

Image Enlargement Using Multiple Sensors

Guest Editors: Wei Wu, Gwanggil Jeon, Marco Anisetti, and Chehri Abdellah

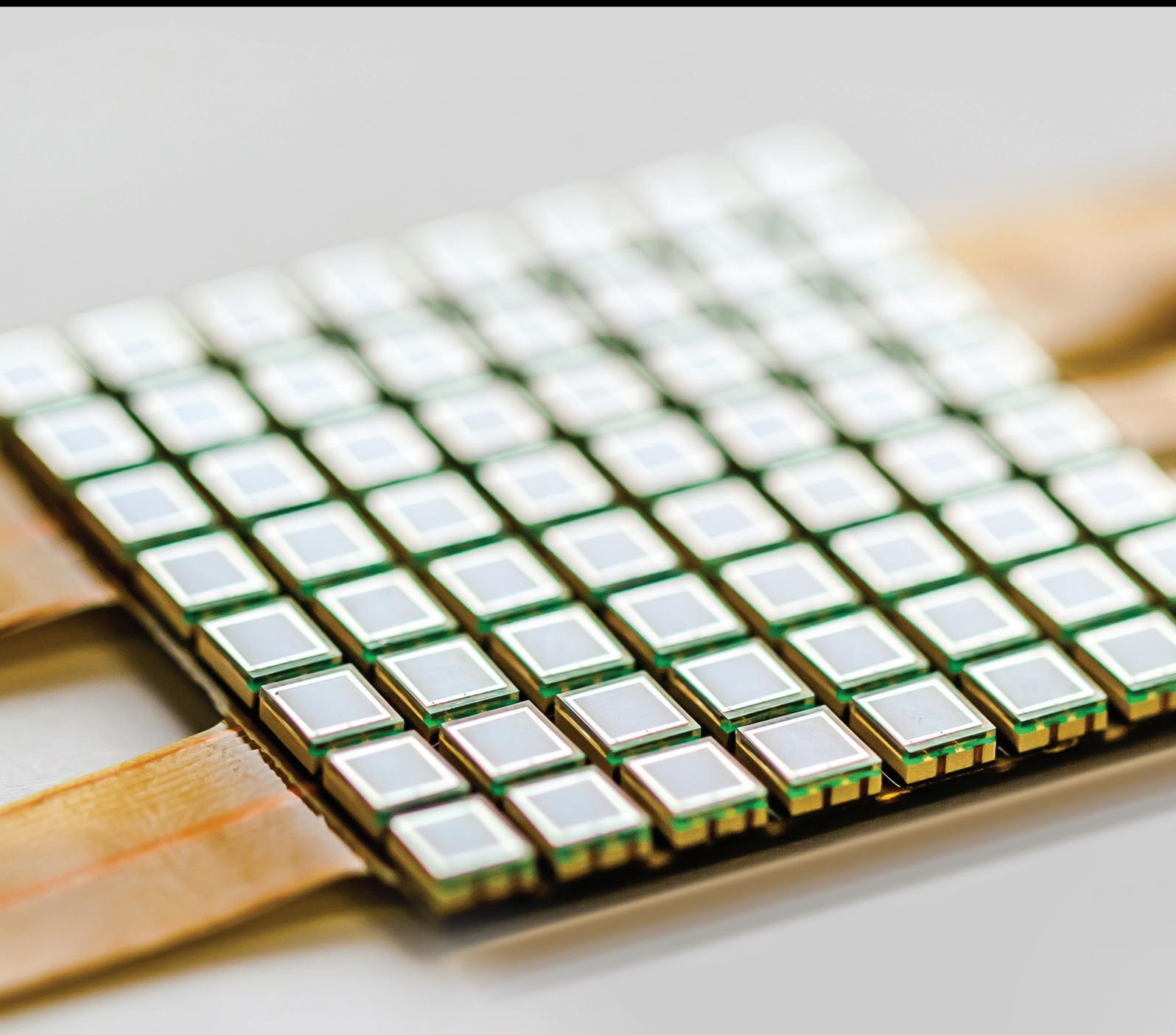




Image Enlargement Using Multiple Sensors

Journal of Sensors

Image Enlargement Using Multiple Sensors

Guest Editors: Wei Wu, Gwanggil Jeon, Marco Anisetti,
and Chehri Abdellah



Copyright © 2016 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "Journal of Sensors." 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

Harith Ahmad, Malaysia
Francisco J. Arregui, Spain
Francesco Baldini, Italy
Fernando Benito-Lopez, Ireland
Romeo Bernini, Italy
Shekhar Bhansali, USA
Wojtek J. Bock, Canada
Hubert Brändle, Switzerland
Davide Brunelli, Italy
Belén Calvo, Spain
Stefania Campopiano, Italy
Jian-Nong Cao, Hong Kong
Sara Casciati, Italy
Chi Chiu Chan, Singapore
Nick Chaniotakis, Greece
Nicola Cioffi, Italy
Elisabetta Comini, Italy
Marco Consales, Italy
Jesus Corres, Spain
Andrea Cusano, Italy
Dzung V. Dao, Japan
Catherine Dehollain, Switzerland
Manel del Valle, Spain
Ignacio Del Villar, Spain
Utkan Demirci, USA
Junhang Dong, USA
Abdelhamid Errachid, France
Stephane Evoy, Canada
Vittorio Ferrari, Italy
Luca Francioso, Italy
Laurent Francis, Belgium
Mohammad Reza Ganjali, Iran

Wei Gao, Japan
Michele Giordano, Italy
Banshi D. Gupta, India
María del Carmen Horrillo, Spain
Wieslaw Jakubik, Poland
Hai-Feng Ji, USA
Kouros Kalantar-Zadeh, Australia
Sher Bahadar Khan, KSA
Sang Sub Kim, Republic of Korea
Challa Kumar, USA
Hiroki Kuwano, Japan
Laura M. Lechuga, Spain
Chengkuo Lee, Singapore
Chenzhong Li, USA
Eduard Llobet, Spain
Yu-Lung Lo, Taiwan
Oleg Lupan, Moldova
Eugenio Martinelli, Italy
Jose R. Martinez-de Dios, Spain
Yasuko Y. Maruo, Japan
Ignacio R. Matias, Spain
Mike McShane, USA
Igor L. Medintz, USA
Fanli Meng, China
Aldo Minardo, Italy
Joan Ramon Morante, Spain
Lucia Mosiello, Italy
Masayuki Nakamura, Japan
Vincenzo Paciello, Italy
Marimuthu Palaniswami, Australia
Alberto J. Palma, Spain
Gyuhae Park, Republic of Korea

