–2754, 2009.
- [18] K. Guo, D. Labate, and W.-Q. Lim, "Edge analysis and identification using the continuous shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 27, no. 1, pp. 24–46, 2009.
- [19] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.
- [20] S. Li, "Existence of solutions to a superlinear p -Laplacian equation," *Electronic Journal of Differential Equations*, vol. 2001, no. 66, pp. 1–6, 2001.
- [21] L. Luo, F. Wu, S. Li, Z. Xiong, and Z. Zhuang, "Advanced motion threading for 3D wavelet video coding," *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 601–616, 2004.
- [22] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

- [23] Ö. N. Gerek and A. E. Çetin, "A 2-D orientation-adaptive prediction filter in lifting structures for image coding," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 106–111, 2006.
- [24] C. Zhu, X. Sun, F. Wu, and H. Li, "Video coding with spatio-temporal texture synthesis and edge-based inpainting," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '08)*, pp. 813–816, June 2008.
- [25] Y. Lu, W. Gao, and F. Wu, "Efficient background video coding with static sprite generation and arbitrary-shape spatial prediction techniques," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 5, pp. 394–405, 2003.
- [26] F. Pan, X. Lin, S. Rahardja, E. P. Ong, and W. S. Lin, "Measuring blocking artifacts using edge direction information," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '04)*, vol. 2, pp. 1491–1494, June 2004.

Research Article

Image Sequence Fusion and Denoising Based on 3D Shearlet Transform

Liang Xu, Junping Du, and Zhenhong Zhang

Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Junping Du; junpingdu@126.com

Received 22 December 2013; Accepted 8 February 2014; Published 17 March 2014

Academic Editor: Feng Gao

Copyright © 2014 Liang Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a novel algorithm for image sequence fusion and denoising simultaneously in 3D shearlet transform domain. In general, the most existing image fusion methods only consider combining the important information of source images and do not deal with the artifacts. If source images contain noises, the noises may be also transferred into the fusion image together with useful pixels. In 3D shearlet transform domain, we propose that the recursive filter is first performed on the high-pass subbands to obtain the denoised high-pass coefficients. The high-pass subbands are then combined to employ the fusion rule of the selecting maximum based on 3D pulse coupled neural network (PCNN), and the low-pass subband is fused to use the fusion rule of the weighted sum. Experimental results demonstrate that the proposed algorithm yields the encouraging effects.

1. Introduction

Video sensors have been applied extensively to video monitor and machine vision fields with the performance improvement and the reduced cost. In real applications, because of the effect of objects movement, occlusion, and illumination variation factors, single video sensor may not satisfy the requirements [1]. To get the complete information of a scene, multiple different sensors are employed simultaneously to capture the content of the same scene [2]. To utilize the information captured from multiple sensors sufficiently and efficiently, we need to combine the contents from different video sensors into an image sequence. This can be easily satisfied by the image sequence fusion methods, which can merge multiple image sequences from different sensors into a single image sequence containing all important information, eliminating redundancy and improving the availability of the information.

In the past decade, a variety of image fusion approaches have been developed for different applications [3–5]. The simplest fusion method is the weighted average of source images. However, the method is brittle and easy to lead to the contrast reduction and introduce artifacts. Now, fusion methods based on multiscale decomposition in the transform

domain are increasingly popular because of better robustness and reliability [6–8]. The most existing fusion methods are designed for static images; however, the fusion methods especially for image sequences and videos are seldom. The fusion methods for static images can even ensure the quality of single frame by the frame-by-frame fusion [9], but the temporal consistency and stability of the image sequence is hard to be preserved. So far, there are several fusion methods for image sequences and videos. By utilizing the information of temporal axis, in recent years, several state-of-the-art video fusion approaches [10, 11] have been developed. Based on the three-dimensional surfacelet transform, Zhang et al. [12] propose a video fusion framework which fuses multiframe images of input videos as a whole procedure. However, the method is still insufficient in extracting the spatial-temporal information from videos with dynamic background. Zhang et al. [13] propose a multisensor video fusion method based on three-dimensional uniform discrete curvelet transform (3D-UDCT) and spatial-temporal structure tensor. However, these schemes did not still obtain the satisfactory results due to the insufficient capability of representing motion information.

In addition, the most existing image fusion methods still only pay attention to combine the useful pixels of source

images and seldom consider dealing with the artifacts. The noises are easily introduced into the fusion image together with the useful pixels if the further processes are not performed for eliminating noises. There are some fusion methods that consider the enhancement and denoising when the fusion procedure is performed. Piella [14] presents a variational model to perform the fusion of input images while preserving the salient information and enhancing the contrast. Yang and Li [15] propose a sparse representation-based multifocus image fusion method which can simultaneously carry out denoising and fusion of noised source images.

In the paper, we propose a novel image sequence fusion algorithm which implements fusion and denoising simultaneously in 3D shearlet transform domain. Input image sequences are first decomposed by 3D shearlet transform into high- and low-pass subbands, and then the coefficients are merged and denoised in frequency domain. Finally, inverse 3D shearlet transform is applied on the fused coefficients to reconstruct the image sequence. Multiple frame images can be decomposed to different frequency scales once in 3D space by 3D shearlet, which has certain direction selectivity and can avoid aliasing and instability for the coefficients of neighbor frames. 3D shearlet considers the motion feature in temporal axis, such that the decomposed coefficients can approximate sufficiently the spatial-temporal features of image sequences in different scales. For merging coefficients, to preserve the consistency and stability of the interframe coefficients, we propose a spatial-temporal fusion rule based on 3D PCNN [16], which can extract the spatial-temporal information of the corresponding subbands from the neighbor frames. In addition, the most existing image fusion methods still only pay attention to combine the pixels of source images and seldom consider dealing with the noises. Here, we combine the fusion and denoising together and directly deal with noises on the coefficients using the recursive filter. Superior to the separate fusion and denoising based on multiscale transforms, our method will reduce the error only due to the need to perform the decomposition and reconstruction once.

The remainder of the paper is organized as follows. Section 2 reviews basic 3D shearlet transform theory in brief. Section 3 describes the proposed image sequence fusion and denoising algorithm in detail. Section 4 presents and discusses the experimental results. Section 5 concludes.

2. 3D Shearlet Transform

In this section, we briefly review theory and properties of 3D shearlet transform, which will be used in the rest of this paper (see [17] for details).

The shearlet approach inherits the general advantages of curvelets and surfacelets. During the last decade, to overcome the limitations of wavelets and other traditional methods, a new class of multiscale methods is introduced through a novel framework. The framework can effectively combine the standard multiscale decomposition and efficiently capture anisotropic features. The shearlet just belongs to part of the new class of multiscale methods. The shearlet representation is a multiscale pyramid of well-localized waveforms defined

at various locations and orientations. This representation can break through the limitations of traditional multiscale systems in dealing with multidimensional data.

The 3D shearlet transform is constructed by a shearlet system associated with the pyramidal regions. Three pyramidal regions P_1 , P_2 , and P_3 are obtained by partition of the Fourier space \hat{R}^3 , defined as follows (shown in Figure 1):

$$\begin{aligned} P_1 &= \left\{ (\xi_1, \xi_2, \xi_3) \in R^3 : \left| \frac{\xi_2}{\xi_1} \right| \leq 1, \left| \frac{\xi_3}{\xi_1} \right| \leq 1 \right\}, \\ P_2 &= \left\{ (\xi_1, \xi_2, \xi_3) \in R^3 : \left| \frac{\xi_1}{\xi_2} \right| < 1, \left| \frac{\xi_3}{\xi_2} \right| \leq 1 \right\}, \\ P_3 &= \left\{ (\xi_1, \xi_2, \xi_3) \in R^3 : \left| \frac{\xi_1}{\xi_3} \right| < 1, \left| \frac{\xi_2}{\xi_3} \right| < 1 \right\}. \end{aligned} \quad (1)$$

The directionality of the shearlet systems is controlled through the use of shearing matrices. The 3D shearlet systems for $L^2(R^3)$ are a Parseval frame, which is obtained by using an appropriate combination of the systems of shearlets associated with the pyramidal regions P_d ($d = 1, 2, 3, l = (l_1, l_2) \in Z^2$). In this way, the 3D shearlet systems are defined as the collections consisting of the coarse-scale shearlets, the interior shearlets, and the boundary shearlets:

$$\begin{aligned} &\{\tilde{\psi}_{-1,k} : k \in Z^3\} \cup \{\tilde{\psi}_{j,l,k} : j \geq 0, l_1, l_2 = \pm 2^j, k \in Z^3\} \\ &\{\tilde{\psi}_{j,l,k,d} : j \geq 0, |l_1| < 2^j, |l_2| \leq 2^j, k \in Z^3, d = 1, 2, 3\}, \end{aligned} \quad (2)$$

where the shearing parameters l_1 and l_2 control the orientations of the support regions in 3D shearlet systems. Figure 2 illustrates a typical support region. The orientation of the support region is controlled by $l = (l_1, l_2)$. It can be seen that the support region is becoming more elongated as j increases.

A numerical implementation of the 3D discrete shearlet transform takes advantage of the sparsity properties of the corresponding continuous representation. The 3D digital shearlet transform can preserve the discrete integer lattice and enable a natural transition from the continuous to the discrete setting due to the use of shearing matrices rather than rotations. The 3D digital shearlet transform algorithm is implemented through a cascade of a multiscale decomposition and a directional filtering stage. The multiscale decomposition is first implemented using the Laplacian pyramid algorithm. And then the directional components are obtained using shearing matrices to control orientations in the pseudospherical domain.

3. Proposed Image Sequence Fusion and Denoising Algorithm

In this section, the proposed image sequence fusion and denoising algorithm is presented in detail. The main idea is that image sequence fusion and denoising are implemented simultaneously in 3D shearlet transform domain. The framework of the proposed algorithm is shown in Figure 3. For the clearness of the presentation, we assume that two registered image sequences are combined.

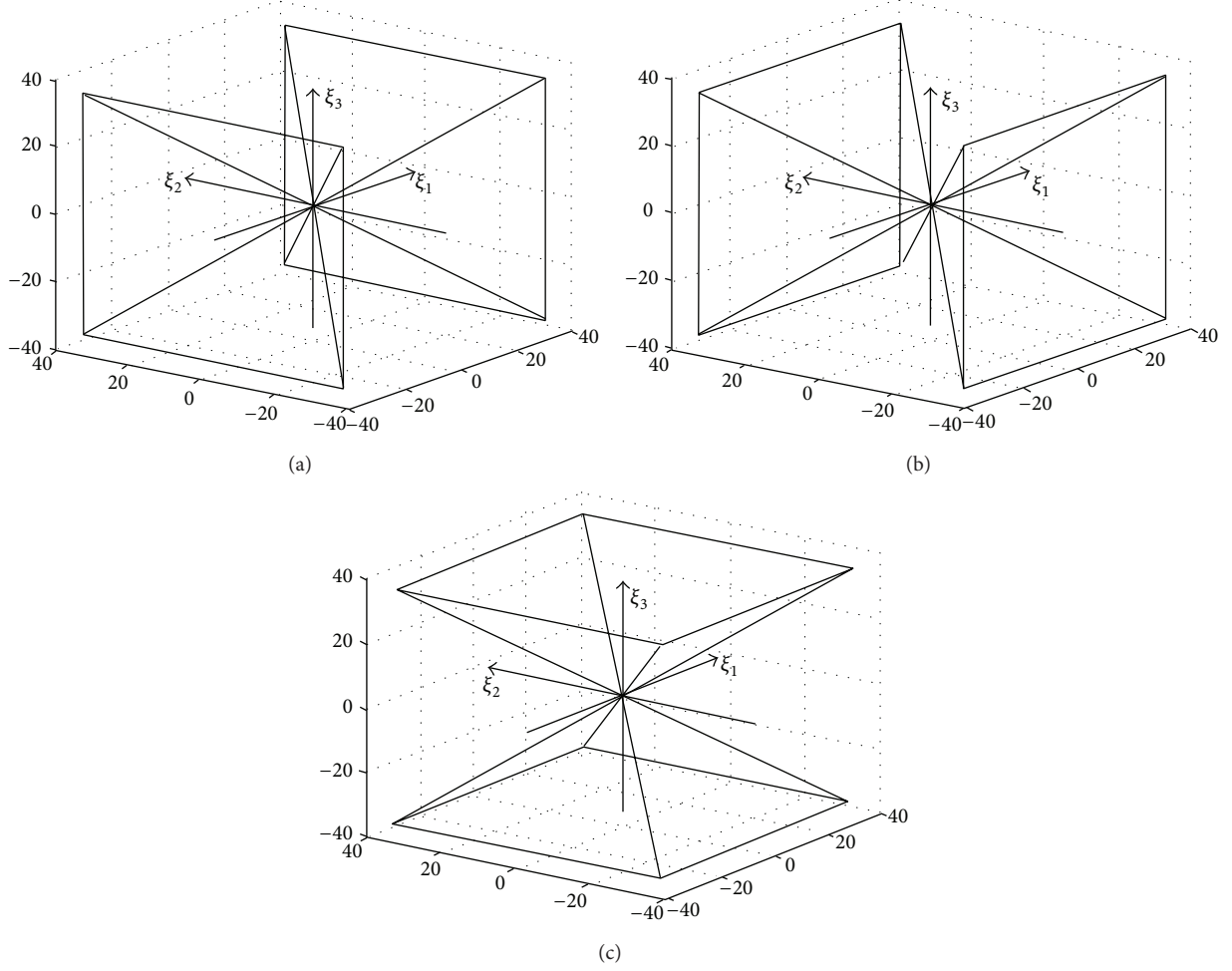


FIGURE 1: ((a)–(c)) The pyramidal regions P_1 , P_2 , and P_3 in frequency domain \hat{R}^3 are illustrated.

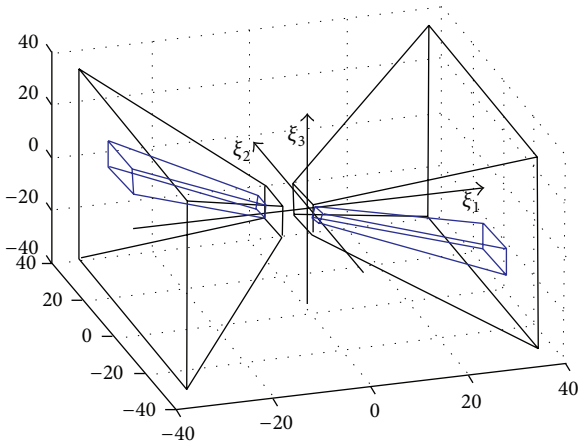


FIGURE 2: Frequency support of a representative shearlet function $\psi_{j,l,k}$, inside the pyramidal region P_1 .

Two groups of image sequences Va and Vb are decomposed by 3D shearlet transform into two groups of high-pass subbands and two groups of low-pass subbands for

several frames. Then, high-pass subbands are denoised by the recursive filter (RF) [18]. Next, the fused subband coefficients of all frames are obtained to employ a 3D PCNN-based spatial-temporal saliency fusion rule. Finally, the fused image sequence is reconstructed by inverting 3D shearlet transform on the fused coefficients of all frames. The steps of the proposed algorithm are presented as follows.

Step 1. Input image sequences are transformed using 3D shearlet transformation to the frequency domain and produce the low- and high-pass subbands $\{C_{j_0}^{Va}(x, y, t), C_{j,l}^{Va}(x, y, t)\}$ and $\{C_{j_0}^{Vb}(x, y, t), C_{j,l}^{Vb}(x, y, t)\}$ for several frames, where $C_{j_0}(x, y, t)$ denotes the low-pass subband coefficients at the coarsest scale of the t th frame and $C_{j,l}(x, y, t)$ denotes the high-pass subband coefficients at the j th scale and in the l th direction and at the t th frame.

Step 2. For the high-pass subband coefficients, the recursive filter is performed on the coefficients of each frame for eliminating noises. Then, 3D PCNN is used to compute the spatial-temporal activity levels of the denoised high-pass coefficients to obtain the spatial-temporal activity maps,

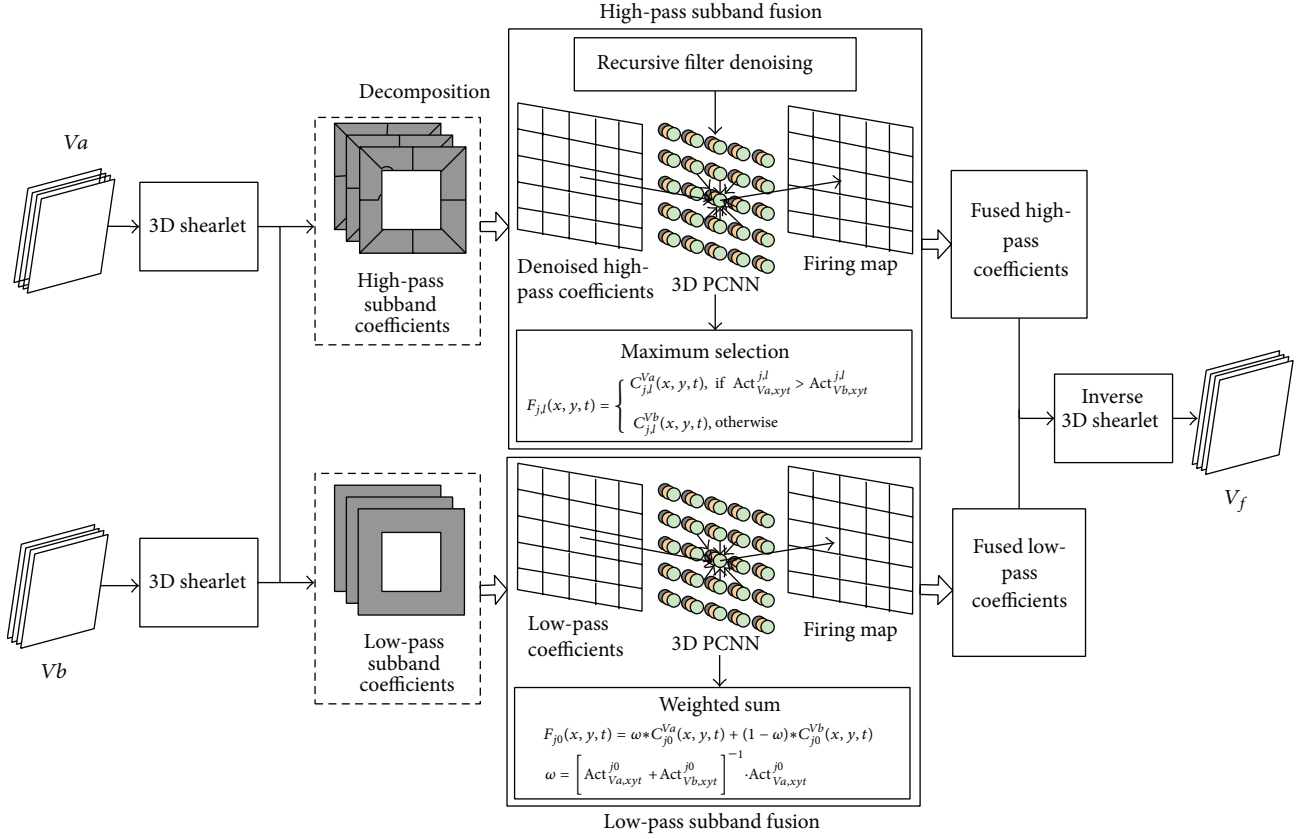


FIGURE 3: The diagram of the proposed image sequence fusion and denoising algorithm.

which are employed to merge the fused high-pass subband by the selecting maximum.

Step 3. The low-pass subband coefficients are merged using a spatial-temporal energy weighted fusion rule based on the activity maps yielded by 3D PCNN.

Step 4. Apply the inverse 3D shearlet transform to the fused coefficients $\{F_{j0}(x, y, t), F_{jl}(x, y, t)\}$ for all frames and then obtain the fused image sequence V_f .

3.1. 3D PCNN-Based Spatial-Temporal Fusion. This section discusses the proposed 3D PCNN-Based Spatial-Temporal Fusion in detail. Based on the experimental observations of synchronous pulse bursts in cat and monkey visual cortex, a novel biological neural network called PCNN is developed. PCNN neuron consists of receptive field, modulation field, and pulse generator [19, 20]. Different from the traditional neural network, PCNN is a feedback network and does not need to be trained. A secondary receptive field of PCNN, known as the linking field, integrates inputs from adjacent neurons to modulate the primary feeding field. In image processing, 2D PCNN is a single layer and two-dimensional connection neural network. Considering the spatial features of 2D image plane, 2D PCNN utilizes the output of the spatial neighborhood pixels as the inner input of next iteration. The similar neurons in PCNN generate pulses simultaneously to

compensate effectively the spatial incoherence and the slight amplitude changes, so that PCNN can measure the salient object regions completely. It has been successfully used to deal with the single static image. However, if 2D PCNN is applied directly on an image sequence frame by frame, it may not extract the temporal motion information.

To adapt PCNN to deal with image sequences, 2D PCNN is extended to 3D PCNN utilizing the correlation of neighbor frames. 3D PCNN has been used to the segmentation of stereo images successfully [16]. Here, 3D PCNN is employed to measure the activity energy of the coefficients from 3D shearlet decomposition. Let $C_{jl}(x, y, t)$ indicate the coefficient located at (x, y) in the j th scale at the l th direction and at t th frame. $C_{jl}(x, y, t)$ in each subband is inputted to 3D PCNN as the external feeding input. Both the last output of the spatial neighbor pixels and the corresponding output of the neighbor frames are used as the inner linking input. In this way, 3D PCNN can extract the spatial-temporal information sufficiently in an image sequence. The 3D PCNN is defined as follows:

$$\begin{aligned}
 F_{xyt}^{jl} [n] &= C_{jl}(x, y, t) \\
 L_{xyt}^{jl} [n] &= \exp(-\alpha_L) L_{xyt}^{jl} [n-1] \\
 &\quad + V_L \sum_{pqr} W_{xyt, pqr}^{jl} Y_{xyt, pqr}^{jl} [n-1]
 \end{aligned}$$

$$\begin{aligned}
U_{xyt}^{j,l} [n] &= F_{xyt}^{j,l} [n] * (1 + \beta L_{xyt}^{j,l} [n]) \\
\theta_{xyt}^{j,l} [n] &= \exp(-\alpha_\theta) \theta_{xyt}^{j,l} [n-1] + V_\theta Y_{xyt}^{j,l} [n-1] \\
Y_{xyt}^{j,l} [n] &= \begin{cases} 1, & \text{if } U_{xyt}^{j,l} [n] > \theta_{xyt}^{j,l} [n] \\ 0, & \text{otherwise,} \end{cases}
\end{aligned} \tag{3}$$

where the coefficient $C_{j,l}(x, y, t)$ is inputted to the feeding input $F_{xyt}^{j,l}$. The linking input $L_{xyt}^{j,l}$ is equal to the sum of neurons firing times in linking range, where α_L indicates the decay constants and V_L is the amplitude gain. $W_{xyt,pqr}$ is the weighted coefficient (p, q, r point out the size of linking range in 3D PCNN). $Y_{xyt}^{j,l} [n-1]$ is the output of the neuron from the previous iteration. The internal state signal $U_{xyt}^{j,l}$ is obtained by modulating $F_{xyt}^{j,l}$ and $L_{xyt}^{j,l}$, where β is the linking strength. $\theta_{xyt}^{j,l}$ is the threshold, where α_θ is the decay constants and V_θ is the amplitude gain. n indicates the iteration times. If $Y_{xyt}^{j,l} = 1$, the neuron will generate a pulse, called one firing. If $Y_{xyt}^{j,l} = 0$, the neuron will not generate a pulse.