Alain Pauly, France
Giorgio Pennazza, Italy
Michele Penza, Italy
Andrea Ponzoni, Italy
Biswajeet Pradhan, Malaysia
Ioannis Raptis, Greece
Christos Riziotis, Greece
Maria Luz Rodríguez-Méndez, Spain
Albert Romano-Rodríguez, Spain
Josep Samitier, Spain
Giorgio Sberveglieri, Italy
Luca Schenato, Italy
Michael J. Schöning, Germany
Andreas Schütze, Germany
Woosuck Shin, Japan
Pietro Siciliano, Italy
Vincenzo Spagnolo, Italy
Weilian Su, USA
Tong Sun, UK
Hidekuni Takao, Japan
Isao Takayanagi, Japan
Pierre Temple-Boyer, France
Guiyun Tian, UK
Suna Timur, Turkey
Hana Vaisocherova, Czech Republic
Qihao Weng, USA
Matthew J. Whelan, USA
Stanley E. Woodard, USA
Hai Xiao, USA
Hyeonseok Yoon, Republic of Korea

Contents

Image Enlargement Using Multiple Sensors

Wei Wu, Marco Anisetti, Chehri Abdellah, and Gwanggil Jeon
Volume 2016, Article ID 1498796, 3 pages

An Edge-Preserved Image Denoising Algorithm Based on Local Adaptive Regularization

Li Guo, Weilong Chen, Yu Liao, Honghua Liao, and Jun Li
Volume 2016, Article ID 2019569, 6 pages

Spectrogram Image Analysis of Error Signals for Minimizing Impulse Noise

Jeakwan Kim, Yunseon Choi, and Young-Sup Lee
Volume 2016, Article ID 4935694, 9 pages

Multiscale and Multitopic Sparse Representation for Multisensor Infrared Image Superresolution

Xiaomin Yang, Kai Liu, Zhongliang Gan, and Binyu Yan
Volume 2016, Article ID 7036349, 14 pages

Active Suppression of Narrowband Noise by Multiple Secondary Sources

Seokhoon Ryu, Yun Jung Park, and Young-Sup Lee
Volume 2016, Article ID 6276828, 9 pages

A Tile-Based EGPU with a Fused Universal Processing Engine and Graphics Coprocessor Cluster

Yang Wang, Li Zhou, Tao Sun, Yanhu Chen, Lei Wang, and Shaotao Sun
Volume 2016, Article ID 7281031, 9 pages

Facial Feature Extraction Using Frequency Map Series in PCNN

Rencan Nie, Dongming Zhou, Min He, Xin Jin, and Jiefu Yu
Volume 2016, Article ID 5491341, 9 pages

Visibility Enhancement of Scene Images Degraded by Foggy Weather Conditions with Deep Neural Networks

Farhan Hussain and Jechang Jeong
Volume 2016, Article ID 3894832, 9 pages

RS485 Image Sensor for Digital Cinema System

Eunju Kim, Seokhoon Kang, and Sangsoon Lee
Volume 2016, Article ID 9721820, 8 pages

Feature Coverage Indexes for Dual Homography Estimation in Constructing Panorama Image

Kyungkoo Jun and Sijung Kim
Volume 2016, Article ID 1350750, 7 pages

Patch Classifier of Face Shape Outline Using Gray-Value Variance with Bilinear Interpolation

Seokhoon Kang and Seonwoon Kim
Volume 2016, Article ID 3825931, 10 pages

Efficient Deep Neural Network for Digital Image Compression Employing Rectified Linear Neurons

Farhan Hussain and Jechang Jeong
Volume 2016, Article ID 3184840, 7 pages



Modified Hybrid Freeman/Eigenvalue Decomposition for Polarimetric SAR Data

Shuang Zhang, Shuang Wang, and Bo Chen

Volume 2015, Article ID 168720, 7 pages

Superresolution of Hyperspectral Image Using Advanced Nonlocal Means Filter and Iterative Back Projection

Jin Wang, Zhensen Wu, and Young-Sup Lee

Volume 2015, Article ID 943561, 5 pages

Editorial

Image Enlargement Using Multiple Sensors

Wei Wu,¹ Marco Anisetti,² Chehri Abdellah,³ and Gwanggil Jeon⁴

¹College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China

²Dipartimento di Informatica, Università degli Studi di Milano, Via Bramante 65, 26013 Crema, Italy

³School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Avenue, Ottawa, ON, Canada K1N 6N5

⁴Department of Embedded Systems Engineering, College of Information and Technology, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon 22012, Republic of Korea

Correspondence should be addressed to Gwanggil Jeon; gjeon@inu.ac.kr

Received 10 September 2015; Accepted 13 September 2015

Copyright © 2016 Wei Wu 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.

Image sensing is generally performed with multiple spectral sensors. For example, combination of three sensors (red, green, and blue) is used for color image reproduction, and electrooptical and infrared sensors are used for surveillance and satellite imaging, respectively. The resolution of each sensor can be intensified by taking the other sensors into account and applying correlations between different sensors. There are various successful applications of image enlargement using multiple sensors and even multimodal sensors. However, there still are several open issues in sensor processing which can be explained by signal processing-based image enlargement using redundancy among the sensors.

The articles contained in the present issue include both reviews and basic scientific studies focused on image enlargement using multiple spectral sensors to improve image quality. This issue comprises the description of image enlargement methodologies which uses embedded system, signal processing (audio and feature extraction), denoising, regularization for error minimizing, super resolution, and artificial enhancement such as neural network.

Pulse coupled neural network (PCNN) has been widely used in image processing. The 3D binary map series (BMS) generated by PCNN can be treated as the basis of extracting 1D oscillation time series (OTS) for an image, since they effectively describe image features such as edges and regional distribution.

However, the traditional methods using BMS require improvements for being effectively applied in some specific

field of image processing such as face recognition. The contribution by R. Nie et al. "Facial Feature Extraction Using Frequency Map Series in PCNN" computed 2D entropy for every map in frequency map series (FMS) obtained from BMS, and then 3D FMS was transformed into 1D OTS (OTS-FMS). The authors proved that OTS-FMS feature has good geometry invariance for the facial image and contains the space structure information of the image.