In the paper, the 3D PCNN can be used to measure the spatial-temporal saliency of the coefficients. Through combining the coefficients of the corresponding scales and directions from neighbor frames, we construct a 3D volume whose size is $M \times N \times T$. In this way, a neuron model of 3D PCNN is constructed, where each of the coefficients is the external input of 3D PCNN. The neuron with the maximum coefficient value is first fired. Following, the similar neurons from the 3D space constructed by the internal linking matrix W are motivated to produce synchronization pulse through pulse propagation. The generated pulse sequence $Y[n]$ forms a 3D binary sequence, which contains the saliency information of images, for example, regions, edges, textures, and so forth.

In applications, the firing times are generally employed to represent image information. Firing times $\text{Act}_{xyt}^{j,l} [n]$ can be computed by accumulating all pixels before and including the present iteration, written as follows:

$$\text{Act}_{xyt}^{j,l} [n] = \text{Act}_{xyt}^{j,l} [n-1] + Y_{xyt}^{j,l} [n], \tag{4}$$

where $\text{Act}_{xyt}^{j,l} [n]$ is often used to indicate the total firing times in n iteration. Here, the firing times indicate the activity energy of the coefficients.

The high-pass subbands of 3D shearlet decomposition contain abundant detail information, for example, lines, edges, contours, and so forth. To preserve the detail components in fused images, we propose a fusion rule of spatial-temporal selecting maximum based on 3D PCNN for the high-pass subbands. Coefficients with large firing times are selected as the fused coefficients. Consequently, the fused high-pass subband coefficients located at (x, y) in the j th

scale at the l th direction and at t th frame denoted as $F_{j,l}(x, y, t)$ are defined as

$$F_{j,l}(x, y, t) = \begin{cases} C_{j,l}^{Va}(x, y, t), & \text{if } \text{Act}_{Va,xyt}^{j,l} > \text{Act}_{Vb,xyt}^{j,l} \\ C_{j,l}^{Vb}(x, y, t), & \text{otherwise.} \end{cases} \tag{5}$$

The low-pass subband of 3D shearlet decomposition in the coarsest scale contains the main energy of source images and denotes abundant structural information. The fusion rule of the low-pass subband employs a spatial-temporal weighted fusion rule based on firing times of 3D PCNN. The fused coefficients of low-pass subbands denoted as $F_{j0}(x, y, t)$ employ a weighted fusion rule based on firing times of 3D PCNN on coefficients $C_{j0}^{Va}(x, y, t)$ and $C_{j0}^{Vb}(x, y, t)$, which are defined as

$$F_{j0}(x, y, t) = \omega * C_{j0}^{Va}(x, y, t) + (1 - \omega) * C_{j0}^{Vb}(x, y, t), \tag{6}$$

where ω is the weight of coefficients and Act_{xyt}^{j0} is computed by (3) and (4):

$$\omega = \left[\text{Act}_{Va,xyt}^{j0} + \text{Act}_{Vb,xyt}^{j0} \right]^{-1} \cdot \text{Act}_{Va,xyt}^{j0}. \tag{7}$$

Finally, apply the inverse 3D shearlet transform to the fused coefficients of N frames $\{F_{j0}(x, y, t), F_{j,l}(x, y, t)\}$, and then obtain the fused image sequence V_f .

3.2. Recursive Filter Denoising. The previous section introduces the fusion method of the decomposed coefficients. When input images contain some noises, the fused image will also introduce noise artifacts if the coefficients are only directly merged. So we need first to deal with the coefficients for denoising. This section presents mainly the recursive filter method [18] which is used as the denoising of the coefficients. The recursive filter is a real-time edge-preserving smoothing filter. Comparing with the separate fusion and denoising based on multiscale transform, the fusion and denoising simultaneously on the coefficients decomposed by 3D shearlet transform will reduce the error caused by the decomposition and reconstruction. In addition, the denoising filter is run on the coefficients of different scales and directions such that it enhances the robustness of the algorithm.

The low-pass subband contains the main energy of images and denotes the structural information. The high-pass subbands contain the abundant details, for example, lines, edges, and contours. In general, the noises appear in the high-pass subbands, so here the recursive filter is only performed on the high-pass subband coefficients to obtain the denoised coefficients $\tilde{C}_{j,l}$, which are defined as follows:

$$\tilde{C}_{j,l} = \text{RF}(C_{j,l}), \tag{8}$$

where RF indicates the recursive filter. When the noises are eliminated, the details need to be preserved on the high-pass

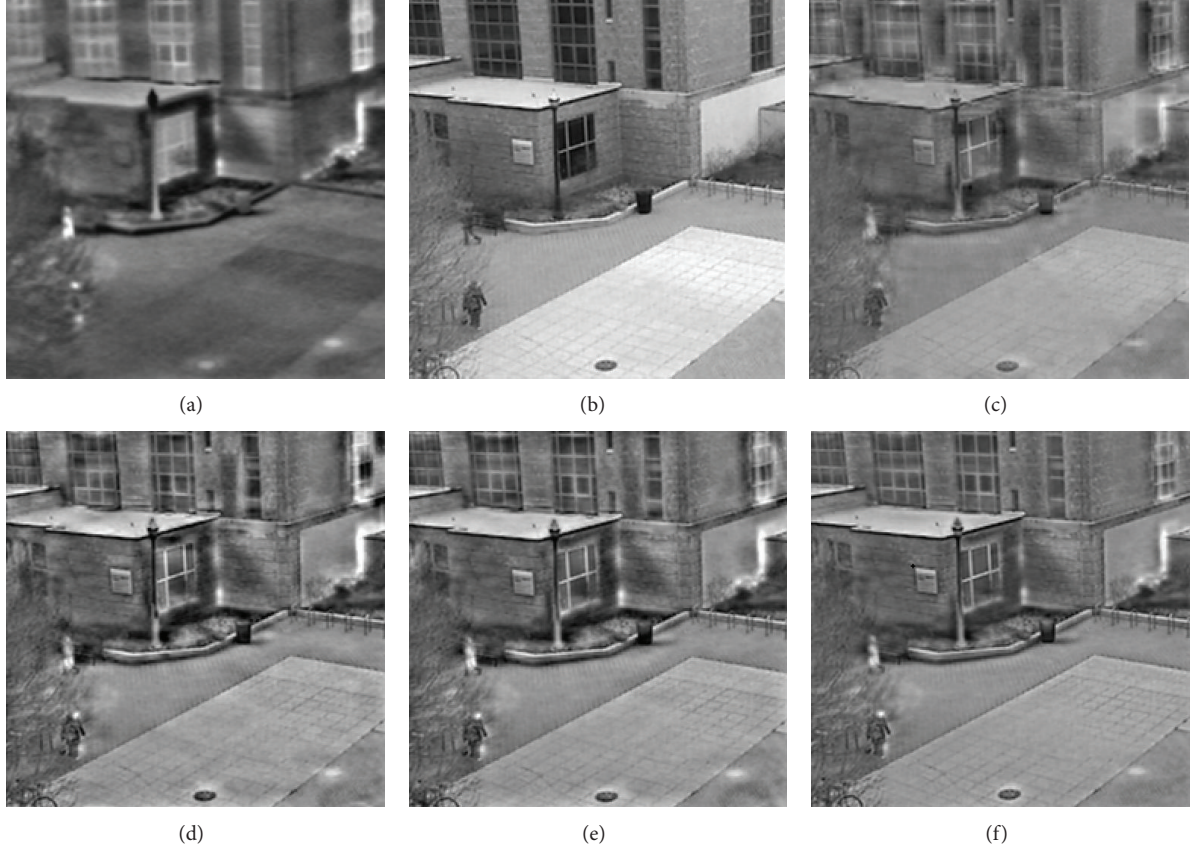


FIGURE 4: Comparisons of different fusion methods. (a) A frame of visible light image sequence; (b) corresponding frame of infrared image sequence; and fused frame images using (c) DWT, (d) 3D DWT, (e) 3D DTCWT, and (f) proposed 3DShearlet-3DPCNN methods.

subbands. This can be easily satisfied by the edge-preserving filter, called the recursive filter:

$$J[n] = (1 - a^d)I[n] + a^d J[n-1], \quad (9)$$

where $a \in [0, 1]$ is a feedback coefficient, $I[n]$ is the value of the n th coefficient of the input high-pass subband, $J[n]$ is the n th coefficient of the filtered high-pass subband, and d is the distance between neighborhood coefficients of the high-pass subband. As d increases, a^d goes to zero, stopping the propagation chain and preserving details on high-pass subbands. More details about the recursive filter can refer to [18].

The recursive filter is performed in all N frame images for obtaining the denoised high-pass coefficients. Following, the fused coefficients can be obtained through merging the denoised coefficients.

4. Experiments and Analysis

In this section, the proposed image sequence fusion algorithm based on 3D shearlet and 3D PCNN (named as 3DShearlet-3DPCNN) is tested on several groups of image sequences. For comparison, besides the fusion scheme proposed in this paper, another three fusion algorithms, the discrete wavelet transform based (DWT), 3D DWT based, and

3D dual-tree complex wavelet transform based (3D DTCWT) methods, are used to fuse the same image sequences. All of these methods use the average sum and absolute maximum selection schemes for merging low- and high-pass subband coefficients, respectively. The decomposition level of all of the transforms is three. It is assumed that source image sequences have been registered. The tested collections contain the clear and the noised image sequences.

4.1. Fusion Results of Clear Image Sequences. The first experiment is a pair of visible light and infrared image sequences without noises. A pair of frames from input image sequences and four fused images produced by DWT, 3D DWT, 3D DTCWT, and proposed 3DShearlet-3DPCNN are shown in Figure 4. Figure 4(a) is an infrared image frame from an input image sequence, and Figure 4(b) is a visible light source image frame from the corresponding visible light image sequence. From Figure 4, we can observe that Figure 4(c) is the worst result, which has the lower contrast, and the motion objects tend to blur. Figures 4(d)–4(f), produced by 3D DWT, 3D DTCWT, and proposed 3DShearlet-3DPCNN, present the better results, which have higher contrast, clearer motion objects. However, Figures 4(d) and 4(e) have some distortion near the windows. The edges of the windows appear warp. The middle wall between windows introduces the dark regions.

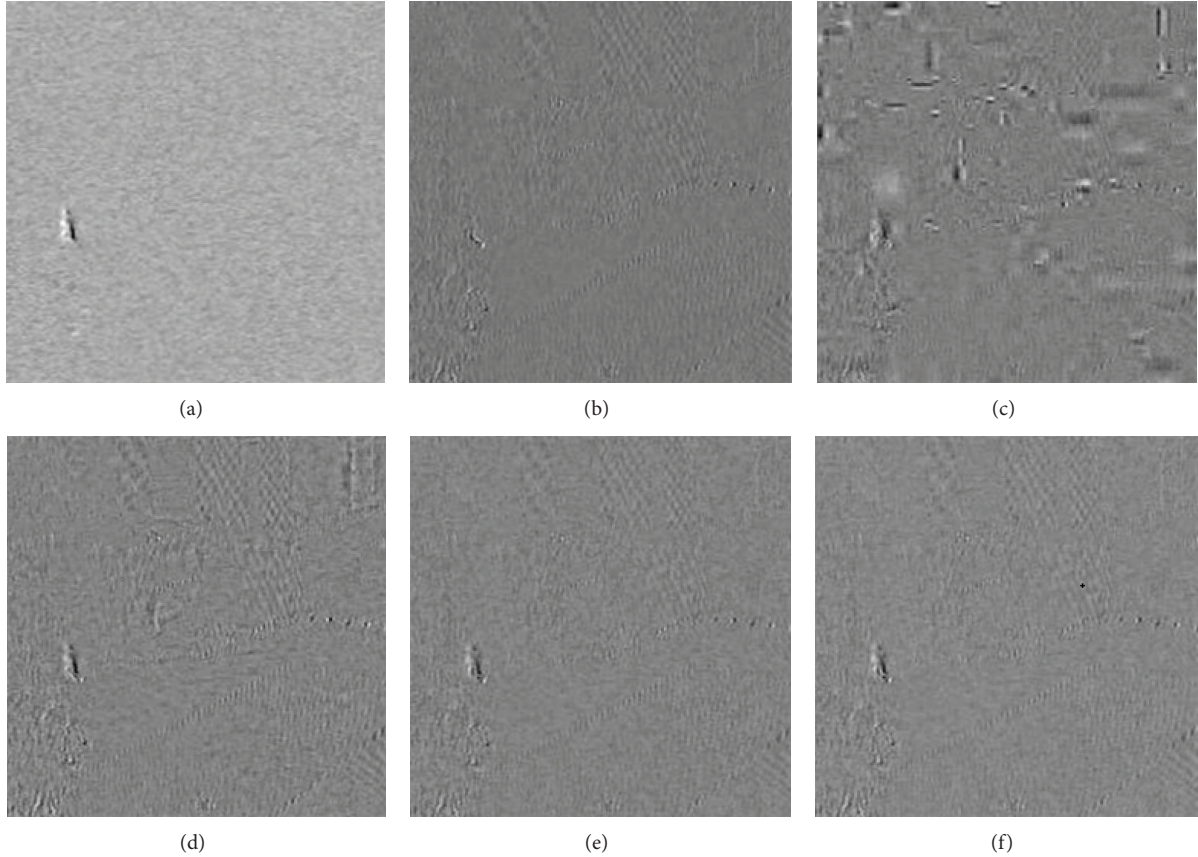


FIGURE 5: One set of IFD images of input and fused image sequences corresponding to (a)–(f), IFD images between the current frame images in Figures 4(a)–4(f), and their corresponding preframe images, respectively.

In contrast, the proposed method (Figure 4(f)) yields the best result. This comparison reveals that the proposed fusion approach effectively determines complementary or redundant information between input image frames. It can preserve all the useful information of the input image frames while avoiding artifacts. In addition, to evaluate the performance of image sequence fusion methods in temporal stability and consistency, a clearer comparison is made by examining the interframe-difference (IFD) between the current frame and the preframe, shown in Figure 5. One set of IFD images for source and fused frames between the current frames in Figure 4 and their corresponding preframes are shown in Figure 5. Obviously, the IFDs in Figure 5(c) introduce many artifacts, which exist neither in Figure 5(a) nor in Figure 5(b). In Figure 5(d), artifacts are greatly reduced. In Figures 5(e) and 5(f), we cannot see nearly any artifacts. This further demonstrates that 3D DTCWT and proposed fused image frames have better temporal stability and consistency.

Figure 6 shows one pair of input image frames and the corresponding fused results. Comparisons of different fused results (Figures 6(c)–6(f)) and the frames fused using DWT, 3D DWT, and 3D DTCWT (Figures 6(c)–6(e)) are not clear enough and have lower contrast. In particular, artifacts around the man were also introduced in Figure 6(c). The image frame fused using the proposed approach (Figure 6(f)) is obviously clearer and has stronger contrast. Experimental

results further demonstrate that the proposed algorithm can effectively improve the quality of the fused image sequence.

For more accurate comparison, besides the visual analysis, the performance of fusion algorithms needs to be further measured by objective quantitative analysis tools. Here two metric tools were used: the spatial-temporal gradient preservation based video fusion performance metric ($DQ^{AB/F}$) [21] and the mutual information of IFD images (IFD_MI) [9]. $DQ^{AB/F}$ indicates how much the spatial-temporal information is extracted and transferred into the fused image sequence. IFD_MI reflects the performance of the image sequence fusion method in temporal stability and consistency. The higher the values are for the two metrics, the better are the fusion results.

The measurement results of the different fusion methods for the image frames in Figure 4 and Figure 6 are shown in Figure 7. The results show that the DWT method has the lowest scores. Compared with the DWT, 3D DWT and 3D DTCWT methods have the improved results. The proposed method represents the highest performance in the listed fusion approaches for $DQ^{AB/F}$ and IFD_MI metrics. The quantitative results are consistent with the visual analysis, illustrating that the proposed fusion method has the superior performance while it can preserve the temporal stability and consistency for image sequences.

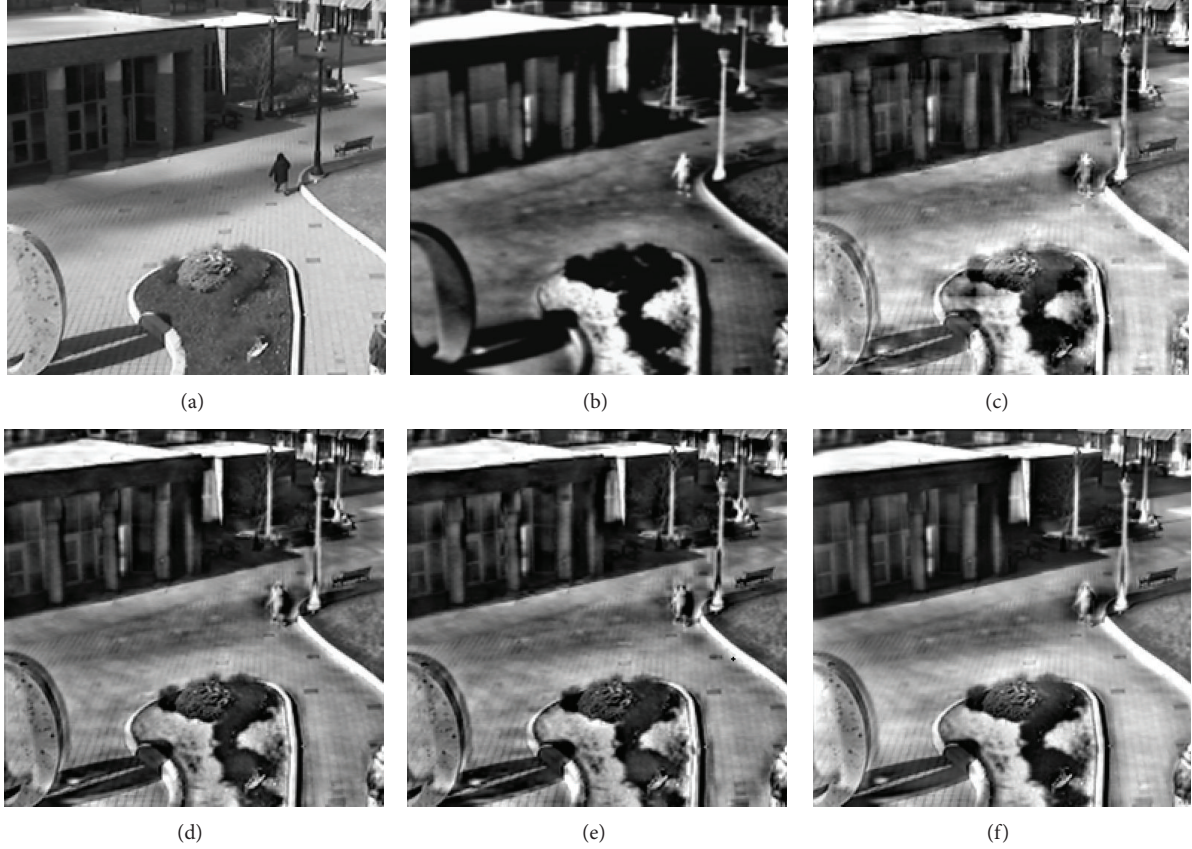


FIGURE 6: Comparisons of different fusion methods. (a) A frame of visible light image sequence; (b) corresponding frame of infrared image sequence; and fused frame images using (c) DWT, (d) 3D DWT, (e) 3D DTCWT, and (f) proposed 3DShearlet-3DPCNN methods.

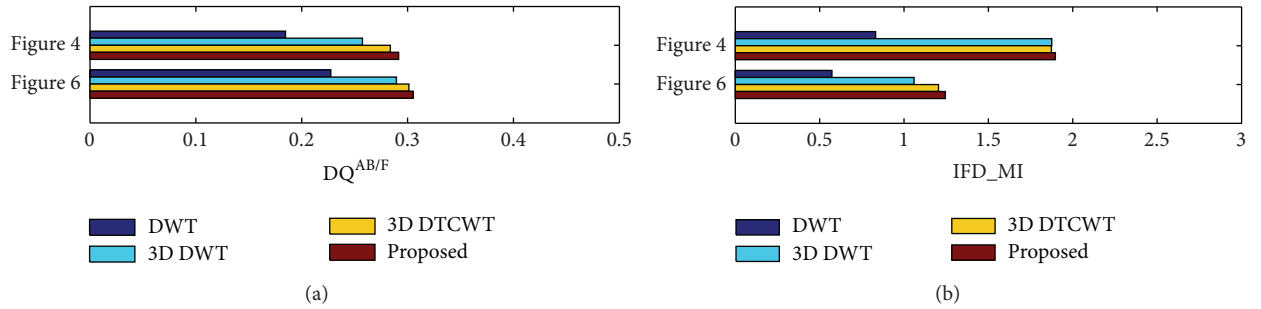


FIGURE 7: Quality measures for the different fusion methods.

4.2. Fusion Results of Noised Image Sequences. In previous discussion, the fusion results of different algorithms for the clear image sequences have been analyzed by both visual and objective metric tools. In the section, the denoising and fusion results of different algorithms for the noised image sequences are compared. Our method is compared to the 3D DTCWT fusion method followed by 3D DTCWT denoising [22], named as 3D DTCWT-FD.

Figure 8 presents the fusion and denoising results yielded by our method and 3D DTCWT-FD method from a pair of real world noise image sequences. Observing the source image frames (Figures 8(a) and 8(b)), the visible light image frame (Figure 8(a)) is clear and the infrared image frame

(Figure 8(b)) contaminated by noises. It can be seen that the fusion image frames (Figures 8(c) and 8(d)) have eliminated the noises. However, the 3D DTCWT-FD result (Figure 8(c)) becomes blurred due to over-smoothing. The proposed denoising and fusion method yields the optimal result which nearly removes all the noises and preserves the high contrast.

In addition, to evaluate the performance of different fusion and denoising schemes, $DQ^{AB/F}$, IFD_MI, and peak signal to noise ratio (PSNR) metrics are used. Table 1 shows the results of $DQ^{AB/F}$ and IFD_MI metrics for Figure 8. We can see that the proposed method gets the higher scores in both $DQ^{AB/F}$ and IFD_MI. This indicates that the proposed



FIGURE 8: Dublin source image frames and denoised and fused results: (a) a visible light image frame; (b) an infrared image frame; and fused image frames using (c) 3D DTCWT-FD and (d) proposed 3DShearlet-3DPCNN.

fusion and denoising simultaneously method contains all the useful information from source images, while eliminating the noises. Figure 9 presents the PSNR results for 1000 frame fusion images associated with Figure 8. Here, to get the PSNR results, the clear visible light source image frames are used as the reference images. From Figure 9, we can observe that the results of the proposed method have higher PSNR scores which indicate that our method yields better the denoised images. Consequently, from the comparisons of $DQ^{AB/F}$, IFD_MI, and PSNR metrics, the experiments further demonstrate that the proposed image sequence fusion and denoising simultaneously method yields the satisfactory results, which merge the important complementary information and remove the artifacts.

5. Conclusion

The paper proposes a novel algorithm for image sequence fusion and denoising based on 3D shearlet transform. The most existing image fusion methods do not deal with the

TABLE 1: Quality measures for the different fusion methods.

| Data set | Metric | 3D DTCWT-FD | Proposed |
|----------|-------------|-------------|----------|
| Figure 8 | $DQ^{AB/F}$ | 0.1841 | 0.1903 |
| | IFD_MI | 0.7294 | 1.4470 |

artifacts when the fusion procedure is performed. If source images contain noises, the noises may be also transferred into the fusion image together with useful pixels. Therefore, we propose that the recursive filter is first performed on the high-pass subbands to obtain the denoised high-pass coefficients in 3D shearlet transform domain. The high-pass subbands are then combined to employ the fusion rule of the selecting maximum based on 3D PCNN, and the low-pass subband is fused to use the fusion rule of the weighted sum. In this way, the fusion and denoising simultaneously can be achieved for the noised image sequences. Experiments demonstrate that the proposed method improves greatly the quality of the fused image sequence.

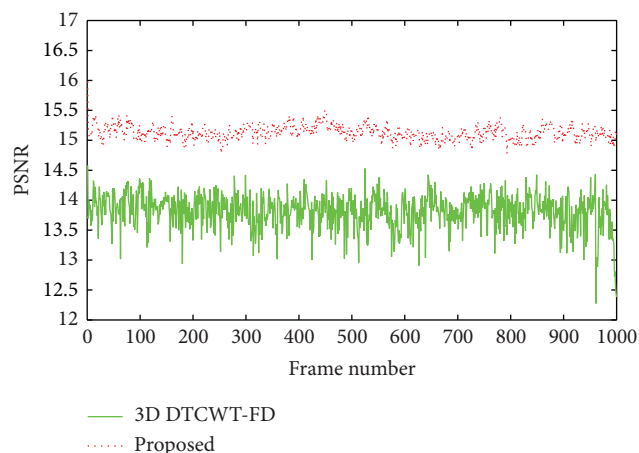


FIGURE 9: PSNR results of the fused and denoising image sequence associated with Figure 8 (frame number: 1000). The red dotted line represents the PSNR scores of our method. The green solid line indicates the PSNR values of the 3D DTCWT-FD.

Conflict of Interests

The authors do not have any actual or potential conflict of interests.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program) 2012CB821200 (2012CB821206) and the National Natural Science Foundation of China (no. 61320106006).

References

- [1] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multi-sensor data fusion: a review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.
- [2] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognition*, vol. 37, no. 9, pp. 1855–1872, 2004.
- [3] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Information Fusion*, vol. 8, no. 2, pp. 131–142, 2007.
- [4] L. Yang, B. L. Guo, and W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, no. 1–3, pp. 203–211, 2008.
- [5] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, 2010.
- [6] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, no. 4, pp. 259–280, 2003.
- [7] V. S. Petrović and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Transactions on Image Processing*, vol. 13, no. 2, pp. 228–237, 2004.
- [8] L. Xu, J. P. Du, Q. Hu, and Q. P. Li, "Feature-based image fusion with a uniform discrete curvelet transform," *International Journal of Advanced Robotic Systems*, vol. 10, pp. 1–11, 2013.
- [9] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 288–291, Santa Barbara, Calif, USA, October 1997.
- [10] G. Xiao, K. Wei, and Z. Jing, "Improved dynamic image fusion scheme for infrared and visible sequence based on image fusion system," in *Proceedings of the 11th International Conference on Information Fusion*, pp. 1–6, Cologne, Germany, July 2008.
- [11] X. H. Zhang, M. Tang, Z. C. Jiang, and R. F. Tong, "Video fusion oriented object motion removing," *Journal of Computational Information Systems*, vol. 8, no. 6, pp. 2317–2324, 2012.
- [12] Q. Zhang, L. Wang, Z. Ma, and H. Li, "A novel video fusion framework using surfacelet transform," *Optics Communications*, vol. 285, no. 13–14, pp. 3032–3041, 2012.
- [13] Q. Zhang, Y. L. Chen, and L. Wang, "Multisensor video fusion based on spatial-temporal salience detection," *Signal Processing*, vol. 93, no. 9, pp. 2485–2499, 2013.
- [14] G. Piella, "Image fusion for enhanced visualization: a variational approach," *International Journal of Computer Vision*, vol. 83, no. 1, pp. 1–11, 2009.
- [15] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, 2010.
- [16] N. Chou, J. Wu, J. B. Bingren, A. Qiu, and K.-H. Chuang, "Robust automatic rodent brain extraction using 3-D pulse-coupled neural networks (PCNN)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2554–2564, 2011.
- [17] P. S. Negi and D. Labate, "3D discrete shearlet transform and video processing," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 2944–2954, 2012.
- [18] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Transactions on Graphics*, vol. 30, no. 4, article 69, 2011.
- [19] R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex," *Neural Computation*, vol. 2, no. 3, pp. 293–307, 1990.
- [20] J. L. Johnson and M. L. Padgett, "PCNN models and applications," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 480–498, 1999.
- [21] V. Petrovic, T. Cootes, and R. Pavlovic, "Dynamic image fusion performance evaluation," in *Proceedings of the 10th International Conference on Information Fusion*, pp. 1–7, Québec, Canada, July 2007.
- [22] I. W. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Wavelets: Applications in Signal and Image Processing X*, vol. 5207 of *Proceedings of SPIE*, no. 2, pp. 607–618, San Diego, Calif, USA, August 2003.

Research Article

Exact CS Reconstruction Condition of Undersampled Spectrum-Sparse Signals

Ying Luo,¹ Qun Zhang,^{1,2} Guozheng Wang,³ and Youqing Bai³

¹ Institute of Information and Navigation, Air Force Engineering University, Xi'an 710077, China

² Key Laboratory for Information Science of Electromagnetic Waves, Ministry of Education, Fudan University, Shanghai 200433, China

³ Institute of Science, Air Force Engineering University, Xi'an 710051, China

Correspondence should be addressed to Qun Zhang; zhangqunnus@gmail.com

Received 15 August 2013; Accepted 18 November 2013

Academic Editor: Feng Gao

Copyright © 2013 Ying Luo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Compressive sensing (CS) reconstruction of a spectrum-sparse signal from undersampled data is, in fact, an ill-posed problem. In this paper, we mathematically prove that, in certain cases, the exact CS reconstruction of a spectrum-sparse signal from undersampled data is impossible. Then we present the exact CS reconstruction condition of undersampled spectrum-sparse signals, which is valuable for digital signal compression.

1. Introduction

In digital signal processing, the Nyquist sampling theorem indicates that the sampling rate must be twice as large as the bandwidth of the analog signal at least for acquiring the intact information of the signal. Restricted by the theorem, it is a challenge to digitize ultrawide bandwidth (UWB) signals because of the unfeasible high sampling rate requirement for the analog-to-digital converter (ADC). On the other hand, the mass sampling data have to be compressed to save the storage, which means that many data are abandoned in the compression processing. Hence, why not to obtain the compressed data of signals directly rather than to sample signal with ultrahigh rate and then abandon most of the samplings?

The emerging *compressive sensing* (CS) theory [1] provides an effective approach to solve this problem, which has attracted much attention recently [2–7]. Consider a signal $\mathbf{x} \in \mathbb{C}^{N \times 1}$ and assume it is sparse on an orthogonal basis $\Psi = \{\psi_i\}$ with K -sparse representation ($K \ll N$) as $\mathbf{x} = \Psi\theta$, where θ is an $N \times 1$ column vector with K nonzero elements. Let Φ denote a measurement matrix and let \mathbf{y} be

the measurements vector of signal \mathbf{x} ; it can be expressed as $\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\theta$, where Φ is a $M \times N$ matrix, M denotes the number of measurements, and $K < M \ll N$. Therefore, the sampling rate is reduced significantly compared with Nyquist rate. Generally, recovery of the signal \mathbf{x} from the measurements \mathbf{y} is ill-posed because $M \ll N$ [8]. However, the CS theory demonstrates that if $\Phi\Psi$ has the *Restricted Isometry Property* (RIP), then it is indeed possible to recover the K largest $\theta_x(i)$ when M is large enough [3, 9]. It is difficult to validate if a measurement matrix satisfies the RIP constraints given in [9] directly, but fortunately, the RIP is closely related to an incoherency between Φ and Ψ , where the rows of Φ do not provide a sparse representation of the columns of Ψ and vice versa [4]. Furthermore, to ensure exact reconstruction, two different K -sparse signals may not be projected by a measurement matrix into the same sampling ensemble [1, 2, 10].

When \mathbf{x} is sparse in spectrum, the N -dimensional inverse discrete Fourier transform (IDFT) matrix (\mathbf{D}_N^{-1}) can be chosen as the sparse representation matrix (Ψ). In this case, an easy way to obtain the compressed data of the signal is to undersample the signal with lower sampling rate than

the Nyquist rate [2]. Therefore the measurement matrix is in fact a partial unit matrix [11]. It is important to investigate the mathematical properties of compressive sensing reconstruction for this kind of undersampled spectrum-sparse signal. In the following, we mathematically prove that the exact CS reconstruction of a spectrum-sparse signal from undersampled data is impossible under certain conditions. In order to reconstruct a spectrum-sparse signal from undersampled data exactly, the corresponding exact CS reconstruction condition is presented, which is valuable for digital signal compression.

2. Inexact CS Reconstruction Cases

When a signal \mathbf{x} is sparse in spectrum, the IDFT matrix \mathbf{D}_N^{-1} and partial unit matrix can be chosen as the sparse representation matrix and the measurement matrix, respectively. By defining the downsampling rate r to be the ratio between the Nyquist rate and the undersampling rate, an inexact CS reconstruction case can be depicted as the following theorem.

Theorem 1. Suppose $\mathbf{x} \in \mathbb{C}^{N \times 1}$ with Nyquist sampling rate f_s is K -sparse in spectrum domain; $\mathbf{y} = \{x_{mr+b} \mid mr+b \leq N, m > 0, m \in \mathbb{N}\}$ is an arbitrary subset of \mathbf{x} , where r is the downsampling rate, $r \geq 2$, and $r \in \mathbb{N}$, b is a constant and $b \in \mathbb{N}$, and \mathbb{N} is the set of all natural numbers. \mathbf{x} cannot be exactly reconstructed from \mathbf{y} by CS.

Proof. According to the expression of \mathbf{y} , we have

$$\mathbf{y}_{M \times 1} = \Phi_{M \times N} \mathbf{x}_{N \times 1} = \Phi_{M \times N} \Psi_{N \times N} \boldsymbol{\theta}_{N \times 1}, \quad (1)$$

where $\boldsymbol{\theta}$ is K -sparse; Φ is the measurement matrix with size $M \times N$:

$$\Phi_{M \times N} = \begin{matrix} \text{Column} & 1 & & r+1 & & 2r+1 & & (M-1)r+1 \\ \text{index} & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \begin{bmatrix} 1 & 0 & \cdots & & & & & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & & & \\ 0 & 0 & \cdots & & & & 0 & \cdots & 1 \end{bmatrix}, \end{matrix}$$

$$\Psi_{N \times N} = \mathbf{D}_N^{-1} = \frac{1}{N} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N^{-1} & W_N^{-2} & \cdots & W_N^{-(N-1)} \\ 1 & W_N^{-2} & W_N^{-4} & \cdots & W_N^{-2(N-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & W_N^{-(N-1)} & W_N^{-2(N-1)} & \cdots & W_N^{-(N-1)^2} \end{bmatrix},$$

$$W_N = \exp\left(-j\frac{2\pi}{N}\right). \quad (2)$$

Equation (1) is equivalent to

$$\boldsymbol{\theta}'_{M \times 1} = \mathbf{D}_M \mathbf{y}_{M \times 1} = \mathbf{D}_M \Phi_{M \times N} \mathbf{D}_N^{-1} \boldsymbol{\theta}_{N \times 1}, \quad (3)$$

where \mathbf{D}_M is the M -dimensional DFT matrix. Because \mathbf{y} is isometrically downsampled from \mathbf{x} , $\boldsymbol{\theta}'$ is also sparse. In (3), we define a new measurement matrix $\Phi_1 = \mathbf{D}_M \Phi_{M \times N}$ and it can be expressed as

$$\Phi_1 = \begin{matrix} \text{Column} & 1 & & r+1 & & 2r+1 & & (M-1)r+1 \\ \text{index} & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 0 & & 0 & W_M & 0 & & 0 & W_M^2 & 0 & & 0 & W_M^{M-1} \\ 1 & 0 & & 0 & W_M^2 & 0 & & 0 & W_M^4 & 0 & & 0 & W_M^{2(M-1)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & W_M^{M-1} & 0 & \cdots & 0 & W_M^{2(M-1)} & 0 & \cdots & 0 & W_M^{(M-1)^2} \end{bmatrix}. \end{matrix} \quad (4)$$

According to CS theory, if $\Phi_1 \mathbf{D}_N^{-1}$ satisfies the RIP, $\boldsymbol{\theta}$ can be exactly reconstructed. Let ϕ_{1a} be the a th row of Φ_1 ; we have

$$\begin{aligned} \phi_{1a}^T &= \mathbf{D}_N^{-1} \mathbf{D}_M \phi_{1a}^T = \mathbf{D}_N^{-1} (\mathbf{D}_M \phi_{1a}^T) \\ &= \mathbf{D}_N^{-1} \left[0, \frac{1 - W_N^{Mr}}{1 - W_M^{a-1} W_N^r}, \dots, \frac{1 - W_N^{(N-1)Mr}}{1 - W_M^{a-1} W_N^{(N-1)r}} \right]^T, \end{aligned} \quad (5)$$

where $W_M = \exp(-j2\pi/M)$. Because $\text{Rank}(\mathbf{D}_N^{-1}) = N$, according to Cramer's Rule, the equation $\mathbf{D}_N^{-1} \boldsymbol{\alpha} = \phi_{1a}^T$ has unique solution; that is,

$$\boldsymbol{\alpha} = \left[0, \frac{1 - W_N^{Mr}}{1 - W_M^{a-1} W_N^r}, \dots, \frac{1 - W_N^{(N-1)Mr}}{1 - W_M^{a-1} W_N^{(N-1)r}} \right]^T. \quad (6)$$

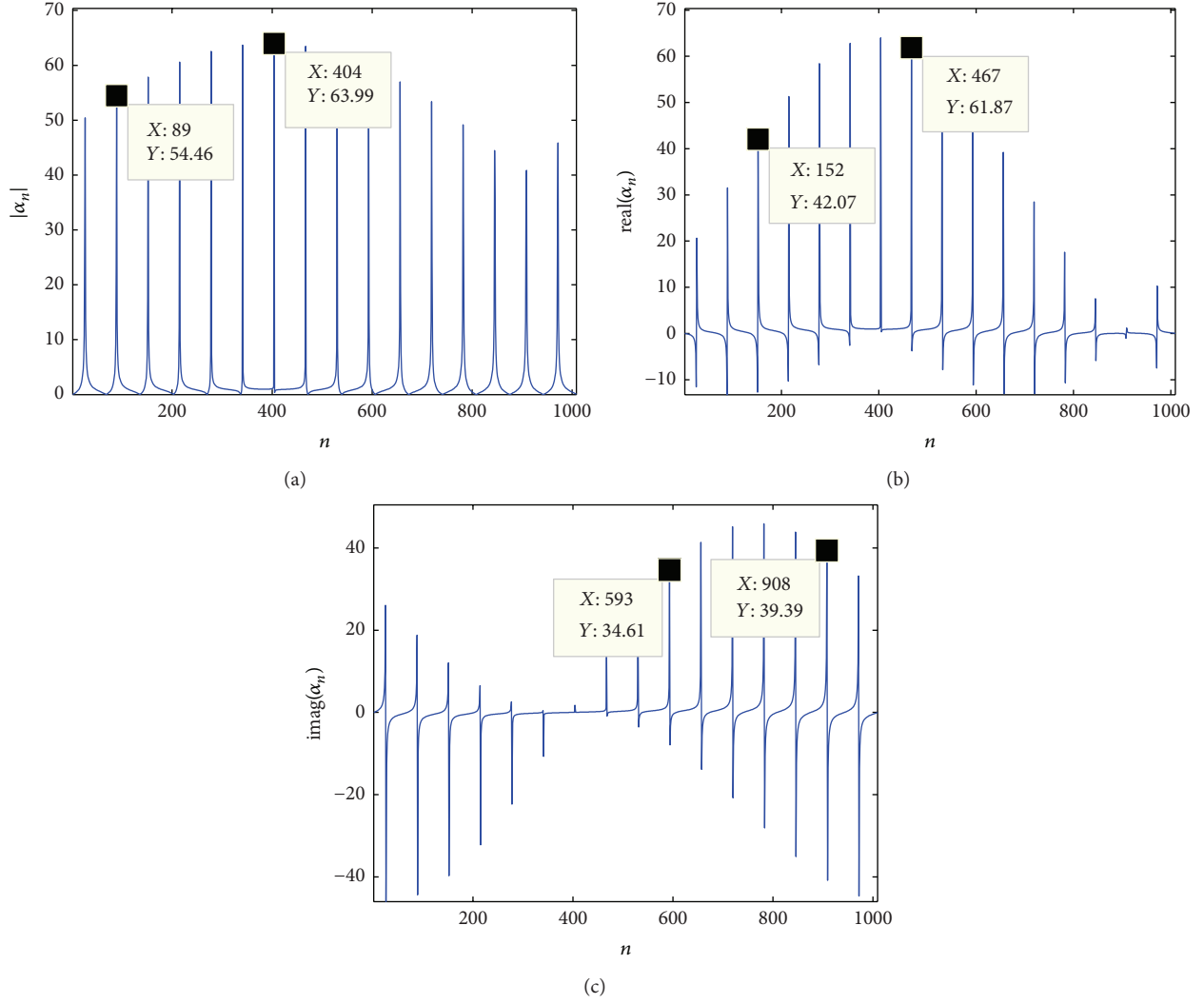


FIGURE 1: The sparsity of α_n when $N = 1024$, $M = 64$, $r = 16$, and $a = 40$. (a) $\{|\alpha_n|\}$; (b) $\{\text{real}(\alpha_n)\}$; (c) $\{\text{imag}(\alpha_n)\}$.

Let α_n be the n th element of α ; it can be obtained

$$|\alpha_n| = \left| \frac{1 - W_N^{(n-1)Mr}}{1 - W_M^{a-1} W_N^{(n-1)r}} \right| \quad (7)$$

$$= \frac{1 - \cos((2\pi/N)(n-1)Mr)}{1 - \cos(2\pi((a-1)/M + ((n-1)r)/N))}.$$

It peaks at

$$\frac{(n-1)r}{N} = -\frac{a-1}{M} + l, \quad l \in \mathbb{Z}, \quad (8)$$

where \mathbb{Z} is set of all nonnegative integers. Because $N = (M-1)r + 1$, we have

$$n = -a + 2 + \frac{(r-1)(a-1)}{Mr} + l \left(M - 1 + \frac{1}{r} \right) \quad (9)$$

and n is a natural number; hence

$$n = \left\lceil -a + 2 + \frac{(r-1)(a-1)}{Mr} + l \left(M - 1 + \frac{1}{r} \right) \right\rceil, \quad (10)$$

where $\lceil \cdot \rceil$ returns the round of (\cdot) . When n does not satisfy (10), $|\alpha_n|$ is relatively quite small. It indicates that $\{|\alpha_n|\}$ is sparse. Similarly, it can be proved that both $\{\text{real}(\alpha_n)\}$ and $\{\text{imag}(\alpha_n)\}$ are also sparse. Therefore, ϕ_{1a} can be sparsely represented by the columns of \mathbf{D}_N^{-1} ; that is, $\Phi_1 \mathbf{D}_N^{-1}$ does not satisfy the RIP, and θ cannot be exactly reconstructed. \square

Figure 1 shows the values of $\{|\alpha_n|\}$, $\{\text{real}(\alpha_n)\}$, and $\{\text{imag}(\alpha_n)\}$ when $N = 1024$, $M = 64$, $r = 16$, and $a = 40$. It is obvious that $\{|\alpha_n|\}$, $\{\text{real}(\alpha_n)\}$, and $\{\text{imag}(\alpha_n)\}$ are sparse and the locations of peaks agree well with the theoretical values determined by (10).

By Theorem 1 and its proof, we can also obtain the following corollary.

Corollary 2. Suppose $\mathbf{x} \in \mathbb{C}^{N \times 1}$ with Nyquist sampling rate f_s is K -sparse in spectrum domain; $\mathbf{x}' = \{x_{mr+b} \mid mr+b \leq N, m > 0, m \in \mathbb{N}\}$ is an arbitrary subset of \mathbf{x} , where r is the downsampling rate, $r \geq 2$, and $r \in \mathbb{N}$ and b is a constant and

$b \in \mathbb{N}$. Let \mathbf{y} be an arbitrary subset of \mathbf{x}' ; then \mathbf{x} cannot be exactly reconstructed from \mathbf{y} by CS.

Proof. Assume the length of \mathbf{x}' is M and we have

$$\mathbf{y} = \Phi_2 \mathbf{x}' = \Phi_2 \Phi \mathbf{D}_N^{-1} \boldsymbol{\theta}, \quad (11)$$

where Φ_2 is a partial unit matrix. Assuming the size of Φ_2 is $L \times M$ ($L \leq M$), the solution set of (1) is a subset of the solution set of (11). According to Theorem 1, the solutions of (1) are not determined; therefore, $\boldsymbol{\theta}$ in (11) cannot be determined and \mathbf{x} cannot be exactly reconstructed. \square

Corollary 2 indicates that, if the set of undersamplings is only a subset of the set of signals' samplings with sampling rate lower than the Nyquist rate, the signal cannot be exactly reconstructed from these undersamplings by CS. Therefore, when designing the ADC with random sampling space, the ADC should better possess the capability with the minimum sampling space of $1/f_s$, where f_s is the Nyquist rate. Taken in this sense, the high sampling rate requirement for ADC is indeed not suppressed even though the CS theory is utilized.