The face shape model is used together with the scope of detection and recognition for many applications; however, there are still some imitations in the current state of the art solutions especially for the aspects related to feature detection and tracking. The contribution by S. Kang and S. Kim "Patch Classifier of Face Shape Outline Using Gray-Value Variance with Bilinear Interpolation" proposed a method to classify whether the outline in a face shape model is properly fitted to feature points. The authors implemented a classifier using this characteristic and achieved good accuracy through the patch classifier.

Image is a 2D signal; thus any type of signal processing approaches even audio processing techniques can be used as inspiration for developing image processing. The contribution by S. Ryu et al. "Active Suppression of Narrowband Noise by Multiple Secondary Sources" extensively investigated in both theory and real-time control experiment on the active suppression of narrowband noise. After analyzing the primary and secondary paths of the duct system, an acoustic narrowband signal was chosen as a primary noise and

the impulse responses were implemented as the secondary path models. The authors proved that the lower quality factor cancels the more primary noise as defined in the theory although the attenuation levels are not exactly and inversely proportional to the quality factor.

Image processing is nowadays applied in many application fields including automotive. Camera-based advanced driver assistance systems (ADAS) have been introduced to assist the drivers and ensure their safety under various driving conditions like in the case of low visibility and lower contrast due to foggy conditions. The contribution by F. Hussain and J. Jeong “Visibility Enhancement of Scene Images Degraded by Foggy Weather Conditions with Deep Neural Networks” presented a new approach to provide a solution to this problem by employing deep neural networks and mathematically modelling fog in an image. This paper achieved two advantages, “real-time operation” and “robustness/generalization for various unseen image data.”

Image enlargement method as the one specific for ADAS system is in general simultaneously applied with image denoising. Image denoising methods are often based on the minimization of an appropriately defined energy function. The contribution by L. Guo et al. “An Edge-Preserved Image Denoising Algorithm Based on Local Adaptive Regularization” adaptively adjusted denoising degree of noisy image by adding spatial variable fidelity term.

The spectrogram image analysis of error signals can be used in the active noise suppression. The contribution by J. Kim et al. “Spectrogram Image Analysis of Error Signals for Minimizing Impulse Noise” presented the theoretical and experimental study on the spectrogram image analysis of error signals for minimizing the impulse input noises in the active suppression of noise. The authors proved that the suggested algorithm worked with a nice stability and performance against impulse noises, applying it for practical active noise control systems.

There have been many image enlargement methods, which are based on sparse coding especially in single image superresolution reconstruction. When applied to infrared (IR) images, the traditional sparse representation-based (SR) image reconstruction methods suffer from (i) lack of detailed information, (ii) fixed size patch-based dictionary, (iii) universal and overcomplete dictionary. The contribution by X. Yang et al. “Multiscale and Multitopic Sparse Representation for Multisensor Infrared Image Superresolution” overcame the above issues by combining the information from multisensors to improve the resolution of the IR image. In addition, they used multiscale patches to represent the image in a more efficient manner. Finally they partition the natural images into documents. Then they group such documents to determine the inherent topics using pLSA and to learn the sparse dictionary of each topic.

The Embedded Graphics Processing Unit (EGPU) has assumed more processing tasks within embedded devices, requiring high computational capacity. The contribution by Y. Wang et al. “A Tile-Based EGPU with a Fused Universal Processing Engine and Graphics Coprocessor Cluster” proposed general-purpose computing and 3D graphics rendering. The authors demonstrated that the proposed EGPU can be used in

a System on Chip (SoC) configuration connected to sensors with the scope of accelerating its processing and creating a proper balance between performance and cost.

Stitching is one of efficient methods for images enlargement by stitching adjacent images which contain overlapping regions even though they are obtained through separate image sensors. The contribution by K. Jun and S. Kim “Feature Coverage Indexes for Dual Homography Estimation in Constructing Panorama Image” proposed dual homography method. The authors presented three feature coverage indexes which evaluate the stitching performance of feature detectors and predict the outcomes of the stitching.

The theoretical resolution of 35 mm digital cinema system is greater than that of 2K (2048×1080) digital cinema. Today’s digital cinema system equips hundreds of 4D chairs and the environmental directors; it is nearly infeasible for the legacy system to control. The contribution by E. Kim et al. “RS485 Image Sensor for Digital Cinema System” designed and implemented a new system, which makes hundreds of 4D chairs and the environmental directors be controlled simultaneously by exploiting RS485 network topology and its repeaters.

Digital image compression plays an extremely important role in the transmission and storage of digital image data. Image compression is the process of effectively coding digital images to reduce the number of bits required to represent an image. The contribution by F. Hussain and J. Jeong “Efficient Deep Neural Network for Digital Image Compression Employing Rectified Linear Neurons” proposed a compression technique for still digital images which uses deep neural networks (DNNs) and rectified linear units (ReLU). The authors proved that ReLUs well establish an efficient gradient propagation and the proposed system is efficient in terms of computations making these networks suitable for real-time compression systems.

Due to the rapid advancement of the airborne sensors and spaceborne sensors, large volumes of fully polarimetric synthetic aperture radar (PolSAR) data are available, but they are too complex to interpret difficultly. The contribution by S. Zhang et al. “Modified Hybrid Freeman/Eigenvalue Decomposition for Polarimetric SAR Data” proposed modified hybrid Freeman/eigenvalue decomposition method for the coherency matrix derived from the fully PolSAR sensors. The proposed method used a real unitary transformation on the coherency matrix to release correlations between the copolarized term and cross polarized term.

Hyperspectral image obtains terrestrial information in various contiguous and narrow spectral bands, and it is attracting more and more researchers due to its wide applications in aerial and space imagery fields. The contribution by J. Wang et al. “Superresolution of Hyperspectral Image Using Advanced Nonlocal Means Filter and Iterative Back Projection” introduced an efficient superresolution algorithm based on advanced nonlocal means (NLM) filter and iterative back projection for hyperspectral image. The authors proved that their method is able to recover the high-resolution image by iteratively minimizing the reconstruction error from the given low-resolution image which is blurred due to the noise.

We hope that this special issue would shed light on major developments in the area of image enlargement using multiple sensors and attract attention by the scientific community to pursue further investigations leading to the rapid implementation of these sensor technologies.

Acknowledgments

We would like to express our appreciation to all the authors for their informative contributions and the reviewers for their support and constructive critiques in making this special issue possible. This work was supported under the framework of international cooperation program managed by National Research Foundation of Korea (no. 2014000756) and by National Natural Science Foundation of China (no. 61271330 and no. 6141101009).

*Wei Wu
Marco Anisetti
Chehri Abdellah
Gwanggil Jeon*

Research Article

An Edge-Preserved Image Denoising Algorithm Based on Local Adaptive Regularization

Li Guo,¹ Weilong Chen,² Yu Liao,¹ Honghua Liao,¹ and Jun Li¹

¹Information Engineering Department, Hubei University for Nationalities, Enshi 445000, China

²Digital Media College, Sichuan Normal University, Chengdu 610068, China

Correspondence should be addressed to Weilong Chen; 312790467@qq.com

Received 19 March 2015; Revised 1 July 2015; Accepted 12 July 2015

Academic Editor: Marco Anisetti

Copyright © 2016 Li Guo 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.

Image denoising methods are often based on the minimization of an appropriately defined energy function. Many gradient dependent energy functions, such as Potts model and total variation denoising, regard image as piecewise constant function. In these methods, some important information such as edge sharpness and location is well preserved, but some detailed image feature like texture is often compromised in the process of denoising. For this reason, an image denoising method based on local adaptive regularization is proposed in this paper, which can adaptively adjust denoising degree of noisy image by adding spatial variable fidelity term, so as to better preserve fine scale features of image. Experimental results show that the proposed denoising method can achieve state-of-the-art subjective visual effect, and the signal-noise-ratio (SNR) is also objectively improved by 0.3–0.6 dB.

1. Introduction

In the process of image acquisition and transmission, all recording devices have traits which make them susceptible to noise. Noise deteriorates the quality of image and causes difficulty in image observation, feature extraction, and image analysis. In order to effectively reduce noise, some filters such as mean filter and Gaussian filter are applied on the noisy image. It is limited that filters can lose large edge and texture information of image in denoising process. In order to avoid this problem, many researchers have been working on different denoising methods. Weickert proposed a partial differential equation (PDE) method which can effectively remove noise [1]. Chan and Esedoglu proposed a total variation method based on L1 norm; it produced stair effect in smoothing area [2]. Bo and Li used a symmetric four-order PDE method to achieve acceptable denoising result [3]. Xu and Wang introduced nonlocal means into regularization to obtain a denoising version; this method oversmoothed image edges and details [4]. Gupta and Kumar proposed a generalized total variation denoising model, which can remove false edge, but it is sensitive to the selection of P factor [5]. Liu and

Zeng proposed a map image adaptive regularization denoising method to get a good visual effect [6]. These methods can improve the image denoising quality in some extent but destroy high-frequency information of image inevitably. In recent years, Liu and Huang proposed a new nonlocal total variation regularization algorithm for image denoising [7]. Chen et al. proposed an adaptive denoising model by regulating regularization and fidelity total variation [8]. Suman used adaptive median filter into image denoising and got good result [9]. Yan and Lu added least squares fidelity in imaging denoising by generalized total variation regularization [10]. Anilet and Hati combined curvelet transform and wiener filter to effectively reduce noise in an image [11]. Liu et al. discussed many methods in image fusion and image denoising, which all are useful and popular [12]. Abovementioned image denoising methods achieve acceptable denoising effect, but the selection of an appropriate regularized factor and reasonable iteration is still a problem. For this reason, the goal of this study is to examine efficient and reliable image denoising algorithms. This paper proposed an image denoising method based on local adaptive regularization, which can adaptively adjust denoising according to different area of noisy image

and better protect the texture and details of image, so as to achieve state-of-the-art denoising effect.

The rest of this paper is organized as follows. The introduction of regularization (especially for total variation regularization) is described in Section 2. Further, Section 3 details main idea of the proposed method in this paper. And experimental results are presented in Section 4. Finally, summary and outlook are discussed in Section 5, that is, the conclusion of this paper.

2. Total Variation Denoising

In signal processing, total variation denoising is remarkably effective at simultaneously preserving edges whilst smoothing away noise in flat regions, even at low signal-to-noise ratios. It is based on the principle that signals with excessive and possibly spurious detail have high total variation (TV); that is, the integral of the absolute gradient of the signal is high. According to this principle, reducing the total variation of the signal subject to its being a close match to the original signal removes unwanted detail whilst preserving important details such as edges. The concept of total variation was pioneered by Rudin et al. [13]. In area of image processing, suppose f is the original image, f_0 is the noisy version of f , and this relation can be mathematically expressed by

$$f_0(x, y) = f(x, y) + n(x, y). \quad (1)$$

Here $n(x, y)$ is random noise with zero mean and σ^2 variation. At the same time, we can, for example, define the total variation as $V(x, y) = \sum_n |f^{n+1}(x, y) - f^n(x, y)|$; the goal of TV denoising is to find an approximation, which is smaller but close to the one before. That is, TV denoising is a minimization process; it explores the equal state of energy function relative to the TV norm of f and the regularization of f_0 , expressed in

$$E_{\text{TV}} = \int_{\Omega} \left[(|\nabla f|) + \frac{1}{2} \lambda (f - f_0)^2 \right] dx dy. \quad (2)$$

Here Ω represents the domain of image, all pixels $(x, y) \in \Omega$. Normally the TV of ideal image is smaller than noisy image, so minimizing TV can reduce the noise of image. Based on this principle, (2) can be equaled as follows:

$$\frac{\lambda}{2} \int_{\Omega} (f - f_0)^2 dx dy + \int_{\Omega} \sqrt{f_x^2 + f_y^2} dx dy = 0. \quad (3)$$

The first term in (3) is data fidelity term, which can retain characteristics of the original image and reduce distortion. The second term in (3) is regularization term, which depends on noise level and balance denoising and smoothing. The Euler-Lagrange equation derivate from (3) is represented by

$$\lambda (f - f_0) - \nabla \cdot \left(\frac{\nabla f}{|\nabla f|} \right) = 0. \quad (4)$$

Here $1/|\nabla f|$ is a diffusion coefficient. In edge of image, large ∇f will lead to small diffusion coefficient, so the diffusion along edge is weak to preserve edge of image. In smoothing area of image, small $|\nabla f|$ will lead to large diffusion

coefficient, so the diffusion in smoothing area is strong to remove noise in image. The regularization parameter λ plays a critical role in the denoising process. When $\lambda = 0$, there is no denoising and the result is identical to the input signal. As $\lambda \rightarrow \infty$, however, the total variation term plays an increasingly strong role, which forces the result to have smaller total variation, at the expense of being less like the input (noisy) signal. Thus, the choice of regularization parameter is critical to achieving just the right amount of noise removal.