3. Exact CS Reconstruction Condition

In the following, we present the exact CS reconstruction condition of undersampled spectrum-sparse signals.

Theorem 3. Suppose $\mathbf{x} \in \mathbb{C}^{N \times 1}$ with Nyquist sampling rate f_s is K -sparse in spectrum domain; and the frequency indexes of K nonzero points in spectrum are f_h ($h = 1, 2, \dots, K$). Undersampling \mathbf{x} with rate $f_s/r_1, f_s/r_2, \dots, f_s/r_i, \dots, f_s/r_I$ ($i = 1, 2, \dots, I, r_i \geq 2$, and $r_i \in \mathbb{N}$), respectively, all the samples consist of $\mathbf{y}_{M \times 1}$. The necessary and sufficient condition for \mathbf{x} exactly reconstructed from \mathbf{y} by CS is

$$\begin{aligned} & \left\{ f_h + \frac{k_1 f_s}{r_1} \mid h = 1, 2, \dots, K; k_1 \in \mathbb{Z}; \left| f_h + \frac{k_1 f_s}{r_1} \right| \leq \frac{f_s}{2} \right\} \\ & \cap \left\{ f_h + \frac{k_2 f_s}{r_2} \mid h = 1, 2, \dots, K; k_2 \in \mathbb{Z}; \left| f_h + \frac{k_2 f_s}{r_2} \right| \leq \frac{f_s}{2} \right\} \\ & \cap \dots \\ & \cap \left\{ f_h + \frac{k_I f_s}{r_I} \mid h = 1, 2, \dots, K; k_I \in \mathbb{Z}; \left| f_h + \frac{k_I f_s}{r_I} \right| \leq \frac{f_s}{2} \right\} \\ & = \{f_h \mid h = 1, 2, \dots, K\}. \end{aligned} \quad (12)$$

Proof. Reconstructing \mathbf{x} from \mathbf{y} is in fact to solve the following underdetermined equation system:

$$\begin{aligned} \mathbf{y}_1 &= \Phi_1 \mathbf{x} = \Phi_1 \mathbf{D}_N^{-1} \boldsymbol{\theta}, \\ \mathbf{y}_2 &= \Phi_2 \mathbf{x} = \Phi_2 \mathbf{D}_N^{-1} \boldsymbol{\theta}, \\ &\vdots \\ \mathbf{y}_I &= \Phi_I \mathbf{x} = \Phi_I \mathbf{D}_N^{-1} \boldsymbol{\theta}, \end{aligned} \quad (13)$$

where Φ_i is the measurement matrix according to the undersampling rate f_s/r_i ; \mathbf{y}_i is composed of the undersamplings

of \mathbf{x} with rate f_s/r_i . Assume the frequency indexes of the solution set of the i th equation are $\{\hat{f}_{hi}\}$; because \mathbf{y}_i is the isometric downsampling from \mathbf{x} , we have

$$\begin{aligned} & \{\hat{f}_{hi}\} \\ & \subseteq \left\{ f_h + \frac{k_i f_s}{r_i} \mid h = 1, 2, \dots, K; k_i \in \mathbb{Z}; \left| f_h + \frac{k_i f_s}{r_i} \right| \leq \frac{f_s}{2} \right\}. \end{aligned} \quad (14)$$

Hence, the solution set of (13) is

$$\begin{aligned} & \cap \{\hat{f}_{hi}\} \subseteq \cap \left\{ f_h + \frac{k_i f_s}{r_i} \mid h = 1, 2, \dots, K; k_i \in \mathbb{Z}; \right. \\ & \left. \left| f_h + \frac{k_i f_s}{r_i} \right| \leq \frac{f_s}{2} \right\}. \end{aligned} \quad (15)$$

Only if (12) holds true, under the constraint of K -sparse, it yields $\cap \{\hat{f}_{hi}\} = \{f_h \mid h = 1, 2, \dots, K\}$; therefore, \mathbf{x} is exactly reconstructed. \square

Theorem 3 indicates that it is possible to reconstruct a spectrum-sparse signal exactly from its multirate downsamplings. It means that, when condition (12) is satisfied, the multirate downsampling can be used to compress the data of digital signals. The conclusion is valuable to the design of the ADC when the analog signal is a priori known sparse in spectrum.

In order to validate Theorem 3, an experiment is given as follows. In the experiment, the signal $x(t)$ with Nyquist sampling rate $f_s = 1$ Hz is expressed as $x(t) = \exp(0.2 \times 2\pi t) + 0.8 \exp(0.45 \times 2\pi t) + 0.6 \exp(0.3 \times 2\pi t)$, $t \in [0, 1023]$. The sparsity of $x(t)$ in spectrum domain is 3 and $f_h = \{0.2, 0.45, 0.3\}$. The spectrum of $x(t)$ is shown in Figure 2(a). Let $r_1 = 7$ and $r_2 = 10$. Therefore, we have

$$\begin{aligned} & \left\{ f_h + \frac{k_1 f_s}{r_1} \mid h = 1, 2, \dots, K; k_1 \in \mathbb{Z}; \left| f_h + \frac{k_1 f_s}{r_1} \right| \leq \frac{f_s}{2} \right\} \\ & = \left\{ 0.2 + \frac{k_1}{7} \mid k_1 \in \mathbb{Z}; \left| 0.2 + \frac{k_1}{7} \right| \leq \frac{1}{2} \right\} \\ & \cap \left\{ 0.45 + \frac{k_1}{7} \mid k_1 \in \mathbb{Z}; \left| 0.45 + \frac{k_1}{7} \right| \leq \frac{1}{2} \right\} \\ & \cap \left\{ 0.3 + \frac{k_1}{7} \mid k_1 \in \mathbb{Z}; \left| 0.3 + \frac{k_1}{7} \right| \leq \frac{1}{2} \right\} \\ & = \left\{ 0.2, 0.2 + \frac{1}{7}, 0.2 + \frac{2}{7}, 0.2 - \frac{1}{7}, 0.2 - \frac{2}{7}, \right. \\ & \quad \left. 0.2 - \frac{3}{7}, 0.2 - \frac{4}{7}, 0.45, 0.45 - \frac{1}{7}, \right. \\ & \quad \left. 0.45 - \frac{2}{7}, 0.45 - \frac{3}{7}, 0.45 - \frac{4}{7}, 0.45 - \frac{5}{7}, \right. \\ & \quad \left. 0.45 - \frac{6}{7}, 0.3, 0.3 + \frac{1}{7}, 0.3 - \frac{1}{7}, 0.3 - \frac{2}{7}, \right. \\ & \quad \left. 0.3 - \frac{3}{7}, 0.3 - \frac{4}{7}, 0.3 - \frac{5}{7} \right\}, \end{aligned}$$

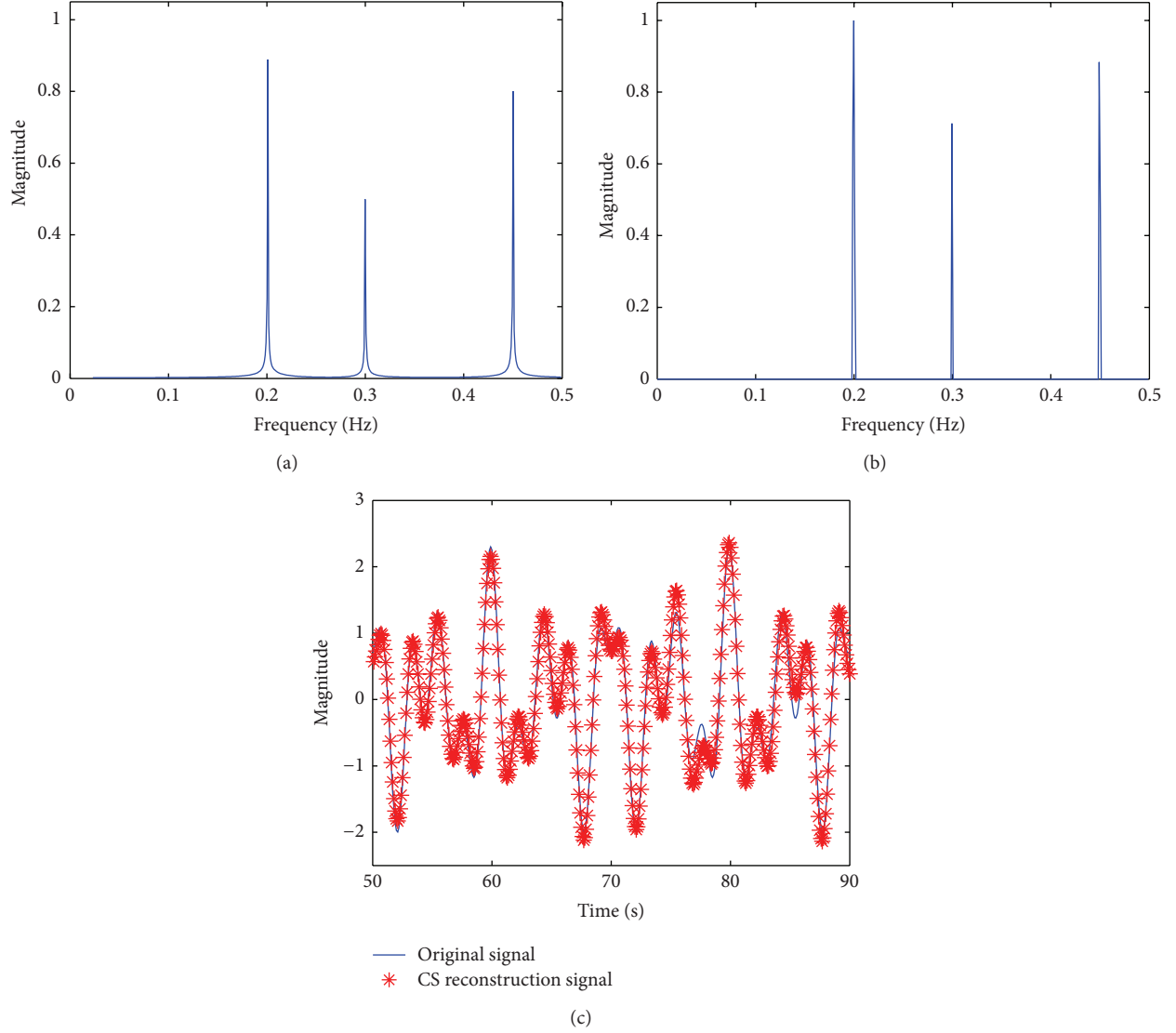


FIGURE 2: Experiment for validation of Theorem 3. (a) The spectrum of the signal $x(t)$ with Nyquist sampling rate; (b) the spectrum of the reconstructed signal by CS; (c) the comparison of $x(t)$ and the reconstructed signal in time domain.

$$\begin{aligned}
 & \left\{ f_h + \frac{k_2 f_s}{r_2} \mid h = 1, 2, \dots, K; k_2 \in \mathbb{Z}; \left| f_h + \frac{k_2 f_s}{r_2} \right| \leq \frac{f_s}{2} \right\} \\
 &= \left\{ 0.2 + \frac{k_2}{10} \mid k_2 \in \mathbb{Z}; \left| 0.2 + \frac{k_2}{10} \right| \leq \frac{1}{2} \right\} \\
 &\cap \left\{ 0.45 + \frac{k_2}{10} \mid k_2 \in \mathbb{Z}; \left| 0.45 + \frac{k_2}{10} \right| \leq \frac{1}{2} \right\} \\
 &\cap \left\{ 0.3 + \frac{k_2}{10} \mid k_2 \in \mathbb{Z}; \left| 0.3 + \frac{k_2}{10} \right| \leq \frac{1}{2} \right\} \\
 &= \{-0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, \\
 &\quad 0.4, 0.5, 0.45, 0.35, 0.25, 0.15, 0.05, -0.05, \\
 &\quad -0.15, -0.25, -0.35, -0.45\}.
 \end{aligned}
 \tag{16}$$

Obviously, the intersection of (16) is $f_h = \{0.2, 0.45, 0.3\}$. Hence, $x(t)$ can be exactly reconstructed from the down-samplings. In the experiment, we choose the orthogonal matching pursuit (OMP) algorithm [12] to reconstruct $x(t)$ from the downsamplings. The spectrum of the reconstructed signal by CS is shown in Figure 2(b), which is very close to that in Figure 2(a). The comparison of the original signal $x(t)$ with Nyquist sampling rate and the reconstructed signal in time domain is also given in Figure 2(c). From the figure, it can be found that the reconstructed signal is close to the original signal, which validates the correctness of Theorem 3.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China under Grant 61201369 and 61172169 and in part by the Natural Science Foundation Research Program of Shaanxi Province under Grant 2013JQ8008.

References

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [3] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [4] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *IEEE Radar Conference*, pp. 128–133, Boston, Mass, USA, 2007.
- [5] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.
- [6] Y. Zhang, "Theory of compressive sensing via \mathcal{L}_1 -minimization: a non-RIP analysis and extensions," *Journal of the Operations Research Society of China*, vol. 1, no. 1, pp. 79–105, 2013.
- [7] T. Strohmer, "Measure what should be measured: progress and challenges in compressive sensing," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 887–893, 2012.
- [8] B. Adcock, A. C. Hansen, E. Herrholz, and G. Teschke, "Generalized sampling: extension to frames and inverse and ill-posed problems," *Inverse Problems*, vol. 29, no. 1, Article ID 015008, p. 27, 2013.
- [9] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [10] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [11] Y. Luo, Q. Zhang, W. Hong, and Y.-R. Wu, "Waveform design and high-resolution imaging of cognitive radar based on compressive sensing," *Science China Information Sciences*, vol. 55, no. 11, pp. 2590–2603, 2012.
- [12] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

Research Article

A Novel Optimization-Based Approach for Content-Based Image Retrieval

Manyu Xiao,¹ Jianghu Lu,^{1,2} and Gongnan Xie³

¹ Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

² Computer Science and Technology, Beihang University, Beijing 100191, China

³ Engineering Simulation and Aerospace Computing (ESAC), Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

Correspondence should be addressed to Manyu Xiao; manyuxiao@gmail.com

Received 4 October 2013; Accepted 5 December 2013

Academic Editor: Feng Gao

Copyright © 2013 Manyu Xiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Content-based image retrieval is nowadays one of the possible and promising solutions to manage image databases effectively. However, with the large number of images, there still exists a great discrepancy between the users' expectations (accuracy and efficiency) and the real performance in image retrieval. In this work, new optimization strategies are proposed on vocabulary tree building, retrieval, and matching methods. More precisely, a new clustering strategy combining classification and conventional K -Means method is firstly redefined. Then a new matching technique is built to eliminate the error caused by large-scaled scale-invariant feature transform (SIFT). Additionally, a new unit mechanism is proposed to reduce the cost of indexing time. Finally, the numerical results show that excellent performances are obtained in both accuracy and efficiency based on the proposed improvements for image retrieval.

1. Introduction

Nowadays, content-based image retrieval (CBIR) has more and more applications and constitutes one of the core problems in computer vision. Its features were thoroughly discussed by Smeulders et al. [1]. One of the most popular methods that yield results of content-based image retrieval is based on visual contents of an image. The visual features of images, such as color [2], texture [3], and shape features [4] have been extensively explored to represent and index image contents, resulting in a collection of research prototypes and commercial systems [5, 6]. Therefore, the performance of a CBIR system mainly depends on the particular image representation and similarity matching function employed [7]. Due to the rapid development and improvement of the internet, image capture devices and computer hardware cause the problem of storage and manipulation of images [8]. That is the reason that the relevant techniques developed by Google Inc. and Baidu Inc. did not perform adequately. The main limitation occurs on either retrieval accuracy or real-time or sometimes both. In order to overcome this limitation, in this work a novel optimization-based approach for content-based image

retrieval is proposed. The conventional procedure of image retrieval is firstly introduced, as shown in Figure 1. It can be divided into three parts: the vocabulary tree building, the storage of test images, and their retrieval. The descriptions of these three parts are briefly discussed as follows.

1.1. Building Vocabulary Tree. In state-of-art techniques, a tree structure is usually built to store the image database. In other words, a training set is needed to get a discriminative and representative tree. The training set is a group of images, which are first transformed into SIFT [9] (scale-invariant feature transform) 128-dimensional descriptor vectors. SIFT features are distinctive invariant features that are used to robustly describe and match digital image content among different views of a scene. While being invariant to scale and rotation, and robust to other image transforms, the SIFT feature description of an image is typically large and slow to be computed [10] (curse of dimensionality). After that, a vocabulary tree [11] must be built. The traditional classification of the large features of image databases is often carried out by the Hierarchical K -Means method (HKM). Assuming

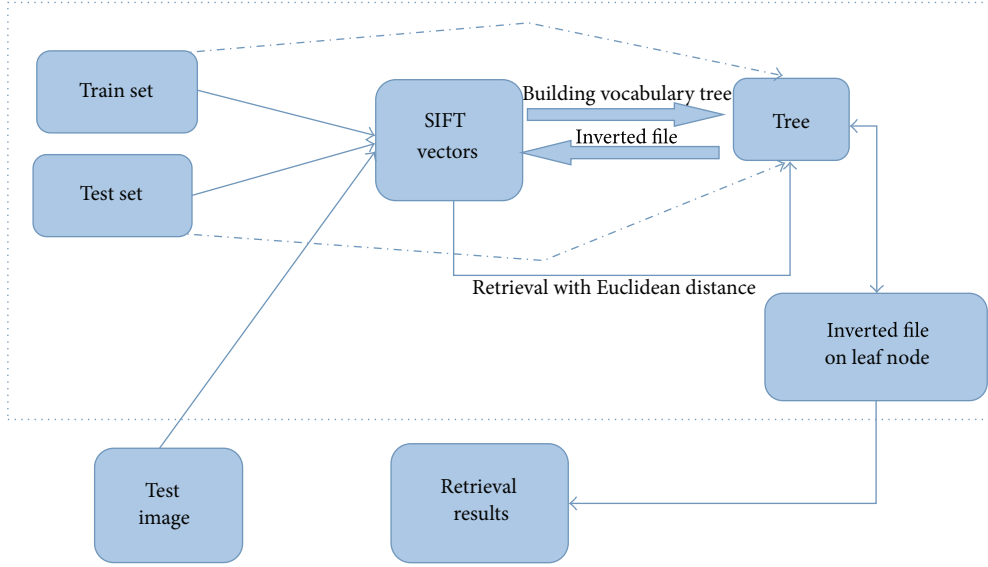


FIGURE 1: Content-based image retrieval procedure.

a priori N the number of branches and H the height of the tree, the all descriptors can be clustered into N parts, and N cluster centers can be obtained from N nodes at the first level, and then each part eventually reclustered into N new subparts to get all new N^2 nodes. This process is repeated until converged to the defined height and obtention of a complete tree.

1.2. Storage of Test Image and Database Information-Inverted File. Once a well-organized data structure is built, the image database can be stored. From the image database, the SIFT descriptors must also be extracted [12, 13]. For each image in the database, all its descriptors undergo the same classification as follows: comparison with the children nodes of the root node by Euclidean distance and then selection of the nearest one as new root node. This process has to be repeated until the leaf node is reached. In order to match and retrieve the image at the next search, a shortest path to the leaf node has to be constructed. This means an inverted file [11, 14, 15] is needed to be built for identifying the relationship between the image database and the test image. When a descriptor of the test image reaches a leaf node, the relevant inverted-file will record the corresponding information of this descriptor.

After all descriptors of a test image are stored in the vocabulary tree, a weight of a leaf node is calculated based on TF-IDF strategy so as to test the effectiveness of different leaf nodes accurately by the following formula:

$$w_i = \log \frac{N}{\sum_{j=1}^m f(id_{ij})}, \quad (1)$$

$$f(id_{ij}) = \begin{cases} 1 & id_{ij} > 0, \\ 0 & id_{ij} = 0, \end{cases}$$

where N is the number of images in the database and m is the dimension of the vectors in the inverted-file. If there only

exist several images descriptors obtained by $\sum_{i=1}^m f(id_{ij})$ and its value is small, this means this leaf node is discriminative. This means the leaf node is discriminative and has a good retrieval feature. Otherwise, this value would be much bigger.

1.3. Retrieval and Rank. After organizing the information of test images and database images, two vectors can be obtained for the test image and the database images as follows.

For the test image,

$$S = (w_1 \times n_1, w_2 \times n_2, \dots, w_n \times n_n). \quad (2)$$

For the database images,

$$D_i = (w_1 \times l_{i1}, w_2 \times l_{i2}, \dots, w_n \times l_{in}), \quad (3)$$

where w_i is the weight of leaf node and n_i the number of test image nodes reaching the i th leaf node, while l_{ij} is the number of the i th database image nodes found in the j th leaf node during building image base.

Now, the ranking results can be obtained by the following formula:

$$h(S, D_i) = \left\| \frac{S}{\|S\|} - \frac{D_i}{\|D_i\|} \right\|. \quad (4)$$

$\|\cdot\|$ usually refers to L_1 -norm or L_2 -norm.

With the usual approach, accuracy issues often occur. First, as numbers of the test images increase, more noises and clutters are brought into the information database, which undoubtedly results in decreasing the retrieval accuracy. Secondly, more information in the image database leads to more time required to search the similar images from the database, which usually cannot satisfy the real-time demands. Finally, after dozens of trial-and-error tests, it is found that the norms of calculating the match degree cannot remove the magnitude of different image SIFT numbers, which reveals a loss

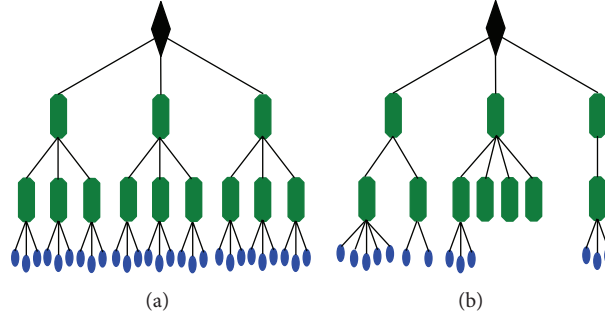


FIGURE 2: Traditional vocabulary tree (a) and new vocabulary tree (b).

of accuracy. That is what motivates to propose the improvements detailed hereafter.

The paper is organized as follows: three improvements are described in Section 2. Then, the image retrieval application is tested in Section 3, followed by the discussion and conclusion in Section 4.

2. Three Improvements

2.1. Improving the Vocabulary Tree Building. From the process described above, it is known that the height and the branch number of the tree are both predefined, namely, a complete tree (see Figure 2(a)). After building each clustering, all the descriptors will be divided equivalently into several parts. Due to differences of distances, there may be different numbers of descriptors in different parts, and there may even happen that a part only includes a few descriptors. On the other hand, it might also occur that the biggest distance is already small in a certain part, but due to the limit of pre-defined level and branch, this part has to be divided continuously. These are not affordable.