3. Image Denoising Method Based on Local Adaptive Regularization

Referring to the classical TV model which described in Section 2, it can reduce noise by energy constraint, however, the selection of an appropriate regularization factor is also a difficult problem. High λ brings oversmoothing and small λ reduces noise ineffectively. In order to handle impulsive noisy image, an appropriate regularization factor λ must be obtained to reduce different noise in image, which aims at balance of data fidelity term and regularization term. For these discussions, a novel image denoising method based on local adaptive regularization is proposed here. According to noise level in different image area, it defines a space variable energy function and can adaptively adjust denoising degree. In order to clearly elaborate idea, the main steps described are as follows.

Step 1 (global residual noisy energy computation). This step produces the residual error of noise in input image. Suppose input noisy image is $f_0(x, y)$. Estimated denoising version of $f_0(x, y)$ is obtained by classical TV regularization, which is expressed as $f'_0(x, y)$. Then, global residual noisy energy f_r can be computed by expression $f_r = f_0(x, y) - f'_0(x, y)$. Therefore, the mean value of global residual error f_r is computed and named as M_r .

Step 2 (local energy computation). On the basis of first step, local variance of residual image can be expressed by $P_{\text{LV}}(x, y) = (1/|\Omega|) \int_{\Omega} f_r(x, y) w(x, y) dx dy$, where $w(x, y)$ is a normalized and radial symmetric smoothing-window and $\int_{\Omega} w(x, y) dx dy = 1$. Suppose $S(x, y) = \sigma^4/P_{\text{LV}}$ to get prior information of noise energy in noisy image; here σ is noise standard deviation of input noisy image $f_0(x, y)$.

Step 3 (iteratively compute regularization factor to achieve local adaptive regularization). Define $n = 1$ and suppose $f^{(n)'}_0 = f'_0$ ($n = 1$). Compute $Q^n = (f^{(n)'}_0 - f_0) \text{div}(\phi'(\nabla f^{(n)'}_0 / |f^{(n)'}_0|))$; regularization factor λ is then computed by $\lambda^n(x, y) = Q^n(x, y)/S(x, y)$. In this way, each iteration of λ can be applied to (4) to obtain an estimated denoising version. When condition of convergence is satisfied, the final denoising result can be achieved.

4. Experiments and Discussion

In this section, we validate the potential of the proposed method by simulated noisy image experiment and real noisy

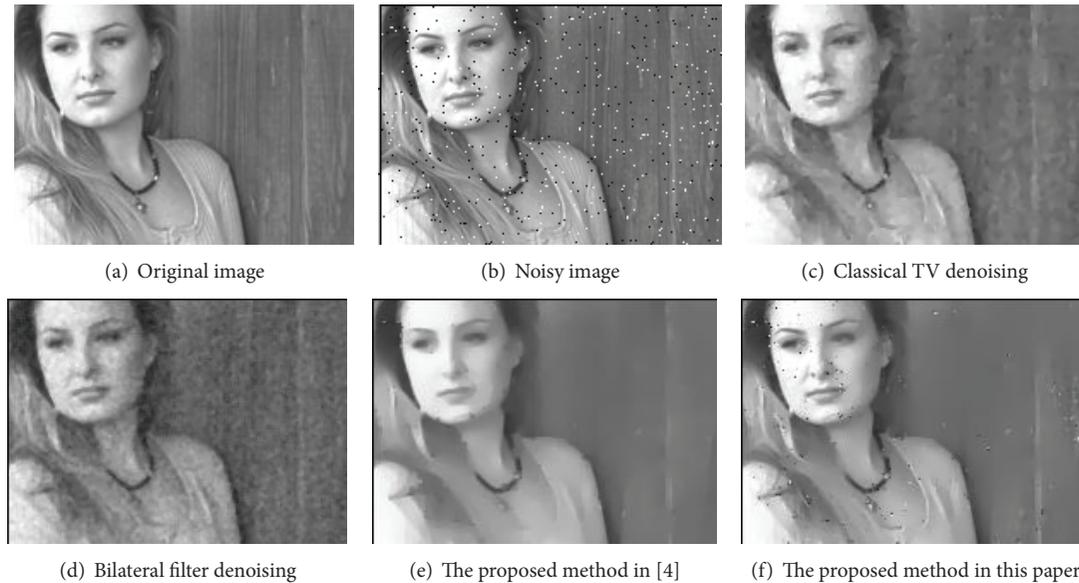


FIGURE 1: Denoising comparison of salt and pepper noisy image. (The SNR value of this comparison is as follows: noisy image SNR = 8.1867, classical TV denoising SNR = 15.2426, bilateral filter denoising SNR = 15.0410, the proposed method in [4] SNR = 15.7141, and the proposed method in this paper SNR = 15.9832.)

image experiment. The comparison we provide here is the denoising results by classical TV denoising, bilateral filtering, and the denoising method proposed in [4].

4.1. Simulated Noisy Image

4.1.1. Salt and Pepper Noise. Firstly, the test image woman was additive with salt and pepper noise, at which variance is 0.02. In order to better show different denoising comparison, experimental result shown here is the enlarged local area of woman.

Figure 1(a) is the clear original woman image, Figure 1(b) is seen to contain a large number of salt and pepper noises and many noise points contaminate image, Figure 1(c) is the denoising result of classical TV method, here texture in the right smoothing part of image is preserved, and conversely woman's face detailed information is lost and edge of image is not ideal, Figure 1(d) is the denoising result by bilateral filter, which is more fuzzy and noisy, and the visual effect is bad, and Figure 1(e) is the denoising result by method in [4] and noise is removed in this result; however the overall effect is oversmoothing. The denoising result in Figure 1(f) is clear and details are preserved well, although there are still a few grain noises, but it does not affect the general viewing effect.

4.1.2. Gaussian Noise. In this experiment, test image Lena is added into Gaussian noise with 0 mean and variation is 0.06. As the same way to experiment 1, Figure 2 also shows the extracted local Lena's face. From this way, the different denoising result is compared more clearly here.

When compared to these different denoising methods, Figure 2(a) is the local part of the original Lena image, Figure 2(b) is local part of noisy image added with Gaussian

noise and a large number of noises can be seen in it, Figure 2(c) is the local part of denoising result by classical TV method and the visual effect is bad, Figure 2(d) is the denoising result of bilateral filter and it also contains quite a few noises in denoising result, and Figure 2(e) is the denoising result of method in [4] and it well reduces noise in noisy image, but the detailed information is lost and overall appearance is too smoothing. Figure 2(f) is the local part of denoising result by the proposed method in this paper, it can be seen that the detailed information (such as feather in cap, eye, and mouth edge) is kept well, the people's viewing effect is very good, and noise is reduced selectively in a natural manner.

4.1.3. Random Noise. In order to further comparison of denoising method, standard test image Lena is added into random noise with variance being 20. From this way, a simulated noisy image is obtained in this experiment. For better illustrating the effectiveness of the proposed algorithm, the local part of denoising result is extracted to be shown in Figure 3.

Figures 3(a) and 3(b) show the local part of original version and noisy version of Lena image, and the noisy version is additive with random noise. Figure 3(c) is the local part of denoising result by classical TV method and it looks more fuzzy, Figure 3(d) maintains a noisy state, which contains a large number of unremoved noises, Figure 3(e) is the local part of denoising result by denoising method in [4] and it can effectively reduce noise but smoothed the details simultaneously, and Figure 3(f) is the local part of denoising result by the proposed method in this paper, it can be seen that the detailed information is protected well (such as eye and the edge of mouth), and this visual effect is much acceptable and outweighs other denoising results.



FIGURE 2: Comparison of Gaussian noisy image and denoising results by different methods. (The SNR value of this comparison is as follows: noisy image SNR = 7.5325, classical TV denoising SNR = 13.5687, bilateral filter denoising SNR = 13.3146, the proposed method in [4] SNR = 14.1413, and the proposed method in this paper SNR = 14.2835.)

TABLE 1: Comparison of denoising results by different methods in terms of SNR value.

	Salt and pepper noise	Gaussian noise	Random noise	Real noisy image	Average SNR
Noisy image	8.1867	7.5325	7.1122	6.3736	7.3013
Classical TV	15.2426	13.5687	14.3264	11.4326	13.6425
Bilateral filter	15.0410	13.3146	14.2720	11.2724	13.4750
Method in [4]	15.7141	14.1413	14.6837	11.7538	14.0732
Method proposed in this paper	15.9832	14.2835	14.8013	12.1015	14.2925

4.2. Real Noisy Image. The comparable experiments provided in Section 4.1 reflect favorite denoising result of the proposed method in this paper. In the following experiment, a real noisy image experiment would be provided in this section. From this way, the denoising method proposed in this paper can be considered more effective and available in common applications. In order to better reflect the impact of noise, we use camera directly for the display of computer. The main content of shoot image is text, including a large number of edges and detailed information. These experimental results are shown in Figure 4. The comparable denoising methods are the same as those in Section 4.1.

In Figure 4(a) is original noisy image, Figure 4(b) is denoising result by classical TV method, its visual effect is bad, Figure 4(c) is the denoising result by bilateral filter, it obviously contained much noise, and the denoising result of Figure 4(d) is a little smoothing. Figure 4(e) is local part of denoising result by the proposed method in this paper, here

noise is well removed, and the edge of text is well preserved, so that the visual effect is better than other methods.

In above experiments including three different noises, pepper and salt noise is a pulse noise which approximately is equal to amplitude but is randomly distributed in image; there are some clean points and contaminated points. Gaussian noise is almost equally distributed in each part of an image. The random noise is also known as background noise, which is the accumulation of a lot of random fluctuations caused by time; the value is unpredictable. The different noises have various characteristics, so many denoising methods can effectively reduce one kind but cannot deal with the others. Besides subjective visual effect, with the aim of deeply discussing the experiments and further show the effectiveness of the proposed methods, the objective SNR values of different denoising methods are also given in Table 1. From the average of SNR value provided in Table 1, we can see the average SNR value of proposed method is the highest; it rises about



FIGURE 3: Comparison of denoising results by different methods. (The SNR value of this comparison is as follows: noisy image SNR = 7.1122, classical TV denoising SNR = 14.3264, bilateral filter denoising SNR = 14.2720, the proposed method in [4] SNR = 14.6837, and the proposed method in this paper SNR = 14.8013.)

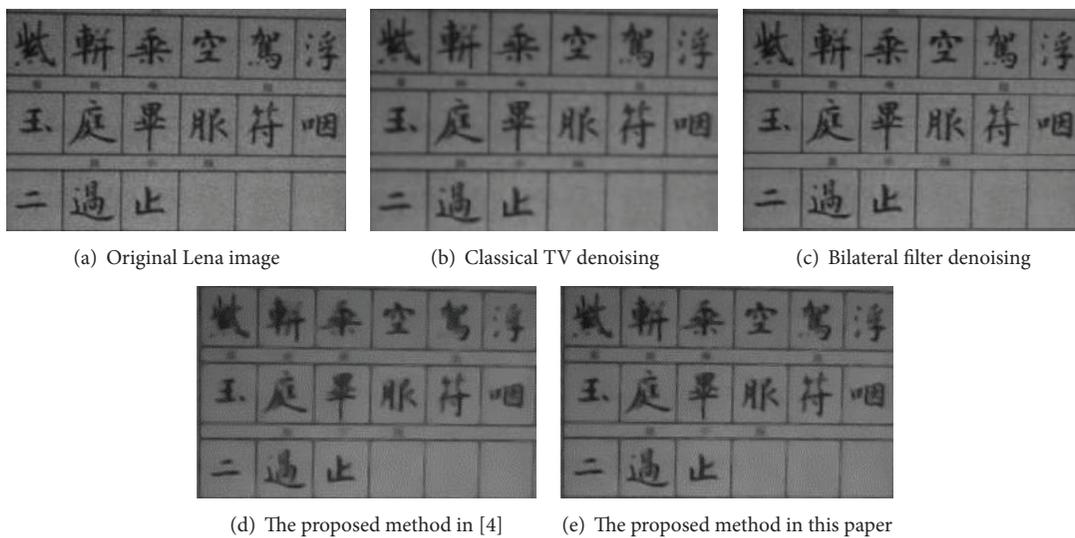


FIGURE 4: Comparison of different denoising methods of real noisy image. (The SNR value of this comparison is as follows: noisy image SNR = 6.3736, classical TV denoising SNR = 11.4326, bilateral filter denoising SNR = 11.2724, the proposed method in [4] SNR = 11.7538, and the proposed method in this paper SNR = 12.1015.)

0.3–0.6 dB compared to other referred methods. From the comparable results in Table 1, it also can be seen that the objective result is consistent with the subjective value. The result of best visual effect also has the highest SNR value.