In practical applications, the quantities of information in different test sets are different, and different trees are therefore needed. When the tree need not even be a complete tree, the conventional method certainly leads to some errors. In order to reduce or even eliminate these errors, the conventional K -Means processes and classification are combined to make sure that the height and branch number of vocabulary tree are defined automatically.

The proposed technique called Hierarchical Classification method (HCM) is done with two thresholds: one is for the number of descriptors in a part and the other for the distance inside a part. These two thresholds can determine when the clustering operations terminate; thus we will not know how many levels the tree has and will not know how many children nodes a parent node owns. The structure of two different trees can be shown as follows, respectively (Figure 2(b) improved tree): this new model provides not only improvements of efficiency but also in precision.

2.2. Shorter Time Spent on Retrieval. In previous works, a classification was often obtained by Euclidean distance of the children nodes of the root node, not the information of root node directly in the left of Figure 3. Obviously, it took much time for calculation. This distance was not bigger than the

sum of the distances between descriptor-root node and the distance obtained from root node-relevant child node. It can be written as follows:

$$\|P - Q\| \leq \|P - R\| + \|R - Q\|, \quad (5)$$

where P denotes a descriptor, Q is one child node of root node, and R is root node.

As the distances between root node and its children nodes are all calculated in advance and reserved in the root node position, the proposed clustering technique consists in finding the next children tree by using only the first term of (5). The second part is used for next clustering rank shown in Figure 3(b).

Define M as the height of vocabulary tree, and K is the number of child nodes at each rank and T is computational time for each clustering. The total time with the traditional classification to find the nearest leaf node is about $(M - 1) \times K \times T$, while the improved method only takes a time of $M \times T$; the later consumes about only $M / ((M - 1) \times K)$ of time needed before, which significantly reduces the retrieval time.

2.3. Improvement on Scoring Mechanism. In the traditional method, $\|\cdot\|$ usually is L_1 -norm or L_2 -norm. However, for lots of practical experiments, it is found that the results are not very reasonable because of the different memory requests. When the image is more complex, it is more possible to match with retrieval image as the large number of features (large number of eigenvectors of dimension 128). On the contrary, it is of poor effect for simple images. The third improvement will concentrate on dealing with this problem. The following *unit norm* is proposed to eliminate the weight of test image and great different database image memory:

$$h(S, D_i) = \left\| \frac{S}{N} - \frac{D_i}{N_i} \right\|, \quad (6)$$

where N stands for the number of test image vectors and N_i the number of the i th database image. This *unit norm* can make sure that different images stay at the same level; a more reasonable score will hence be produced. This new *unit norm* is called *unit norm*.

Important notation: when programming, there may be millions of pictures in the image resource, and there will be

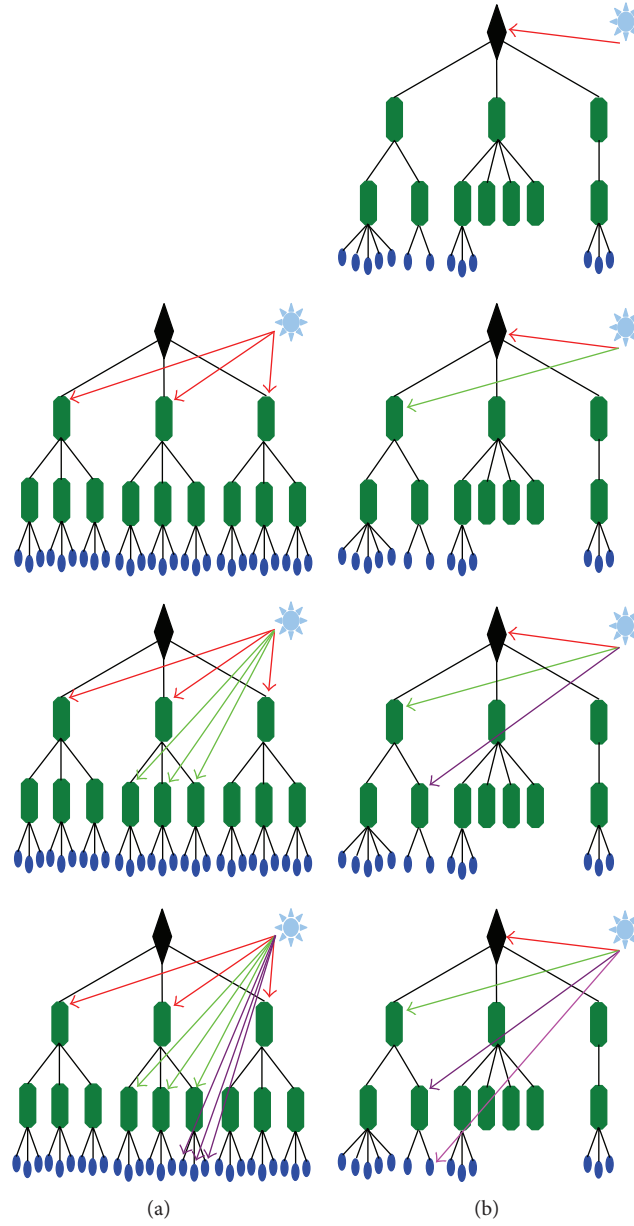


FIGURE 3: Traditional classification (a) and new clustering technique (b).

even more than 106 leaf nodes, while for indexing each image in the database, there will be thousands of dimensions equal to 0 in vectors. In order to save the memory space, assigning storage dynamically is proposed.

3. Test Examples

Ukbench image database contains 2550 groups of images, and every group includes 4 similar images. More precisely, these 4 images with a much similar characters are snapshots in a same image but in different illumination intensity and orientation. Analyzing the expectations of users, the following strategies are taken: indexing one image from the database images, if three of four similar images can be exhibited in the ranking

10 results, this image retrieval implementation is a successful process. The index frequency is calculated by

$$f_i = \frac{n_i}{4}, \quad n_i = 0, 1, 2, 3, 4, \quad (7)$$

where n_i is the number of similar images shown in ranking 10 images and f_i an index accuracy for i th test image.

The average of this accuracy is finally used to test the effectiveness of different *unit norm*. Different quantities of images retrieval are shown in Figure 4.

From Figure 4, it is observed that a more reasonable score h will be obtained with *unit norm* (that means matching degree is bigger between similar images and smaller between irrelevant ones). The index accuracy is thence much better with *unit norm* than with L_1 -norm and L_2 -norm.

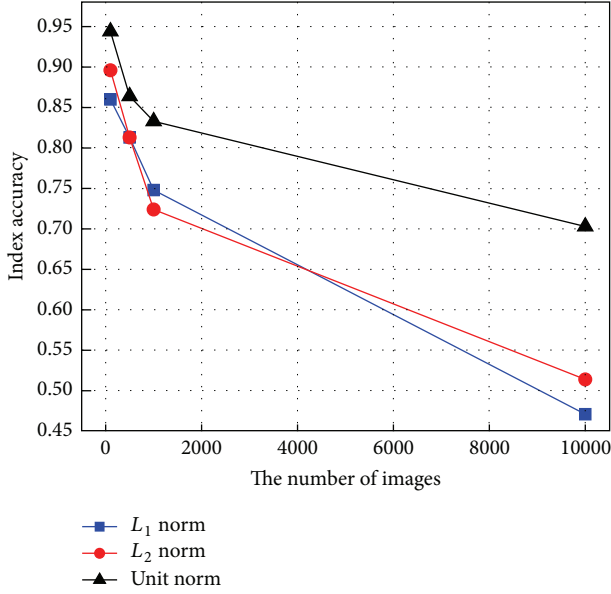


FIGURE 4: Comparison of three different norms.

TABLE 1: Improvements on discarding invalid descriptors.

| Discard | Quantity | | | |
|---------|----------|-------|--------|-------|
| | 100 | 500 | 1000 | 10000 |
| No | 90.6% | 82.3% | 76.2% | 53.9% |
| Yes | 94.3% | 86.3% | 83.25% | 70.2% |

TABLE 2: Comprehensive comparison of new and traditional mechanism.

| Method | Quantity | | | |
|--------|----------|-------|--------|-------|
| | 100 | 500 | 1000 | 10000 |
| HKM | 89.5% | 84.4% | 80.25% | 62.3% |
| HCM | 94.3% | 86.3% | 83.25% | 70.2% |

In building test image base, taking the strategy of discarding into action, much better performance is achieved as listed in Table 1.

Based on these two important improvements, the efficiencies of classification with the famous HKM method and HCM (the proposed method) are compared. The results on Table 2 show that the proposed improvements are more feasible and efficient.

4. Conclusions and Discussion

In this work, three improvements are proposed during the content-based image retrieval: strategy in image classification, mechanism to calculate the Euclidean distance of eigenvectors between source images and research image, and development of the inverse file. As a result, the index accuracy can be greatly enhanced. Furthermore, we can get a faster index procedure, which satisfies the real-time image retrieval quite well. In the point of theoretical view, the proposed technique takes only about one-sixth time of traditional method

needed. In the future work, it is necessary to verify the efficiency of proposed improvement in a practical situation as the time is really very short in the above example.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has been supported by the Basic Project Foundation of Northwestern Polytechnical University (Grants no. JC20120241) and by the National Natural Science Foundation of China (Grants no. 11302173).

References

- [1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.
- [2] R. O. Stehling, M. A. Nascimento, and A. X. Falcao, "On shapes of colors for content-based image retrieval," in *Proceedings of the ACM International Workshop on Multimedia Information Retrieval (ACM MIR '00)*, pp. 171–174, 2000.
- [3] M. Flickner, H. Sawhney, W. Niblack et al., "Query by image and video content: the QBIC system," *Computer*, vol. 28, no. 9, pp. 23–32, 1995.
- [4] D. S. Zhang and G. Lu, "Generic Fourier descriptors shape-based image retrieval," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '02)*, vol. 1, pp. 425–428, 2002.
- [5] F. Jing, M. Li, H. Zhang, and B. Zhang, "An effective region-based image retrieval framework," in *Proceedings of the 10th ACM International Conference on Multimedia*, pp. 456–465, December 2002.
- [6] A. Gupta and R. Jain, "Visual information retrieval," *Communications of the ACM*, vol. 40, no. 5, pp. 71–79, 1997.
- [7] H. B. Kekre, S. D. Thepade, A. Athawale, A. Shah, P. Verlekar, and S. Shirke, "Energy compaction and image splitting for image retrieval using kekre transform over row and column feature vectors," *International Journal of Computer Science and Network Security*, vol. 10, no. 1, 2010.
- [8] M. K. Mandal, F. Idris, and S. Panchanathan, "Critical evaluation of image and video indexing techniques in the compressed domain," *Image and Vision Computing*, vol. 17, no. 7, pp. 513–529, 1999.
- [9] D. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [10] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 2161–2168, 2006.
- [11] P. Indyk and R. Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC '98)*, pp. 604–613, 1998.
- [12] J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," in *Proceedings of the 9th*

IEEE International Conference on Computer Vision (ICCV '03), pp. 1470–1477, Nice, France, October 2003.

- [13] J. S. Beis and D. G. Lowe, “Shape indexing using approximate nearest-neighbour search in high-dimensional spaces,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pp. 1000–1006, June 1997.
- [14] J. Zobel, A. Moffat, and K. Ramamohanarao, “Inverted files versus signature files for text indexing,” *ACM Transactions on Database Systems*, vol. 23, no. 4, pp. 453–490, 1998.
- [15] A. C. Berg, T. L. Berg, and J. Malik, “Shape matching and object recognition using low distortion correspondences,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 26–33, June 2005.

Research Article

Adaptive Self-Occlusion Behavior Recognition Based on pLSA

Hong-bin Tu, Li-min Xia, and Lun-zheng Tan

School of Information Science and Engineering, Central South University, ChangSha, HuNan 410075, China

Correspondence should be addressed to Li-min Xia; xlm@mail.csu.edu.cn

Received 31 July 2013; Accepted 24 October 2013

Academic Editor: Feng Gao

Copyright © 2013 Hong-bin Tu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Human action recognition is an important area of human action recognition research. Focusing on the problem of self-occlusion in the field of human action recognition, a new adaptive occlusion state behavior recognition approach was presented based on Markov random field and probabilistic Latent Semantic Analysis (pLSA). Firstly, the Markov random field was used to represent the occlusion relationship between human body parts in terms an occlusion state variable by phase space obtained. Then, we proposed a hierarchical area variety model. Finally, we use the topic model of pLSA to recognize the human behavior. Experiments were performed on the KTH, Weizmann, and Humaneva dataset to test and evaluate the proposed method. The compared experiment results showed that what the proposed method can achieve was more effective than the compared methods.

1. Introduction

Automatic recognition of human actions from video is a challenging problem that has attracted the attention of researchers in the recent decades. It has applications in many areas such as entertainment, virtual reality, motion capture, sport training [1], medical biomechanical analysis, ergonomic analysis, human-computer interaction, surveillance and security, environmental control and monitoring, and patient monitoring systems.

Occlusion state recognition has been traditionally tackled by applying statistical prediction and inference methods. Unfortunately, basic numerical methods have proved to be insufficient when dealing with complex occlusion scenarios that present interactions between objects (e.g., occlusions, unions, or separations), modifications of the objects (e.g., deformations), and changes in the scene (e.g., illumination). These events are hard to manage and frequently result in tracking errors, such as track discontinuity, inconsistent track labeling.

The Pictorial structure method [2], which represents the human body as a set of linked rectangular regions, does not take occlusion into account. Sigal et al. [3] argue that the self-occlusion problem can be reduced by an occlusion-sensitive likelihood model. This works well if the occlusionstates (i.e., the depth ordering of parts) is known; for example, if it

is specified at the start of the motion and then does not change over time. But, in practice, the depth order of object parts—for example, right arm, torso. Estimating 2D human pose is difficult because of image noises (e.g., illumination and background clutter), self-occlusion, and the varieties of human appearances (i.e., clothing, gender, and body shape) [3–5]. Estimating and tracking 3D human pose is even more challenging because of the large state space of the human body in 3D and our indirect knowledge of 3D depth [6]. In contrast, our approach focuses on self-occlusion. While all of the above methods are modeled to estimate poses from still images, there exists only limited research on the same task in videos. Guo et al. [7] applied the BOW model with human action recognition in video sequence. Niebles et al. [8] successfully applied this model to classify the video sequence of the human action. Wang and Mori [9] assigned each frame of an image sequence to a visual word by analyzing the motion of the person it contains. Sy et al. [10] applied the CRF with a hidden state structure to predict the label of the whole sequence of human gestures. Sigal et al. [3] modeled self-occlusion handling in the PS framework as a set of constraints on the occluded parts, which are extracted after performing background subtraction which renders it unsuitable for dynamic background scenes.

Our work follows literatures [3, 7, 9, 11] by producing a framework for articulated pose estimate-on robust to

cluttered backgrounds and self-occlusion without relying on background subtraction models. The step of rectifying occluded body parts via a GPR model is inspired by recent work by Asthana et al. [12] who used GPR for modeling parametric correspondences between face models of different people. Our problem is more difficult because the human body includes more parameters to be rectified and has more degrees of freedom than faces.

In order to overcome the shortcomings mentioned above, we propose an adaptive self-occlusion state recognition method that estimates not only everybody configuration but also the occlusion states of body parts.

Firstly, the Markov random field was used to represent the occlusion relationship between human body parts in terms of occlusion state variable by phase space obtained. Then, we proposed a hierarchical area variety model. Finally, we inferred human behavior by pLSA. Experiments on Human Eva data set were performed to test and evaluate the proposed algorithm. The experiment results have shown that the proposed method is effective in action recognition.

2. Human Trajectory Reconstruction

A tree structure movement of the human body skeleton structure is used by creating visual invariant model [13], the human body is divided into 15 key points; namely, 15 joint point represents the human body structure, and the 15 joints trajectory represents the human body behavior and then uses Markov random field (MRF) by calculating the observation, spatial relations, and the motion relationship and ultimately determines the occlusion positions of the body joints and restores the missing trajectory. Specific steps described below.

The Markov random field (MRF) was used with a state variable representing the occlusion relationship between body parts. Formally, the MRF was a graph $G = (V, E)$, where V was the set of nodes and E was the set of edges. The graph nodes V represented the state of a human body part and graph edges E model the relationships between the parts [11]. The probability distribution over this graph was specified by the set of potentials defined over the set of edges. The MRF structural parameters are defined as follows: $X_i = (x_i, y_i, z_i)$: The i th joint point coordinates; $X = \{X_1, X_2, \dots, X_{15}\}$: extract the key points of the body 15; $\gamma(X_i)$ ($i \leq 15$): the i th joints visible parts, this parameter is used to determine occlusion relation between nodes. When occlusion occurred, trajectories intersected between

$$X_i(X_i(x_i, y_i, z_i)), \quad X_j(X_j(x_j, y_j, z_j)); \quad (1)$$

$\Lambda = \{\Lambda_{ij}\}$ ($i \leq 15, j \leq 15$): the occlusion relation among the 15 body joints. When $\Lambda_{i,j} = 0$, the i th and j th joints do not occluded. When $\Lambda_{i,j} = 1$, the i th occluded j th. When $\Lambda_{i,j} = -1$, the j th occlude i th; $\lambda_i = \{\lambda_1, \dots, \lambda_{15}\}$: the i th occlude joints node; then, potential of kinematic relationship is calculated as follows:

$$\psi_{ij}^K(X_i, X_j) = N(d(x_i, x_j); \mu_k, \delta_K) f(\theta_i, \theta_j). \quad (2)$$

This function indicates the position of two adjacent joints, and the angles among joints.

$d(x_i, x_j)$ is the Eucidean distance between two adjacent joints. $N()$ is the normal distribution with $\mu_k = 0$ and standard deviation δ_K .

$E_{O|\Lambda}$: occlusion area belong to joints; $W_i = \{w_i\}$: If i joint is occluded, $w_i = 1$, if i joint is not occluded, $w_i = 0$; I : input image; v_{ij} : Indicator for overlapping body parts; $\phi_i(I, X_i; \Lambda_i)$: potential of observation; $\phi_i^C(I, X_i; \Lambda_{ij})$: potential of the color; $\phi_i^E(I, X_i; \Lambda_i)$: potential of the edge; $\phi_i^{C_{\text{visible}}}(I, X_i; \Lambda_{ij})$: the motion state of X_i (the i th body joint) in the viewing area; $\phi_i^{C_{\text{occluded}}}(I, X_i; \Lambda_{ij})$: the motion state of X_i (the i th body joint) in the occluded area; ϕ_i : potential of observation; ψ_{ij}^K : potential of kinematic relationship; ψ_i^T : potential of temporal relationship. Defining a model, similar to [12] for calculating three potential function as follows.

Firstly, we get the observation potential function:

$$\phi_i(I, X_i; \Lambda_i) = \phi_i^C(I, X_i; \Lambda_i) + \phi_i^E(I, X_i; \Lambda_i). \quad (3)$$

The potential of the color

$$\phi_i^C(I, X_i; \Lambda_{ij}) = \phi_i^{C_{\text{visible}}}(I, X_i; \Lambda_{ij}) + \phi_i^{C_{\text{occluded}}}(I, X_i; \Lambda_{ij}), \quad (4)$$

where the first term is X_i of probability of occurrence of color in the visible area and the second term is for the occluded area. The visible term is formulated as

$$\begin{aligned} \phi_i^{C_{\text{visible}}}(I, X_i; \lambda_i) &= \prod_{u \in (\gamma(X_i) - (\gamma(X_i) \cap \gamma(X_j)))} P_C(I_u) \\ &= \prod_{u \in (\gamma(X_i) - (\gamma(X_i) \cap \gamma(X_j)))} \frac{P(I_u | \text{foreground})}{P(I_u | \text{background})}, \end{aligned} \quad (5)$$

where $P(I_u | \text{foreground})$ and $P(I_u | \text{background})$ are the distributions of the color of pixel u given the foreground and background.

$$\begin{aligned} \phi_i^{C_{\text{occluded}}}(I, X_i; \lambda_i) &= \prod_{u \in (\gamma(X_i) \cap \gamma(X_j))} [z_i(I_u) + (1 - z_i(I_u)) P_C(I_u)] \end{aligned} \quad (6)$$

and $z_i(I)$ is calculated as follows:

$$z_i(I) = \frac{1}{N} \sum (\phi_j^C(I_u, x_j(t); \lambda_j)) \quad (7)$$

$u \in (\gamma(X_i) \cap \gamma(X_j))$: the occlusion area is determined by the calculated overlapping region of X_i and X_j , N is the sum of all occlusion nodes.

When $f(\theta_i, \theta_j) = 1$, $T_{\text{lower}} \leq \theta_i - \theta_j \leq T_{\text{upper}}$, where T_{lower} and T_{upper} are the lower and upper bound of motion area between X_i and X_j defined by kinesiology.

Finally, potential of temporal relationship is calculated as follows:

$$\psi_i^T(X_i^t, X_i^{t-1}) = p\left(X_i^t - X_i^{t-1}; \mu_i, \sum_i\right), \quad (8)$$

where μ_i is the dynamics of X_i at the previous time step and Σ_i is a diagonal matrix with a diagonal element is identical to $|\mu_i|$, which similar to a Gaussian distribution with the time.

In this paper, the posterior distribution of model X conditioned on all input images up to the current joint s structure, the current time step τ and occlusion state variable $\Lambda^{1:\tau}$ is

$$p(X^\tau | I^{1:\tau}; \Lambda^{1:\tau}) = \frac{1}{Z} \exp \left\{ - \sum_{i \in X^{1:\tau}} \phi_i^C(I, X_i; \lambda_i) - \sum_{ij \in E_k^{1:\tau}} \psi_{ij}^K(X_i, X_j) - \sum_{i \in E_{T_i}^{1:\tau}, t \in 1:\tau} \psi_i^T(X_i^t, X_i^{t-1}) \right\}, \quad (9)$$

where Z is a normalization constant.

In a word, we put ϕ_i , ψ_{ij}^K , ψ_i^T into (4), and get body occluded joints positions,

$$\hat{X}^t = \underset{X^t}{\operatorname{argmax}} p(X^t | I^{1:t}; \hat{\Lambda}^{t-1}), \quad (10)$$

where X^t is X joint location at t time.

The occluded relation among joints can be obtained by formula (2).

$$\hat{\lambda}_{i,j}^t = \underset{\Lambda_{ij} \neq 0}{\operatorname{argmax}} \phi_i(I^t, \hat{X}_i^t; \Lambda_{ij}), \quad (11)$$

where \hat{X}_i^t is X_i position at t time.

The occluded joints can be calculated by MRF at the entire time of motion. In this paper, we connect missing data in order to restore missing coordinate position.

3. Feature Representation

The human action can be recognized in terms of hierarchical area model, relative velocity, and relative acceleration.