Based on the subjective visual effect and objective SNR value in simulated noisy and real noisy experiments, it can be seen that the denoising result by bilateral filter is worse than other methods. Above several comparable methods use

global and fixed denoising pattern, which results in some details being missing and oversmoothing. The denoising method proposed in this paper is more effective and robust than other several methods. According to characteristic of different noisy image, the proposed method can adaptively adjust the regularization factor λ , the solution of λ through iteration process, and each iteration updates the value of λ by input image and current denoising image, so as to achieve the optimal regularization factor λ . Furthermore, the appropriate λ can better balance the fidelity term and regularization term and then ensure reliable reduction of noise as well as protecting more fine details in image.

5. Conclusion

The widely used denoising algorithm based on global energy constraint can obtain good denoising effect on the simple structural image. But these methods can easily lose key detailed information when removing noise. For these reasons, an image denoising method based on local adaptive regularization is proposed in this paper, it can effectively control denoising degree in different area of noisy image by constraining residual local energy of energy function, and state-of-the-art denoising result can be obtained by this way. In this method, detailed information of image can also be well preserved at the process of denoising. From experiments of simulated and real noisy image, it is proved that the proposed denoising method is effective and robust. However, there are still some drawbacks in this method, such as time complexity. All of these experiments are operated on 4M computer and platform is Matlab. For an $800 * 600$ image, the consuming time is 25 S. It means that the time complexity of the proposed method is high, although the visual effect outweighs other methods. Therefore, we would focus on the improvement of time complexity of the proposed method and then inspire new insights for the further research in this aspect.

Conflict of Interests

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

Acknowledgments

The work is supported by Science Research Program of Hubei Provincial Science & Technology Department of China (no. 2015CFC781, no. 2014CFB612, and no. 2011cdb088), Culture & Technology Promotion Program of China (no. 201307), Program of National Natural Science Foundation of China (nos. 61263030, 61463014), Innovation Projection of Culture Ministry (no. 201217), Ph.D. Technology Program (no. MY2014B018), and Program of Sichuan Provincial Department of Education (no. 15ZB0039).

References

- [1] J. Weickert, "A review of nonlinear diffusion filtering," in *Scale-Space Theory in Computer Vision*, vol. 1252, pp. 1–28, Springer, 1997.
- [2] T. F. Chan and S. Esedoglu, "Aspects of total variation regularized L^1 function approximation," Tech. Rep. 04-07, UCLA Mathematics Department, 2004.
- [3] C. Bo and Z. Li, "Symmetric four order PDE denoising algorithm," *Computer Engineering*, vol. 34, no. 13, pp. 188–189, 2008.
- [4] D. H. Xu and R. S. Wang, "Imaging denoising based on non-local regularization," *The Research and Application of Computer*, vol. 26, no. 12, pp. 4830–4832, 2009.
- [5] M. D. Gupta and S. Kumar, "Non-convex P-norm projection for robust sparsity," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1593–1600, December 2013.
- [6] G. J. Liu and X. P. Zeng, "Map image adaptive regularization denoising," *Journal of Chongqing University*, vol. 35, no. 10, pp. 63–67, 2012.
- [7] X. W. Liu and L. H. Huang, "A new nonlocal total variation regularization algorithm for image denoising," *Mathematics and Computers in Simulation*, vol. 97, no. 3, pp. 224–233, 2014.
- [8] M. J. Chen, P. X. Yang, and J. Wang, "Adaptive image denoising model based regularization and TV fidelity," *Journal of Chongqing Post and Telecommunication*, vol. 23, no. 5, pp. 621–625, 2011.
- [9] S. Suman, "Image denoising using new adaptive based median filter," *Signal & Image Processing*, vol. 5, no. 4, pp. 1–13, 2014.
- [10] J. Yan and W. S. Lu, "Imaging denoising by generalized total variation regularization and least squares fidelity," *Multidimensional Systems and Signal Processing*, vol. 20, no. 1, pp. 89–97, 2015.
- [11] A. Anilet and C. Hati, "Image denoising method using curvelet transform and wiener filter," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 3, no. 1, pp. 6943–6950, 2014.
- [12] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganriere, and W. Wu, "Objective assessment of multiresolution fusion algorithms for context enhancement in night vision: a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 94–109, 2011.
- [13] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.

Research Article

Spectrogram Image Analysis of Error Signals for Minimizing Impulse Noise

Jeakwan Kim, Yunseon Choi, and Young-Sup Lee

Department of Embedded Systems Engineering, University of Incheon, 119 Academy-ro, Yeonsu-gu, Incheon 406-772, Republic of Korea

Correspondence should be addressed to Young-Sup Lee; ysl@incheon.ac.kr

Received 16 March 2015; Accepted 8 July 2015

Academic Editor: Marco Anisetti

Copyright © 2016 Jeakwan Kim 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.

This paper presents the theoretical and experimental study on the spectrogram image analysis of error signals for minimizing the impulse input noises in the active suppression of noise. Impulse inputs of some specific wave patterns as primary noises to a one-dimensional duct with the length of 1800 mm are shown. The convergence speed of the adaptive feedforward algorithm based on the least mean square approach was controlled by a normalized step size which was incorporated into the algorithm. The variations of the step size govern the stability as well as the convergence speed. Because of this reason, a normalized step size is introduced as a new method for the control of impulse noise. The spectrogram images which indicate the degree of the attenuation of the impulse input noises are considered to represent the attenuation with the new method. The algorithm is extensively investigated in both simulation and real-time control experiment. It is demonstrated that the suggested algorithm worked with a nice stability and performance against impulse noises. The results in this study can be used for practical active noise control systems.

1. Introduction

Active noise control (ANC) systems are usually seeking to maximize the attenuation of a primary noise by cancelling the unwanted noise by taking advantage of the principle of superposition [1, 2]. They use adaptive feedforward algorithms such as the filtered reference least mean square (FxLMS) to compensate for the effect of the secondary path in order to ensure convergence [1, 2]. This ANC has been widely applied successfully to a number of applications such as airplanes, cars, headsets, mobile devices, and other consumer electronics [3, 4].

However, as it is necessary to suppress a sudden impulse noise, an advanced ANC system, therefore, needs to have a capability to attenuate the sudden impulse noise to a certain level [5–8]. The spectrogram images give good information on how much the impulse input noise is suppressed after the application of the ANC algorithm [9–15].