3.1. Hierarchical Area Model. For describing the human motion pose (e.g., jogging, running, and walking), we make use of hierarchical area model and extract human facial area S^H , upper limbs area S^U and leg area S^L . To human facial area S^H are extracted in the following way.

- (1) According to Canny algorithm, each of the facial contour point set is extracted, and denoted as C_k , where k is the number of contour point.
- (2) The face contour can be least square fitting by C_k , which obtained in step 1.
- (3) According to step 1 and step 2, if the body movement to make the front, the face area is the largest, if

the human turned sideways, the face area will change. Thus, face area in coordinate is

$$\begin{aligned} \Delta S_x^H(y, z) &= \frac{\sum_{i=1}^n S^H(x^i, y, z)}{\sum_{i=1}^n \delta(S^H(x^i, y, z))}, \\ \Delta S_y^H(x, z) &= \frac{\sum_{i=1}^n S^H(x, y^i, z)}{\sum_{i=1}^n \delta(S^H(x, y^i, z))}, \\ \Delta S_y^H(x, y) &= \frac{\sum_{i=1}^n S^H(x, y, z^i)}{\sum_{i=1}^n \delta(S^H(x, y, z^i))}, \end{aligned} \quad (12)$$

where n is the frames, $S^H(x^i, y^i, z^i)$ is the set of face contour in all frames, $\delta(v(x^i, y^i, z^i))$ is the set of contour in all frames.

- (4) By Repeat Steps 1~3, the face area can be calculated in all frames.

Calculating S^U and S^L is similar to S^H .

Figure 1 shows that the curve for some area features of pedestrian walking. Figure 1(a) is the area variation curve of S^H . Figure 1(b) is the area variation curve of S^U . Figure 1(c) is the area variation curve of S^L .

3.2. Relative Velocity and Relative Acceleration. We can get the relative velocity and relative acceleration by the trajectory of each joint.

Each point' weight can be considered as the same, and build statistical model to calculate the relative velocity and relative acceleration among relative motion joints (e.g., hands and legs) in order to reason the initial state of motion.

$$\Delta_{i,j} = \frac{p(x_i(t)_v, x_j(t)_v)}{\sum_{k=1}^n p(x_i(t)_v, x_j(t)_v)}, \quad (13)$$

where $\Delta_{i,j}$ is the relative velocity among i and j .

The area-velocity goodness T_j is obtained as follow.

- T1: jogging, Δv (the left knee, the right knee), Δv (the left foot, the right foot), Δv (the right knee, the right foot), Δv (the left foot, the left ankle), Δv (the right foot, the right ankle) $> t1$, and $\Delta \alpha$ (the left foot, the left knee) $> t2$.
- T2: running, Δv (the left foot, the left knee), Δv (the right foot, the right knee), Δv (the left foot, the left ankle), Δv (the right foot, the right ankle) $> t3$, and $\Delta \alpha$ (the left foot, the left knee), $\Delta \alpha$ (the left foot, the right knee), and $\Delta \alpha$ (the left foot, the right foot) $> t4$.
- T3: walking, Δv (the left foot, the left knee), Δv (the right foot, the right knee), Δv (the left foot, the left ankle), and Δv (the right foot, the right ankle) $> t5$.
- T4: jumping, Δv (the left foot, the left knee), Δv (the right foot, the right knee), Δv (the left foot, the left ankle), Δv (the right foot, the right ankle) $> t6$, and $\Delta \alpha$ (the left foot, the left ankle), and $\Delta \alpha$ (the right foot, the right ankle) $> t7$.

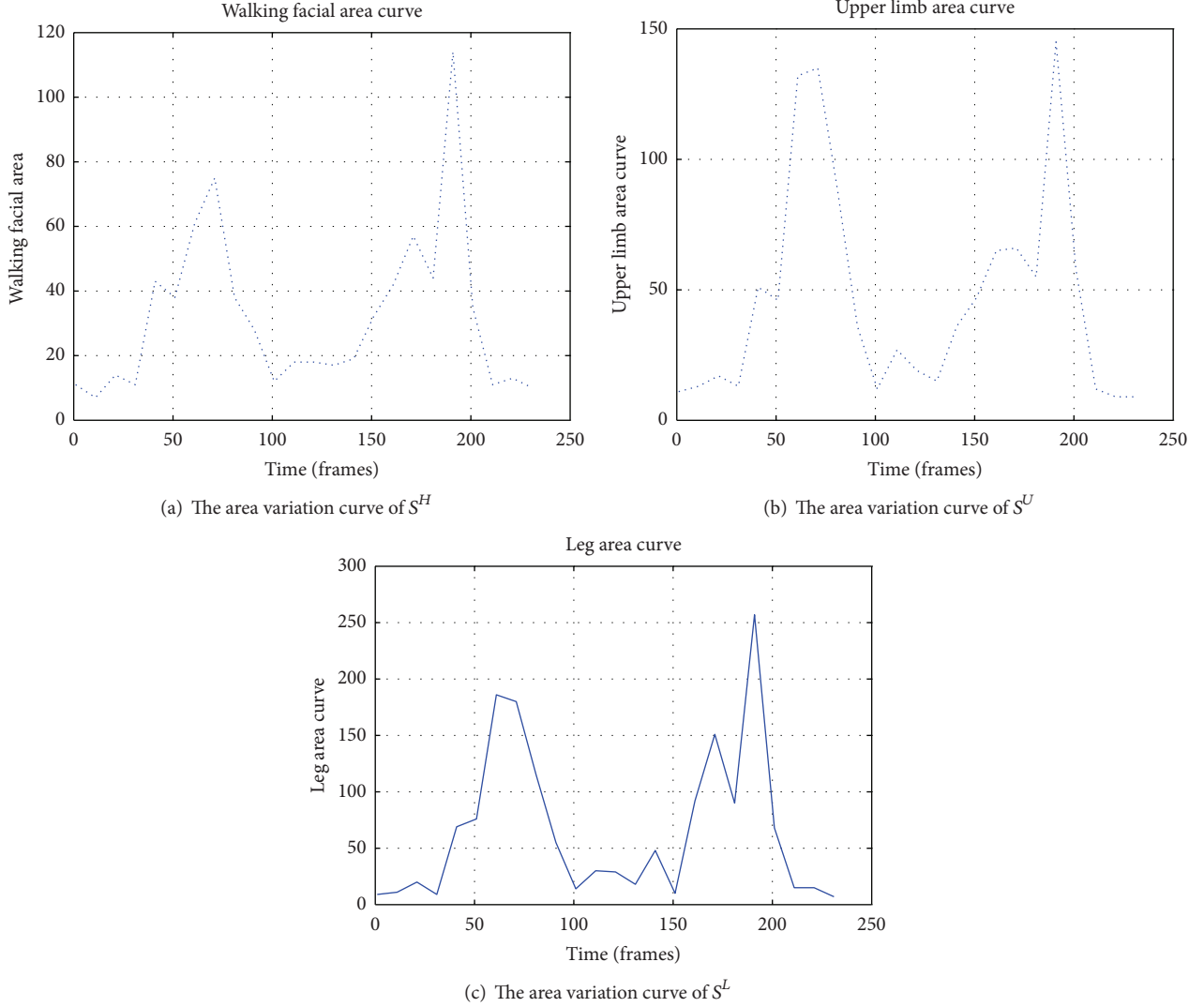


FIGURE 1: The curve for some area features of pedestrian walking.

T5: boxing, Δv (the left foot, the left knee), Δv (the right foot, the right knee), Δv (the left foot, the left ankle), Δv (the right foot, the right ankle) $> t_8$ and, $\Delta \alpha$ (the left hand, the left elbow), $\Delta \alpha$ (the right hand, the right elbow), $\Delta \alpha$ (the left foot, the left ankle), and $\Delta \alpha$ (the right foot, the right ankle) $> t_9$.

Thresholds (t_1, t_2, \dots, t_9) are determined empirically as 1.5, 40, 5.5, 60, 3.5, 5.0, 40, 7.0, and 30.

We cluster the extract feature, which meet the threshold requirement, and extract the typical behavior of the action dataset as a standard action: jogging, running, walking, jumping and boxing. Above 5 kinds of common action decomposition, we get relative velocity among joints, when some action occurred. For example, an jogging operation, the relative velocity of the left leg and the right leg and the relative velocity of the left leg and the left knee are more than others joints.

3.3. Codebook Formulation. In order to construct the codebook, we use the k -means algorithm based on the Euclidean

distance to cluster all the features (hierarchical area model, relative velocity and relative acceleration) extracted from the training frames. The center of each cluster is defined as a codeword. All the centers clustered from the training frames produce the codebook for the pLSA model. A frame in the training videos or in the test videos is assigned to a specific codeword in the codebook which has the minimal Euclidean distance to the frame. In the end, a video is encoded in a bag-of-words way, that is, a video is represented using a histogram of codewords, removing the temporal information.

4. pLSA-Based Human Action Recognition

pLSA is a statistical generative model that associates documents and words via the latent topic variables, which represents each documents as a mixture of topics. Our approach uses the bag of words representation as in papers [14–16]. What's difference is that we use the local spatial-temporal maximum value of hierarchical area model, relative velocity and relative acceleration as our features. We suppose that

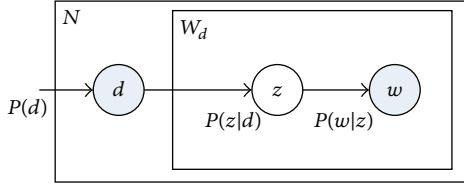


FIGURE 2: Graph model of pLSA.

the words are independent of the temporal order but related to the spatial order, for the k -means clustering approach with all of the features may lead to the mismatch of the words. Similar local features appearing at different position may be clustered together. When we calculate the frequency of the words, the mismatch appears. And this phenomenon may reduce the precision of the classify approach. In order to solve the problem, we assign spatial information to each word. In the classify approach, we use the pLSA models to learn and recognize human action.

In the context of action categorization, the topic variable z_k correspond to action categories, and each video d_i can be treated as a collection of space-time words w_j . The joint probability of video d_i , action category z_k and space-time word w_j can be expressed as

$$p(d_i, z_k, w_j) = p(w_j | z_k) p(z_k | d_i) p(d_i), \quad (14)$$

where $p(w_j | z_k)$ is the probability of word w_j occurring in action category z_k , $p(z_k | d_i)$ is the probability of topic z_k occurring in video d_i , and $p(d_i)$ can be considered as the prior probability of d_i . The conditional probability of $p(w_j | d_i)$ can be obtained by marginalizing over all the topic variables z_k :

$$p(w_j | d_i) = \sum_k p(z_k | d_i) p(w_j | z_k). \quad (15)$$

Denote $n(d_i, w_j)$ as the occurrence of word w_j in video d_i , the prior probability $p(d_i)$ can be modeled as

$$p(d_i) \propto \sum_j n(d_i | w_j). \quad (16)$$

A maximum likelihood estimation of $p(w_j | z_k)$ and $p(z_k | d_i)$ is obtained by maximizing the function using the Expectation Maximization (EM) algorithm, which the graph model is shown in Figure 2. The objective likelihood function of the EM algorithm is:

$$L = \prod_i \prod_j p(w_j | d_i)^{n(w_j, d_i)}. \quad (17)$$

The EM algorithm consists of two steps: an expectation (E) step computes the posterior probability of the latent variables, and a maximization (M) step maximizes the completed data likelihood computed based on the posterior probabilities obtained from E-step. Both steps of the EM algorithm for pLSA parameter estimate are listed below.

E-step: given $p(w_j | z_k)$ and $p(z_k | d_i)$ estimate $p(z_k | d_i, w_j)$

$$p(z_k | d_i, w_j) \propto p(w_j | z_k) p(z_k | d_i). \quad (18)$$

M-step: given the estimated $p(z_k | d_i, w_j)$ in E-step, and $n(d_i, w_j)$, estimate $p(w_j | z_k)$ and $p(z_k | d_i)$

$$\begin{aligned} p(w_j | z_k) &\propto \sum_i n(d_i, w_j) p(z_k | d_i, w_j), \\ p(z_k | d_i) &\propto \sum_i n(d_i, w_j) p(z_k | d_i, w_j). \end{aligned} \quad (19)$$

For the task of human motion classification, our goal is to classify a new video to a specific activity class. During the inference stage, given a testing video test, the document specific coefficients $p(z_k | d_{\text{test}})$.

We can treat each aspect in the pLSA model as one class of activity. So, the activity categorization is determined by the aspect corresponding to the highest $p(z_k | d_{\text{test}})$. The action category k of d_{test} is determined as

$$k = \arg \max_k p(z_k | d_{\text{test}}). \quad (20)$$

In this paper, we treat each frame in a video as a single word and a video as a document. The probability distribution $p(z_k | d_{\text{test}})$ can be regarded as the probability of each class label for a new video. The parameter in the training step defines the probability of a word w_j drawing from an aspect z_k . The aforementioned standard EM training procedure for pLSA is to replace

$$p(z_k | d_i, w_j), \quad p(w_j | z_k), \quad (21)$$

with their optimal possible values at each iteration.

For action recognition with large amount of training data, this would result in long training time. This paper presents an incremental version of EM to speed up the training of PLSA without sacrificing performance accuracy. Assuming the observed data are independent of each other, we propose an incremental EM algorithm presented in Algorithm 1.

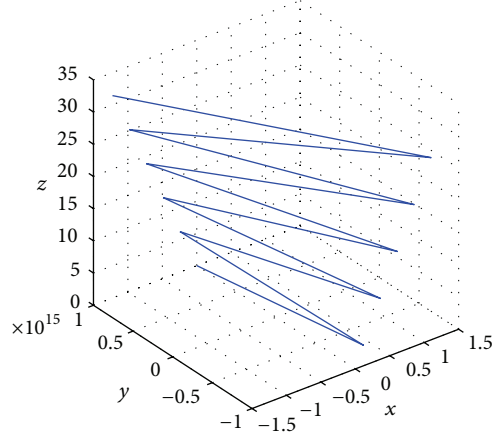
Algorithm 1. Incremental EM Algorithm for PLSA Parameter Estimation is as follows.

- (1) Inputs;
 - (2) K —the number of action categories;
 - (3) D —the number of training videos;
 - (4) S —the number of videos in each subset;
 - (5) M —the size of the codebook of spatial-temporal words;
 - (6) Outputs;
 - (7) $\hat{U} = \{\hat{p}(z_k | d_i)\}_{k,i}$;
 - (8) $\hat{V} = \{\hat{p}(w_j | z_k)\}_{j,k}$;
 - (9) E-Step;
- for all k and j , calculate

$$p(w_j | z_k) = \frac{n_{j,k}}{n_k}. \quad (22)$$



(a)



(b)

| | Torso(axis) | Head | Right shoulder | Left shoulder | Right elbow | Left elbow | Right hand | Left hand | Right hip | Left hip | Right knee | Left knee | Right ankle | Left ankle | Right foot | Left foot |
|----------------|-------------|------|----------------|---------------|-------------|------------|------------|-----------|-----------|----------|------------|-----------|-------------|------------|------------|-----------|
| Torso(axis) | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Head | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Right shoulder | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Left shoulder | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Right elbow | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Left elbow | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Right hand | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Left hand | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Right hip | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ±1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Left hip | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Right knee | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Left knee | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 |
| Right ankle | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Left ankle | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Right foot | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Left foot | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

(c)

FIGURE 3: Process of restoring missing coordinate position Remarks: Figure 3 the original picture, Figure 3 Reconstruction of knee motion used the method of least squares data fitting in order to restore missing coordinate position, Figure 3 the occlusion diagram. In the diagram, occlusion part pairs, occlusion state value -1 (red cell for occluded one) and 1 (green cell for occluder), ±1 (orange red cell for rigid body), respectively. In this manner, every part pairs get corresponding occlusion state values.

For all (d_{test}, w_j) pairs and $k \in \{1, \dots, K\}$ calculate

$$p(z_k | d_{\text{test}}, w_j) = \frac{p(w_j | z_k) p(z_k | d_{\text{test}})}{\sum_{i=1}^K p(w_j | z_i) p(z_i | d_{\text{test}})}; \quad (23)$$

M-Step: calculate the following:

$$p(z_k | d_{\text{test}}) = \frac{\sum_{j=1}^N n(d_{\text{test}}, w_j) p(z_k | d_{\text{test}}, w_j)}{n(d_{\text{test}})}; \quad (24)$$

(10) Repeat E-steps and M-step until the convergence condition is met;

(11) Calculate activity class

$$k = \arg \max_k p(z_k | d_{\text{test}}). \quad (25)$$

5. Experimental Result

5.1. Datasets. We test our algorithm on two datasets: the Weizmann human motion dataset [17], the KTH human action dataset [18, 19], and the HumanEva dataset [3, 20]. All the experiments are conducted on a Pentium 4 machine with 2 GB of RAM, using the implementation on MATLAB. The dataset and the related experimental results are presented in the following sections.

KTH datasets is provided by Schuldt which contains 2391 video sequences with 25 actors showing six actions. Each action is performed in 4 different scenarios.

The WEIZMANN datasets is provided by Blank which contains 93 video sequences showing nine different people, each performing ten actions, such as run, walk, skip, jumping-jack, jump-forward-on-two-legs, jump-in-place-on-two-legs, gallop sideways, wave-two-hands, wave-one-hand and bend.

The HumanEva dataset [3, 20] is used for evaluation. It contains six different motions: Walking, Jogging, Gestures, Boxing, and Combo.

In order to evaluate and fairly compare the performance, we use the same experimental setting as in [21, 22]. For every dataset, 12 video sequences taken by four subjects (out of the five) are used for training, and the remaining three videos for testing. The experiments are repeated five times.

The performance of different methods is shown using the average recognition rate. We report the overall accuracy on three datasets. In order to evaluate the performance of occlusion state estimation and reconstruct missing coordinate position, we hand-labeled the ground truth of the occlusion states for test motions. Figure 3 shows how the ground truth of occlusion state is specified.

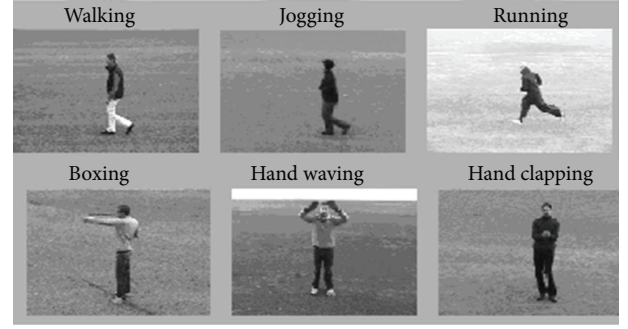
5.2. Comparison. KTH Dataset. It contains six types of human actions (walking, jogging, running, boxing, hand waving, and hand clapping) performed several times by 25 subjects in four different scenarios: outdoors, outdoor-with scale variation, outdoors with different clothes, and indoors. Representative frames of this dataset are shown in Figure 4(a). After the process of restoring missing coordinate position, we use the proposed method, the classification results of KTH dataset obtained by this approach are shown in Figure 5 and indicate quite a small number of videos are misclassified, particularly, the actions, “running” and “handclapping,” are more tended to be confused.

The Weizmann Dataset. The Weizmann human action dataset contains 83 video sequences showing nine different people, and each performing nine different actions: bending (a1), jumping jack (a2), jumping forward on two legs (a3), jumping in place on two legs (a4), running (a5), galloping sideways (a6), walking (a7), waving one hand (a8), waving two hands (a9).

The figures were tracked and stabilized by using the background subtraction masks that come with this data set. Some sample frames are shown in Figure 4(b). The classified results achieved by this approach are shown in Figure 6.

The HumanEva Dataset. The HumanEva dataset is used for evaluation, which are shown in Figure 4(c). It contains five different motions: Walking (a1), Jogging (a2), Gestures (a3), Boxing (a4), and Combo (a5). Each motion is performed by four subjects and recorded by seven cameras (three RGB and four gray scale cameras) with the ground truth data of human joints. The classified results achieved by this approach are shown in Figure 7.

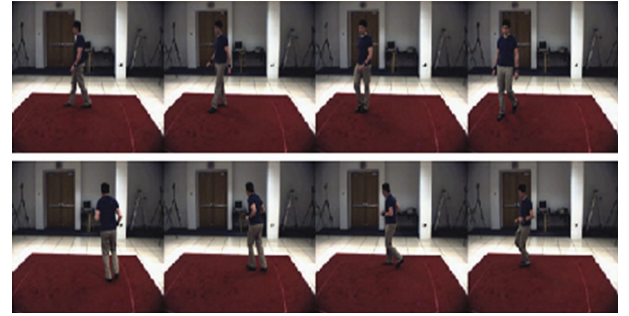
In this paper, we identify jogging, running, walking and boxing and compare the proposed method with the four state-of-the-art methods in the literature: Blank et al. [18], Lu et al. [19], Sigal et al. [3], Chang et al. [20] and Juan Carlos Nieves [21] in three dataset. As shown in the Tables 1, 2 and 3, the existing methods, the low recognition accuracy because these action are not only occlusion situation are complex, but also the legs have complex beat, motion and other group actions. The proposed method can overcome these problems,



(a)



(b)



(c)

FIGURE 4: Sample frames from our datasets. The action labels in each dataset are as follows (a) KTH data set: walking (a1), jogging (a2), running (a3), boxing (a4), and handclapping (a5); (b) Weizmann data set: running, walking, jumping-jack, waving-two-hands, waving-one-hand, and bending; (c) HumanEva dataset: walking(a1), jogging (a2), gestures (a3), boxing (a4), and combo (a5). Each motion is performed by four subjects and recorded by seven cameras (three RGB and four gray scale cameras) with the ground truth data of human joints.

| | | | | | |
|----|------|------|------|------|------|
| a1 | 0.91 | 0.00 | 0.03 | 0.00 | 0.00 |
| a2 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| a3 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 |
| a4 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| a5 | 0.03 | 0.03 | 0.00 | 0.01 | 0.75 |
| | a1 | a2 | a3 | a4 | a5 |

FIGURE 5: Confusion matrix for KTH data set.

| | | | | | | | | | |
|----|------|------|------|------|------|------|------|------|------|
| a1 | 1.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| a2 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| a3 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| a4 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| a5 | 0.03 | 0.03 | 0.00 | 0.01 | 0.75 | 0.00 | 0.31 | 0.00 | 0.00 |
| a6 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.92 | 0.04 | 0.00 | 0.00 |
| a7 | 0.00 | 0.00 | 0.00 | 0.00 | 0.41 | 0.00 | 0.95 | 0.00 | 0.00 |
| a8 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| a9 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| | a1 | a2 | a3 | a4 | a5 | a6 | a7 | a8 | a9 |

FIGURE 6: Confusion matrix for Weizmann data set.

| | | | | | |
|----|------|------|------|------|------|
| a1 | 0.92 | 0.00 | 0.03 | 0.00 | 0.00 |
| a2 | 0.00 | 0.97 | 0.00 | 0.00 | 0.00 |
| a3 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 |
| a4 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| a5 | 0.03 | 0.03 | 0.00 | 0.01 | 0.86 |
| | a1 | a2 | a3 | a4 | a5 |

FIGURE 7: Confusion matrix for HumanEva data set.

TABLE 1: Compared with other approaches on KTH dataset.

| Method | Average recognition rate (%) |
|--|------------------------------|
| The proposed method | 92.50 |
| Lu et al. [19] and Blank et al. [18] | 81.50 |
| Chang et al. [20] and Sigal et al. [3] | 91.20 |
| Niebles et al. [21] | 87.04 |

TABLE 2: Compared with other approaches on Weizmann dataset.

| Method | Average recognition rate (%) |
|--|------------------------------|
| The proposed method | 90.10 |
| Lu et al. [19] and Blank et al. [18] | 89.30 |
| Chang et al. [20] and Sigal et al. [3] | 86.20 |
| Niebles et al. [21] | 88.6 |

and the recognition accuracy and average accuracy are higher than the comparative method.

The experimental results show that the approach proposed in the paper can get satisfactory results and significantly performs better compared the average accuracy with that in [3, 18–21], because of a practical method adopted in the paper.

6. Conclusions and Future Work

In this paper, we proposed an adaptive occlusion state estimation method for 3D human body movement.

Our method successfully recognize without assuming a known and fixed depth order. The proposed method can infer

TABLE 3: Compared with other approaches on HumanEva dataset.

| Method | Average recognition rate (%) |
|--|------------------------------|
| The proposed method | 91.40 |
| Lu et al. [19] and Blank et al. [18] | 88.70 |
| Chang et al. [20] and Sigal et al. [3] | 90.20 |
| Niebles et al. [21] | 90.6 |

state variables efficiently because it separates the estimation procedure into body configuration estimation and occlusion state estimation. More specifically, in the occlusion state estimation step, at first, we reconstruct human trajectory reconstruction which representing the 3D human pose occlusion relationship and detect body parts having an occlusion relationship using the overlapping body parts by using a Markov random field (MRF) with a state variable. Finally, we use the topic model of pLSA to classify. Experimental results showed that the proposed method successfully estimates the occlusion states in the presence of self-occlusion and the average accuracy is about 92.5%, 90.1%, and 91.4% on the KTH dataset, Weizmann dataset, and HumanEva dataset respectively, which is better than other approaches [3, 18–21].

We conjecture that the proposed method can be extended for tracking poses from (two or more) interacting people. Tracking poses of interacting people, however, will involve more complex problems such as dealing with more variable motion, inter-person occlusions, and possible appearance similarity of different people.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper (such as financial gain).

Acknowledgments

This research work was supported by the Grants from the Natural Science Foundation of China (no. 50808025) and the Doctoral Fund of China Ministry of Education (Grant no. 20090162110057).

References

- [1] X. I. A. Li-min, Q. Wang, and W. U. Lian-shi, "Vision based behavior prediction of ball carrier in basketball matches," *Journal of Central South University of Technology*, vol. 19, no. 8, pp. 2142–2151, 2012.
- [2] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.
- [3] L. Sigal, A. O. Balan, and M. J. Black, "HumanEva: synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *International Journal of Computer Vision*, vol. 87, no. 1-2, pp. 4–27, 2010.
- [4] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Strike a pose: tracking people by finding stylized poses," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 271–278, June 2005.
- [5] H. Jiang and D. R. Martin, "Global pose estimation using non-tree models," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.
- [6] M. W. Lee and R. Nevatia, "Human pose tracking in monocular sequence using multilevel structured models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 27–38, 2009.
- [7] P. Guo, Z. Miao, Y. Shen, and H.-D. Cheng, "Real time human action recognition in a long video sequence," in *Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based (AVSS '10)*, pp. 248–255, Boston, Mass, USA, September 2010.
- [8] J. C. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 299–318, 2008.
- [9] Y. Wang and G. Mori, "Human action recognition by semilattent topic models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1762–1774, 2009.
- [10] B. W. Sy, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell, "Hidden conditional random fields for gesture recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 1521–1527, June 2006.
- [11] N.-G. Cho, A. L. Yuille, and S.-W. Lee, "Adaptive occlusion state estimation for human pose tracking under self-occlusions," *Pattern Recognition*, vol. 46, no. 3, pp. 649–661, 2013.
- [12] A. Asthana, M. Delahanty, A. Dhall, and R. Goecke, "Facial performance transfer via deformable models and parametric correspondence," *Proceeding of Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1511–1519, 2012.
- [13] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2nd edition, 2004.
- [14] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proceedings of the 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, VS-PETS*, pp. 65–72, October 2005.
- [15] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, and G. Serra, "Recognizing human actions by fusing spatio-temporal appearance and motion descriptors," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 3569–3572, Cairo, Egypt, November 2009.
- [16] J. Wu and J. M. Rehg, "CENTRIST: a visual descriptor for scene categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1489–1501, 2011.
- [17] C. Schödl, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, pp. 32–36, August 2004.
- [18] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1395–1402, Beijing, China, October 2005.
- [19] W.-L. Lu, K. Okuma, and J. J. Little, "Tracking and recognizing actions of multiple hockey players using the boosted particle filter," *Image and Vision Computing*, vol. 27, no. 1-2, pp. 189–205, 2009.
- [20] J.-Y. Chang, J.-J. Shyu, and C.-W. Cho, "Fuzzy rule inference based human activity recognition," in *Proceedings of the IEEE International Conference on Control Applications (CCA '09)*, pp. 211–215, St Petersburg, Russia, July 2009.
- [21] J. Carlos Niebles, C. -W. Chen, and L. Fei-Fei, "Modeling temporal structure of decomposable motion segments for activity classification," in *Proceedings of the 11th European Conference Computer Vision (ECCV '10)*, vol. 6312 of LNCS, pp. 392–405, 2010.
- [22] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, and G. Serra, "Recognizing human actions by fusing spatio-temporal appearance and motion descriptors," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 3569–3572, Cairo, Egypt, November 2009.

Research Article

Image Matching Using Dimensionally Reduced Embedded Earth Mover's Distance

Fereshteh Nayyeri and Mohammad Faizul Nasrudin

Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Correspondence should be addressed to Fereshteh Nayyeri; f.nayyeri@gmail.com

Received 3 July 2013; Accepted 31 October 2013

Academic Editor: Feng Gao

Copyright © 2013 F. Nayyeri and M. F. Nasrudin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Finding similar images to a given query image can be computed by different distance measures. One of the general distance measures is the Earth Mover's Distance (EMD). Although EMD has proven its ability to retrieve similar images in an average precision of around 95%, high execution time is its major drawback. Embedding EMD into L_1 is a solution that solves this problem by sacrificing performance; however, it generates a heavily tailed image feature vector. We aimed to reduce the execution time of embedded EMD and increase its performance using three dimension reduction methods: sampling, sketching, and Dimension Reduction in Embedding by Adjustment in Tail (DREAT). Sampling is a method that randomly picks a small fraction of the image features. On the other hand, sketching is a distance estimation method that is based on specific summary statistics. The last method, DREAT, randomly selects an equally distributed fraction of the image features. We tested the methods on handwritten Persian digit images. Our first proposed method, sampling, reduces execution time by sacrificing the recognition performance. The sketching method outperforms sampling in the recognition, but it records higher execution time. The DREAT outperforms sampling and sketching in both the execution time and performance.

1. Introduction

One of the interesting problems in database communities is image retrieval from large databases. The fundamental issue is how to design a similarity measure in a manner that shows the concept of similarity between two images, because choosing a proper measure has considerable effects on image retrieval applications. Some of the similarity measures include the Earth Mover's Distance (EMD), Jeffrey's divergence, and Minkowski-form distance [1].

The EMD is a general and flexible metric that has desirable and striking properties for content-based image retrieval [2, 3]. This similarity measure, which applies to weighted point sets, measures the minimum amount of work needed to transform one set into another set by weight transportation. The most significant feature of EMD is that it quantifies perceptual similarity better than other types of distances used for image retrieval [2]. Although EMD can measure the exact distance between images, and by this measure we can retrieve the most similar images from a database, its

execution time is problematic, and this similarity measure is very time consuming.

Another method, called embedded EMD to L_1 , was proposed to solve the EMD problem. This method maps the image matrix to an L_1 norm; therefore, instead of comparing 2-dimensional matrixes, we can compare 1-dimensional vectors. Although this idea is less time consuming, it produces distortion. Sometimes, an exact computation may be practically infeasible; in this situation, an approximation solution is helpful to find the exact result with some distortion. Both execution time and performance are important factors in image retrieval, and we should attempt to reduce distortion as much as possible. In this paper, we propose two methods to improve the performance of embedded EMD. The first method, sampling, reduces the time but decreases performance. In the next proposed method, sketching, we improve performance by sacrificing the time of execution. Finally, in the last method, by solving the problem of sampling, we improve the performance while reducing the execution time.

TABLE 1: Relationship between image's size and array's length.

| | G_1 n^2 | G_2 $n^2/4$ | G_3 $n^2/16$ | G_4 $n^2/64$ | G_5 $n^2/256$ |
|-----------------------------|----------------|------------------|-------------------|-------------------|--------------------|
| Number of elements in array | 256 | 64 | 16 | 4 | 1 |
| Side length | 1 | 2 | 4 | 8 | 16 |

The remainder of this paper is organized as follows. In Section 2 we discuss related previous work. In Section 3 we describe our proposed technique. Section 4 provides the details of our proposed methods. Finally, we discuss the results and our conclusion in Sections 5 and 6.

2. Previous Work

The concept of Earth Mover Distance (EMD) was first explored in [4] to measure perceptual shape similarity. The use of EMD for computing similarity between images was later proposed in [5]. Since then, the EMD has become a trendy similarity measure in computer vision; it has been used effectively in various applications including color-based image retrieval systems, texture signatures [6], shape matching [7–9], and music score matching [10]. The EMD performs quite well in comparison with other similarity measures, such as the Jeffrey divergence and the Minkowski-form distance. In addition, the EMD can be used to measure the differences between vector fields [11].

Some authors in [2] have compared the EMD with other similarity measures and evaluated the retrieval performance of each. The results of the comparisons demonstrate that the EMD is more robust than other measures for the purpose of image retrieval because it matches similarity better than other distances.

The main idea behind the EMD metric is as follows. Suppose that each image is a set of colored points in 2-dimensional space. The minimum amount of work needed to transform one set into another set is defined as the distance of two set points. In recent years, a low-distortion embedding of EMD into L_1 has been developed [12]; although the empirical results show that this distortion is much smaller than what had been estimated previously, the embedding steps themselves decrease the complexity of computing similarity between two images. Other authors [9] have reported on the complexities of querying the time and space of an exact EMD versus an embedded EMD for shape similarity. In this work, we demonstrate how to reduce the complexity of the computing correspondence between two images that are mapped to an L_1 norm by dimension reduction.

The most similar work in this area is that of Grauman and Darrell [9], who show a contour matching algorithm that quickly quantifies the minimum weight matching between sets of descriptive local features using the embedding of the Earth Mover's Distance (EMD) into a normed space. Their method achieves an increase in speed of four orders of magnitude over the exact method at the cost of only a 4% reduction in accuracy.

3. Dimension Reductions in L_1

In modern image retrieval applications, the data is sometimes not only very large relative to the physical memory or even to the disk, but also highly sparse. Accordingly, computing the embedded L_1 on large-scale sparse data can be challenging and time consuming. Various projection methods have been suggested for speeding up these computations. Dimension reduction in the L_1 norm has many applications in information retrieval. The authors of [13] show that dimension reduction by sampling from L_1 does not produce poor results. Additionally, by estimating distances in L_1 from random samples, the original L_1 distances can be recovered. Sampling methods become more important with increasingly large collections [14] because we can use the same set of random samples to estimate any L_1 pairwise distances [15], whereas measuring exact pairwise distances is often too time consuming or sometimes infeasible; however, random sampling often performs poorly when most of the samples are zeros [13]. Additionally, in strictly heavy-tailed data, the estimation errors are sometimes very large.

As another choice of random projection, various sketching algorithms have become popular. In general, a sketching algorithm outperforms random sampling, although random sampling is much more flexible [15]. In the sketching method, after scanning the data, we compute specific summary statistics, and then repeat this step k times.

3.1. Procedures of Sampling and Sketching. Suppose we have a database of n images and we want to compare a particular image with this database. To do so, we need a measurement; this is when we use EMD. Consider that we have 2 images with high similarity, for example, in Figures 1 and 2 apples with spots in different positions.

In this situation, the EMD of two spots in these images is computed as follows.

Euclidean distance between

$$(i) \text{ 1st pixels: } \sqrt{(8-9)^2 + (12-8)^2} = \sqrt{15},$$

$$(ii) \text{ 2nd pixels: } \sqrt{(8-9)^2 + (13-9)^2} = \sqrt{15},$$

$$(iii) \text{ 3rd pixels: } \sqrt{(9-10)^2 + (12-8)^2} = \sqrt{15},$$

$$(iv) \text{ 4th pixels: } \sqrt{(9-10)^2 + (13-9)^2} = \sqrt{15}.$$









































































Therefore, the EMD of two spots is $\sqrt{15} + \sqrt{15} + \sqrt{15} + \sqrt{15} = 4\sqrt{15} = 15.5$.

In the EMD metric, Euclidean distances between all weighted point sets are computed and then the minimum distance between each pair of point sets can be found. There are different methods to solve this type of weighted matching problems; in our case we use the "Hungarian" method [16–19]. This method finds the minimum distances between each pair of points in two images with n points in $O(n^3)$ arithmetic operations; therefore, the typical EMD is very time consuming, which is the biggest drawback for EMD. Another drawback is that when two weighted point sets have unequal total weights, EMD is not an appropriate metric; however,

TABLE 2: Number of vector elements in four methods.

| | | | | | | | | | | | | | |
|---------------------------------|----------------------------------|---|--------------|---|--------------|---|--------------|---|--------------|---|--------------|---|--------------|
| In L_1 -vector (EEMD methods) | | | | | | | | | | | | | |
| | G_1 | | G_2 | | G_3 | | G_4 | | G_5 | | G_6 | | G_7 |
| | \downarrow | | \downarrow | | \downarrow | | \downarrow | | \downarrow | | \downarrow | | \downarrow |
| Elements: | 4096 | + | 1024 | + | 256 | + | 64 | + | 16 | + | 4 | + | 1 |
| = 5461 | | | | | | | | | | | | | |
| In sample and sketch vector | | | | | | | | | | | | | |
| Elements: | 10% of L_1 vector = 546 | | | | | | | | | | | | |
| In DREAT vector | | | | | | | | | | | | | |
| | 10% of G_1 | | 10% of G_2 | | 10% of G_3 | | 10% of G_4 | | 10% of G_5 | | 10% of G_6 | | 10% of G_7 |
| | \downarrow | | \downarrow | | \downarrow | | \downarrow | | \downarrow | | \downarrow | | \downarrow |
| Elements: | 409 | + | 102 | + | 25 | + | 6 | + | 1 | + | 4 | + | 1 |
| = 548 | | | | | | | | | | | | | |

TABLE 3: Some samples of handwritten Persian digit images.

| Class | No. of images | Digit | Correct shape | Samples | | | | |
|----------|---------------|----------|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | 4 | 5 |
| Class 1 | 432 | Zero |  |  |  |  |  |  |
| Class 2 | 1500 | One |  |  |  |  |  |  |
| Class 3 | 1067 | Two |  |  |  |  |  |  |
| Class 4 | 256 | Three |  |  |  |  |  |  |
| Class 5 | 173 | Four (1) |  |  |  |  |  |  |
| Class 6 | 22 | Four (2) |  |  |  |  |  |  |
| Class 7 | 180 | Five |  |  |  |  |  |  |
| Class 8 | 138 | Six (1) |  |  |  |  |  |  |
| Class 9 | 31 | Six (2) |  |  |  |  |  |  |
| Class 10 | 713 | Seven |  |  |  |  |  |  |
| Class 11 | 150 | Eight |  |  |  |  |  |  |
| Class 12 | 657 | Nine |  |  |  |  |  |  |

























it is desirable for robust matching to allow point sets with different total weights and cardinalities [18]. On the other hand, approximation is a good idea because usually exact computation is practically infeasible and an approximate solution can help to find the exact solution more efficiently.

In implementing EMD, in order to embed two sets of contour features with different total weights, we simulate equal weights by eliminating the appropriate number of random points from the larger weight set. For example, in Figure 2, when points are sampled uniformly from the contours of two images of Persian number 3 with a size of 64×64 pixels, the first image has 124 points, while the second image has 131 points. Therefore, the first image has 13 more points than the second one, and 13 points are randomly chosen from its contour to be eliminated.

The next part of the application is implementing embedded EMD into L_1 . We formally show how to construct an

embedded EMD into L_1 . A boundary of $\sqrt{\log n}$ on any L_1 embedding distortion has been defined [20], where n is the number of pixels in the width or height of image (width and height of image are equal). We embed the minimum weight matching of contour features into L_1 via the EMD embedding of [12, 21]. To embed EMD into L_1 , we put bitmap image in a grid whose size is twice bigger than that of the original image and shift grid randomly upon the image. Afterwards, we map pixels of the new image (which are all 0 or 1) to elements of an array in a special orientation starting from the first pixel in the left-top bit of the image to its last pixel in the right-bottom bit. The rest of the array should be set after some computation. For example, in the embedding of a 16×16 image, G_1 is the first grid and it includes 256 elements, each of which has a side length equal to 1. The first 256 elements of the array are set with these elements. In the next step, we add each of the 4 neighbouring elements in G_1 and place the

TABLE 4: Some samples of handwritten Persian letter images.

| Letter | Correct shape | samples | | | | |
|--------|---------------|---|---|--|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Cha | چ |  |  |  |  |  |
| Ja | ج |  |  |  |  |  |
| Zha | ژ |  |  |  |  |  |
| Za | ز |  |  |  |  |  |
| Ta | ت |  |  |  |  |  |
| Sa | ث |  |  |  |  |  |
| Fa | ف |  |  |  |  |  |
| Gha | ق |  |  |  |  |  |





























































```

function Calculate_  $L_1$ (image)
begin
    Scale image to  $64 \times 64$ 
    Initialize an image matrix, original_img, of  $64 \times 64$ 
    Initialize an image matrix, imageMtrx, of  $(2 \times 64) \times (2 \times 64)$ 
    Set imageMtrx to original_img
    Initialize  $U_1$ , a random number between 0 and 64
    Initialize  $V_1$ , a random number between 0 and 64
    Shift each position in imageMtrx to position  $(U_1, V_1)$ 
    Set  $G_1$  array to pixels of imageMtrx
    //Create  $L_1$ -vector of image matrix including  $G_1$  array followed by  $G_2$  array, ...,
    followed by  $G_n$  array (as in Figure 3)
    Initialize  $L_1$ -vector to null
     $L_1$ -vector =  $G_1$ 
    for  $i = 1$  to 6
        Set  $G_{i+1}$  to sum of each 4-neighbour elements of  $G_i$  multiplied by side length
         $L_1$ -vector =  $L_1$ -vector +  $G_{i+1}$ 
    end for
    return  $L_1$ -vector
end

function Calculate_Embedded EMD(image 1, image 2)
begin
    Initialize  $L_1$ -vector 1
    Initialize  $L_1$ -vector 2
    Initialize EEMD
    Set  $L_1$ -vector 1 to Calculate_  $L_1$ (image 1)
    Set  $L_1$ -vector 2 to Calculate_  $L_1$ (image 2)
    Subtract each pair of corresponding elements of  $L_1$ -vector 1 and  $L_1$ -vector 2
    Add all subtractions into EEMD and display it
end

```

TABLE 5: Preprocessing of some test images.

| Original image | 1st process Crop white margin | 2nd process Resize image | 3rd process Get contour | 4th process Remove spots randomly |
|---|---|---|--|---|
| Width: 77 Height: 95 | | Width: 64 Height: 64 | Black spots | Black spots: 150 |
|  |  Width: 20, height: 23 |  |  | 284  |
|  |  Width:13, height: 35 |  |  | 246  |
|  |  Width: 18, height: 39 |  |  | 287  |
|  |  Width: 34, height: 41 |  |  | 265  |
|  |  Width: 22, height: 43 |  |  | 359  |
|  |  Width: 33, height: 50 |  |  | 438  |
|  |  Width: 29, height: 43 |  |  | 334  |
|  |  Width: 36, height: 42 |  |  | 292  |
|  |  Width: 29, height: 43 |  |  | 322  |
|  |  Width: 28, height: 39 |  |  | 342  |
|  |  Width: 25, height: 36 |  |  | 350  |
|  |  Width: 34, height: 43 |  |  | 272  |

results in the corresponding elements of G_2 , which will be the next 64 elements in the array. In the 3rd step, we add each of the 4 neighbouring elements in G_2 and place the results in the corresponding elements of G_3 , which will be the next 16 elements in the array. We continue this process until we have just one element, G_5 , which will be the last element of the array. In Figure 3, you can observe the embedding of an image's pixels to an array in L_1 .

The length of array is the sum of all the grids' lengths. So, in our example length of the array is 341 which is approximately equal to 2×16^2 . As a result, length of the embedding vector is $2\Delta^2$. Table 1 shows the relationship between the size of an image, side length, and the number of each grid's elements in an array.

Pseudocode 1 describes the embedded EMD technique.

Finding the EMD of two images with this method has a complexity of $O(n^2)$, because, for the mapping to L_1 , the vectors are of length $O(n^2)$. Therefore, finding the L_1 of two vectors, that is, vector A and B as in Figure 4, can be done in $O(n^2)$, which is better than $O(n^3)$ in exact EMD. The L_1 mapping is defined as

$$L_1(A, B) = |A_0 - B_0| + |A_1 - B_1| + \dots + |A_i - B_i| + \dots + |A_{2x}^2 - B_{2x}^2|. \quad (1)$$

Note that the exact EMD has a complexity of $O(n^3)$, which is the complexity of the Hungarian algorithm used for its implementation, and that the embedding EMD to L_1 , which computes an approximation instead of the exact EMD, reduces the complexity to $O(n^2)$. We propose two techniques

```

function Calculate_Sampling(image 1, image 2)
begin
  Initialize  $L_1$ _vector 1
  Initialize  $L_1$ _vector 2
  Initialize sampling_  $L_1$ _Vector 1
  Initialize sampling_  $L_1$ _Vector 2
  Initialize sampling_EMD
  Set  $L_1$ _vector 1 to Calculate_  $L_1$ (image 1)
  Set  $L_1$ _vector 2 to Calculate_  $L_1$ (image 2)
  Select 10% indexes of  $L_1$ _vector 1 randomly
  Put the elements of selected indexes of  $L_1$ _vector 1 into sampling_  $L_1$ _Vector 1
  Put the elements of selected indexes of  $L_1$ _vector 2 into sampling_  $L_1$ _Vector 2
  Subtract each pair of corresponding elements in sampling_  $L_1$ _Vector 1 and sampling_  $L_1$ _Vector 2
  Add all subtractions into sampling_EMD and display it
end

```

PSEUDOCODE 2

```

function Calculate_Sketching (image 1, image 1)
begin
  Initialize  $L_1$ _vector 1
  Initialize  $L_1$ _vector 2
  Initialize Sketching Vector Length as 10% of  $L_1$ _vector 1 length
  Initilize Sketching_Mtrx as  $L_1$ _vector 1 length  $\times$  Sketching Vector Length randomly
  // Sketching_Mtrx is as in Figure 6
  Initialize sketching_Vector 1
  Initialize sketching_Vector 2
  Initialize sketching_EMD
  Set  $L_1$ _vector 1 to Calculate_  $L_1$ (image 1)
  Set  $L_1$ _vector 2 to Calculate_  $L_1$ (image 2)
  for  $i = 1$  to Sketching Vector Length
    Multiply each pair of corresponding elements in row  $i$  of Sketching_Mtrx and  $L_1$ _vector 1
    Put the sum of multiplications in sketching_Vector 1
    Multiply each pair of corresponding elements in row  $i$  of Sketching_Mtrx and  $L_1$ _vector 2
    Put the sum of multiplications in sketching_Vector 2
  end for
  Subtract each pair of corresponding elements in sketching_Vector 1 and sketching_Vector 2
  Add all subtractions into sketching_EMD and display it
end

```

PSEUDOCODE 3

to reduce the complexity of EMD to $O(n)$ by using dimension reduction in the L_1 , sampling, and sketching. Concept of the dimension reduction technique from n to predetermined N -dimensional space is based on linear transformation, for example, elements of transformation 2-dimensional matrix A to a 1-dimensional vector [22].

Sampling is an option for dimension reduction in any norm (e.g., L_1 or L_2). In fact, using this technique, distances in L_1 or L_2 from random samples can be estimated by a simple scaling [13, 22]. Although it is a simple and popular method to approximate distances, it does not guarantee accuracy. In this method, as it is shown in Figure 5, we randomly pick k (out of D) columns from the image matrix A and image matrix B . We subtract them and set the results as a corresponding element in the sample vector. Finally, we sum all of the elements of

the sample vector and call the result the sampling EMD of two images A and B .

In order to get the best or, at least, near to the EEMD method, we tested different sampling rates, for example 5%, 20%, 30%, and above the whole vector. Finally, we found that 10% is the best sampling rate. Therefore, we randomly select 10% of elements from L_1 vector that will generate just 546 elements.

Sampling EMD is displayed in Pseudocode 2.

Sketching is another option for dimension reduction. In this method, after scanning the data, we multiply the original data of image matrix A and image matrix B by a random matrix R which has either a 0 or 1 for each element, and the subtraction of the resulting matrices forms one element of the sketch vector. We repeat this step k times. The sum of all

```

function Calculate_DREAT(image 1, image 2)
begin
  Initialize  $L_1$ _vector 1
  Initialize  $L_1$ _vector 2
  Initialize index Vector
  Initialize DREAT_Vector 1
  Initialize DREAT_Vector 2
  Initialize DREAT_EMD
  Set  $L_1$ _vector 1 to Calculate_  $L_1$ (image 1)
  Set  $L_1$ _vector 2 to Calculate_  $L_1$ (image 2)
  For  $i = 1$  to 7
    Select 10% indexes of  $G_i$  randomly
    Put the selected indexes in index Vector
  end for
  Select all elements of  $L_1$ _vector 1 whose indexes are in index Vector
  Put the elements in DREAT_Vector 1
  Select all elements of  $L_1$ _vector 2 whose indexes are in index Vector
  Put the elements in DREAT_Vector 2
  Subtract each pair of corresponding elements in DREAT_Vector 1 and DREAT_Vector 2
  Add all subtractions into DREAT_EMD and display it
end

```

PSEUDOCODE 4

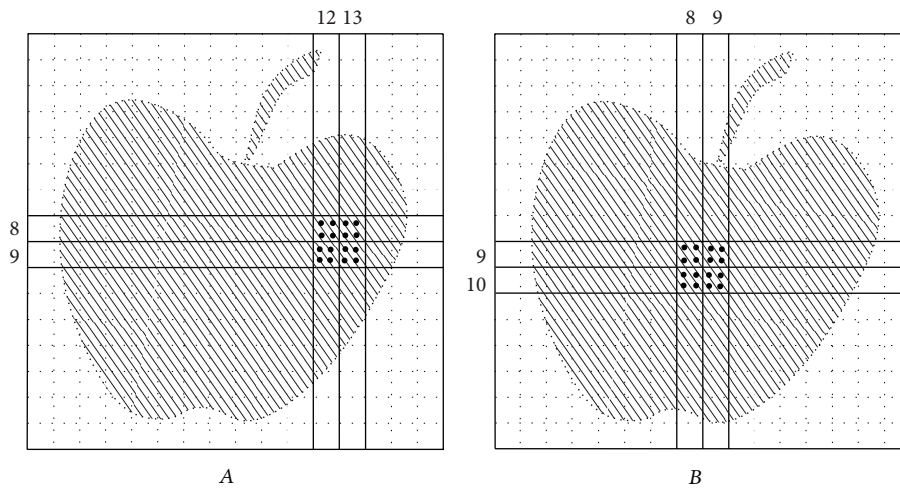


FIGURE 1: Two figures with high similarity.

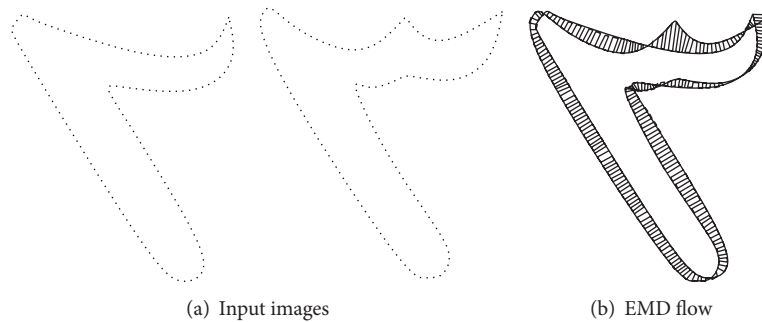


FIGURE 2: Computation of dissimilarity between two input images in Euclidean space and their corresponding EMD flow.

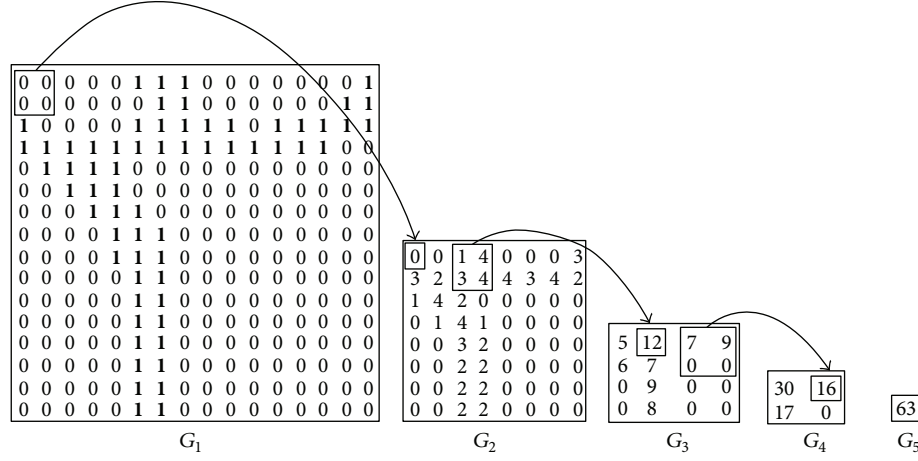
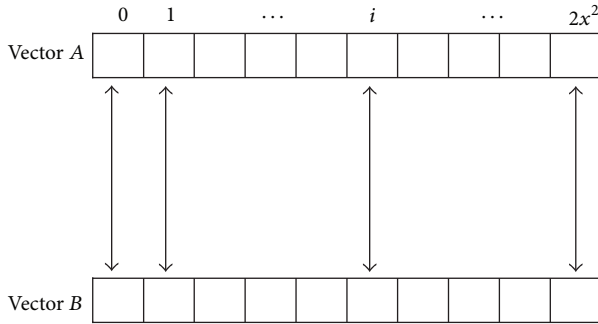
FIGURE 3: Mapping of a 16×16 image into a vector.

TABLE 6: Example of average precision calculation.

| n | Doc no. | Relevance | Precision points |
|-----|---------|-----------|-------------------|
| 1 | 588 | Yes | $P = 1/1 = 1$ |
| 2 | 589 | No | |
| 3 | 576 | Yes | $P = 2/3 = 0.667$ |
| 4 | 590 | No | |
| 5 | 986 | Yes | $P = 3/5 = 0.6$ |
| 6 | 592 | No | |
| 7 | 984 | No | |
| 8 | 988 | Yes | $P = 4/8 = 0.5$ |
| 9 | 578 | Yes | $P = 5/9 = 0.556$ |
| 10 | 985 | No | |

FIGURE 4: Computing $L_1(A, B)$.

elements of the sketch vector is what we call the sketching EMD. Sketching method is illustrated in Figure 6.

Pseudocode 3 shows sketching method.

3.2. Procedures of DREAT. Based on the sampling and sketching experiments, the images' L_1 vectors are heavily tailed where there are many zero elements in former grids and many nonzero elements in latter grids. In the sampling method, we choose samples of the L_1 vector and apply EMD to the samples instead of the whole vector; therefore, the execution

time is reduced. However, the problem is that all elements of vector are sampled at the same rate.

When we go through the vector, most data in the initial sections, such as G_1 and G_2 , contain almost all zeros when compared with the latter sections, such as G_3 , G_4 , and G_5 . We considered this fact to be a heavy-tailed vector. As a result, when we apply the sampling method, the vector might by chance contain almost all zeros, which is meaningless. That is the reason why we need to create a method that will select an equal portion of samples from each part of the grid instead of randomly sampling from the whole.

We called the proposed method as the Dimension Reduction in Embedding by Adjustment in Tail (DREAT), a method that hybrids both the sampling and sketching. For example, suppose we want to select 10% of a vector as a sample vector. In the original sampling method we randomly selected elements of the vector, but, in the DREAT method, we selected only 10% of the elements of each grid part, G_n . In this way, we can select the same portion of all parts of the vector, not only among early elements that have many zeros but also among latter elements with large numbers that are required for recognition.

Table 2 shows the comparison between the number of vector elements in sampling, sketching, and DREAT methods. In DREAT method, we select 10% elements from each grid part because it will produce a number of vector elements that are near to the ones produced by the sampling and sketching methods. Therefore, by using similar idea of adjusting the heavy-tailed vector that we used in the sketching method and combining it with the sampling method we can improve the accuracy without increase of the running time.

This can be expressed in Pseudocode 4.

4. Experiments

In this work, we tested 5 methods: exact EMD, embedded EMD, sampling, sketching, and DREAT. Our image dataset includes bitmap images from Amirkabir University of Iran [23]. The images are scanned from handwritten Persian letters

TABLE 7: Results of 5 methods.

| | MAP | Percentage of first correct recognition | | | | | | |
|--------------|------|---|----------|----------|----------|----------|----------|-----------|
| | | 1st pos. | 2nd pos. | 3rd pos. | 4th pos. | 5th pos. | 6th pos. | >6th pos. |
| Exact EMD | 0.97 | 0.99 | 0.01 | — | — | — | — | — |
| Embedded EMD | 0.85 | 0.90 | 0.01 | 0.02 | 0.03 | 0.01 | 0.01 | 0.02 |
| Sampling | 0.59 | 0.58 | 0.09 | 0.02 | 0.05 | 0.03 | 0.01 | 0.23 |
| Sketching | 0.87 | 0.89 | 0.04 | 0.02 | 0.01 | — | 0.01 | 0.03 |
| DREAT | 0.91 | 0.91 | 0.06 | 0.02 | — | — | — | 0.01 |

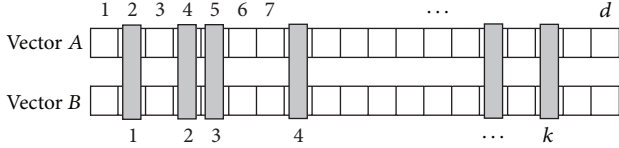


FIGURE 5: Sampling method on vector A and vector B.

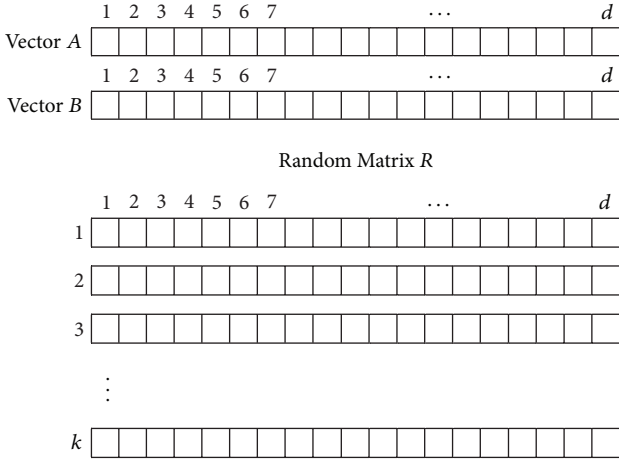


FIGURE 6: Sketching method on vector A and vector B.

and digits. The dataset includes 47 classes which are divided into two parts. The first part of the dataset consists of 35 classes including letter images and the second part of the dataset consists of 12 classes including digit images.

In the dataset, each image is named based on a combination of class number and running number. The first part of image name is its class number as listed in Table 3. Similarity is measured by comparing between the class number of a test image and the queried image. In our work, we only use the second part of the dataset that consists of 12 classes of 5319 handwritten Persian digit images. In Table 3, some of digit images are shown.

We did not use the letter images because Persian letters are very similar in handwritten shape even for a human reader. As some samples are shown in Table 4, letters that have two or three dots are similar to other handwritten letters with a dot. For example in the first sample of Table 4, letter “Cha” is very similar to letter “Ja” because the dots of letter “Cha” stick to each other and they look like one dot. Similarly, letter “Zha” is similar to letter “Za” in some cases. In this case,

the similarity measurement will produce a high distortion. Since that is not a focus of this work, we excluded all the letters.

We divided our dataset into two parts: reference images and test images. The reference set includes 100 images that we randomly selected from the dataset and tested them on the rest of the dataset. In other terms, we remove these 100 reference images from test images’ part. So, we will only find similar images to these reference images not exact ones. For each reference image, we applied the 5 methods and calculated EMD. We then computed the average precision (AP) and placed the results in a table. Finally, we computed the mean average precision (MAP) for all 100 reference images for each method; the results are shown in Table 5.

In each method, some preprocessing steps should be performed. The first step, which is common to all methods, is cropping the white margin of the image. Then, in the next step, the image should be resized to a particular size 64×64 . As a result of pre-processing, we have some images with the same features.

In the first method, exact EMD, what we need is the contour of the image as well as the same number of spots for all of the images. Therefore, 2 additional steps are necessary in this method: getting the contour of the image and removing the additional spots randomly until the same number of spots is achieved, which should be 150 spots.

However, for the methods of embedded EMD, sampling, and sketching, we need the whole image not just its contour. Therefore, preprocessing steps in these methods are only up to the second step. Preprocessing of some of the images is illustrated in Table 5.

5. Results




We computed the mean average precision (MAP) values for the results of 5 different methods applied to 100 query images. Average precision (AP) is the average of the precision values at the points at which each relevant document is retrieved. Precision is defined as

$$\text{Precision} = \frac{\text{number of relevant documents retrieved}}{\text{total number of documents retrieved}}. \quad (2)$$

For example in Table 6 it can be clearly seen that 5 out of 10 documents are relevant; thus, the AP is computed as

$$\text{AP} : \frac{(1 + 0.667 + 0.6 + 0.5 + 0.556)}{5} = 0.665. \quad (3)$$

TABLE 8: Time execution of 3 reference images and estimation for 100 images.

| | Image query 1:  | Image query 2:  | Image query 3:  | Average for 100 images | Total for 100 images |
|--------------|---|---|---|------------------------|----------------------|
| | mm : ss | mm : ss | mm : ss | mm : ss | hh : mm |
| Exact EMD | 22 : 45 | 23 : 11 | 22 : 57 | 22 : 57 | 36 : 41 |
| Embedded EMD | 9 : 18 | 9 : 51 | 9 : 39 | 9 : 36 | 16 : 00 |
| Sampling | 9 : 10 | 9 : 13 | 9 : 11 | 9 : 11 | 15 : 18 |
| Sketching | 11 : 06 | 11 : 20 | 11 : 14 | 11 : 13 | 18 : 41 |
| DREAT | 8 : 45 | 9 : 09 | 9 : 02 | 8 : 58 | 14 : 56 |

Information retrieval systems are frequently judged by their mean average precision (MAP). MAP is the average of the average precision values for a set of queries; it is a performance evaluation measure of information retrieval. Using this measure, we are able to retrieve top-ranked images that are mostly relevant.

The results of all of our experiments are presented in Tables 7 and 8. We compute the AP, which is the average precision of relevant retrieved images among the 10 top-ranked images of the test image sets. In Table 7, the average AP of 100 test images for each method is in the second column. In columns 3 to 8 the percentages of first correct recognition in the first to sixth positions are shown. In the last column, the percentages of first correct recognition in the seventh position and beyond are shown.

In Table 8, the execution times of 5 methods for 3 randomly selected test images are shown. In the fifth column we estimate an average execution time, and in the last column we estimate the execution time for 100 test images for each method. As can be seen in this table, exact EMD has the highest execution time, and embedded EMD reduces this time by half. We can reduce this time by using our proposed methods. The last method, DREAT, achieves the lowest execution time.

6. Conclusion

DREAT is a method that hybrids both the sampling and sketching. In this paper, it shows its usefulness in dimension reduction of sparse and heavy-tailed data. As can be seen in the results, the exact EMD has a MAP value of 0.97; the MAP value is the average of relevant retrieved images among the 10 top-ranked images of 100 images. Although this method is excellent for measuring image similarity, its execution time is very high. By using embedded EMD, a MAP value of 0.85 can be achieved in half the time of exact EMD. Our first proposed method, sampling, reduces the time of execution, but it achieves the poorest MAP value of 0.59. Our second method, sketching, improves the MAP to 0.87 by sacrificing the execution time. Our last method, DREAT, has the lowest execution time and produces one of the best MAP values, which is 0.91.

In general, the results show that dimension reduction techniques like those in this paper, are useful for improving the processing time and matching. DREAT, especially,

combines sketching and sampling where it converts sketches of the data into conditional random samples online in the estimation stage, with the sample size being determined retrospectively. The improvement to the EEMD is useful for overcoming problems with heavily tailed feature vectors.

References

- [1] J. K. K. Samuel, *Lower Bounds for Embedding the Earth Mover Distance Metric into Normed Spaces*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 2005.
- [2] Y. Rubner, C. Tomasi, and L. J. Guibas, "Earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
- [3] J. Xu, Z. Zhang, A. K. H. Tung, and G. Yu, "Efficient and effective similarity search over probabilistic data based on Earth mover's distance," *The VLDB Journal*, vol. 21, no. 4, pp. 535–559, 2012.
- [4] S. Peleg, M. Werman, and H. Rom, "Unified approach to the change of resolution: space and gray-level," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 739–742, 1989.
- [5] D. Mumford, "Mathematical theories of shape: do they model perception," in *Geometric Methods in Computer Vision*, vol. 1570 of *Proceedings of SPIE*, pp. 2–10, International Society for Optics and Photonics, San Diego, Calif, USA, 1991.
- [6] Y. Rubner and C. Tomasi, "Texture metrics," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4601–4607, October 1998.
- [7] S. Cohen and L. Guibas, "Earth mover's distance under transformation sets," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, pp. 1076–1083, September 1999.
- [8] P. Giannopoulos and R. Veltkamp, "A pseudo-metric for weighted point sets," in *Proceedings of the Computer Vision-ECCV*, pp. 89–114, 2002.
- [9] K. Grauman and T. Darrell, "Fast contour matching using approximate Earth mover's distance," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 1, pp. 1220–1227, July 2004.
- [10] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. van Oostrum, "Using transportation distances for measuring melodic similarity," Tech. Rep. uu-cs-2003-024, Computer Science Department, University of Utrecht, Utrecht, The Netherlands, 2003.
- [11] R. Batra and L. Hesselink, "Feature comparisons of 3-D vector fields using Earth mover's distance," in *Proceedings of the IEEE*

- Visualization '99*, pp. 105–114, IEEE Computer Society Press, October 1999.
- [12] P. Indyk and N. Thaper, “Fast image retrieval via embeddings,” in *Proceedings of the International Conference on Computer Vision (ICCV '03)*, 2003.
 - [13] P. Li, T. Hastie, and K. W. Church, *Procedures for Dimension Reduction in l_1* , Department of Statistics, Stanford University, 2006.
 - [14] P. Li and K. W. Church, “Using sketches to estimate associations,” in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 708–715, Association for Computational Linguistics, October 2005.
 - [15] P. Li, K. W. Church, and T. J. Hastie, “Conditional random sampling: a sketch-based sampling technique for sparse data,” in *Advances in Neural Information Processing Systems*, vol. 19, pp. 873–880, 2007.
 - [16] K. Steiglitz and C. H. Papadimitriou, *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1982.
 - [17] C. H. Papadimitriou and U. V. Vazirani, “On two geometric problems related to the travelling salesman problem,” *Journal of Algorithms*, vol. 5, no. 2, pp. 231–246, 1984.
 - [18] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Dover Publications, Prentice-Hall, Englewood Cliffs, NJ, USA, 1998.
 - [19] J. S. Cope and P. Remagnino, “Utilizing the hungarian algorithm for improved classification of high-dimension probability density functions in an image recognition problem,” in *Advanced Concepts for Intelligent Vision Systems*, pp. 268–277, Springer, New York, NY, USA, 2012.
 - [20] A. Naor and G. Schechtman, “Planar earthmover is not in L_1 ,” *SIAM Journal on Computing*, vol. 37, no. 3, pp. 804–826, 2007.
 - [21] W. Hong and T. S. Chen, “A novel data embedding method using adaptive pixel pair matching,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 1, pp. 176–184, 2012.
 - [22] S. Łukasik and P. Kulczycki, “An algorithm for sample and data dimensionality reduction using fast simulated annealing,” in *Advanced Data Mining and Applications*, vol. 7120 of *Lecture Notes in Computer Science*, pp. 152–161, 2011.
 - [23] H. Khosravi and E. Kabir, “Introducing a very large dataset of handwritten Farsi digits and a study on their varieties,” *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1133–1141, 2007.