Impulse noises are important but harmful sources of the input to a number of practical control systems in which the noise level needs to maintain a certain level such as passenger cars or other various vehicles. Passenger cars are frequently exposed to excessive impulse noises when they

are driven on rough roads or various bumps on pavements. Sudden impulses give huge effects to active noise control systems in terms of control stability as they can invoke excessive responses to maintain the performance during control. Therefore, it is necessary to adapt the step size to the magnitude of the impulses to control it, and the FxNLMS (filtered-x normalized LMS) algorithm can be used suitably in this case. For better driving and travelling conditions for drivers and passengers, they need to keep a comfortable noise level. The FxNLMS can offer better calm environment rather than FxLMS algorithm especially against impulse noise inputs. In this study, the algorithm for a duct ANC system is investigated in depth in order to control some impulse signals made of specific half-sine waves.

In this study, thus, the analysis of the spectrogram images of error signals before and after control against three different impulses input noises to a duct system is shown.

The rest of the paper is organized as follows. In Section 2, theoretical considerations are carried out to derive proper equations for impulse noise control based on the FxNLMS algorithm with a normalization factor in a duct. Section 3 presents the experimental setup including the test duct and control board for the real-time control. Also, the modelling

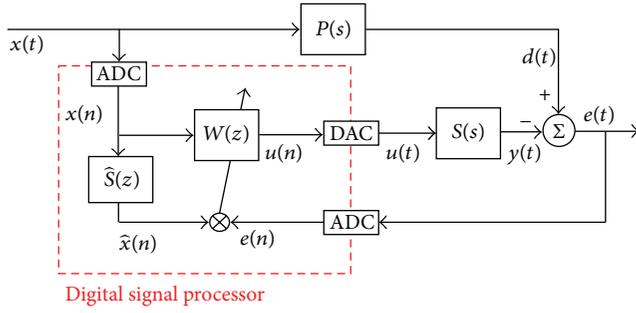


FIGURE 1: FxLMS algorithm.

of the secondary path is achieved to implement it in the control algorithm and the impulse signals for the input signal are designated in the same section. Intensive analysis and discussions from the measured results from real-time control in terms of time and frequency domain are written in Section 4. Finally, conclusions are summarized in Section 5.

2. Control Algorithm for Minimizing Impulse Noises

An impulse input to a system to be controlled causes a sudden response which makes the normal FxLMS cannot work properly. As shown in Figure 1, the FxLMS update equation of the adaptive filter $W(z)$ is expressed as

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \alpha e \hat{\mathbf{x}}, \quad (1)$$

where α is the convergence coefficient and e is the error signal and $\hat{\mathbf{x}}$ is the filtered reference signal vector which passed through the secondary path model $\hat{S}(z)$ as illustrated in Figure 1.

Also, as the large impulse input signal can introduce the instability in the LMS algorithm, one can use a smaller convergence coefficient when the measured reference signal becomes suddenly huge. This can lead to a stable FxLMS algorithm by modifying the step size of the update of the adaptive filter $W(z)$ by replacing α in (1) with α_N which can be defined by

$$\alpha_N = \frac{\alpha}{\hat{\mathbf{x}}^T(n) \hat{\mathbf{x}}(n)}. \quad (2)$$

In (2), α_N becomes smaller when the filtered reference signal $\hat{\mathbf{x}}$ is suddenly large. This is known as the FxNLMS (filtered-x normalized LMS). However, when the interval of each impulse is small enough, the system will respond continuously before the previous response decays completely. Thus, a new suggestion for the convergence coefficient can be written as

$$\alpha_{NC} = k \frac{\alpha}{\hat{\mathbf{x}}^T(n) \hat{\mathbf{x}}(n)}, \quad (3)$$

where the new term k increases the step size. Thus, α_{NC} takes the place of α in (1) and it is rewritten as

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \alpha_{NC} e \hat{\mathbf{x}}. \quad (4)$$

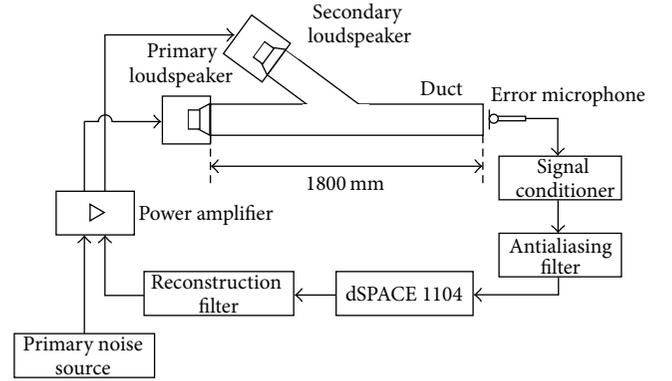


FIGURE 2: Experiment setup for the real-time control.

TABLE 1: The specifications of the components in the duct system.

Component	Specification
Duct	Acryl, 1800 mm
Loudspeakers	Sammi SR-100A50
Microphone	PCB 1/2" 377B11
Signal conditioner	PCB 442B104
Antialiasing filter, reconstruction filter	Low-pass filter, cutoff frequency = 500 Hz
Digital signal processor	dSPACE DS1104
Power amplifier	InterM QD4240

Based on the suggested control algorithm in (4), the active minimization of the impulse input noise in a duct system will be presented in Section 3 in detail. As the algorithm can adaptively follow the huge and sudden changes in the input signal as denoted $x(t)$ in Figure 1, the control experiments were performed to demonstrate the feasibility of suppressing the impulse input noises in terms of the spectrogram images.

3. Experimental Setup

3.1. Hardware Setup. For the real-time impulse noise control, a duct experimental system has been built. The system includes a dSPACE 1104 digital signal processor, an acrylic duct, a primary loudspeaker at the end, a secondary loudspeaker, an error microphone at the other side end, a power amplifier, a signal conditioner, a reconstruction filter, and an antialiasing filter as shown in Figure 2. The distance between the primary loudspeaker and the error microphone is 1800 mm.

The primary noise was originated from an external source and the control signal was generated from dSPACE 1104. The sampling frequency for the experiment was 6,000 Hz. The reconstruction filter and the antialiasing filter were designed as low-pass filters and their cutoff frequencies were both 500 Hz. Table 1 shows the specifications of the components used in the experimental duct system in detail. In the real-time control experiment, the fixed length of the signal $\hat{\mathbf{x}}$ in the denominator of (2) for the FxNLMS was 100.

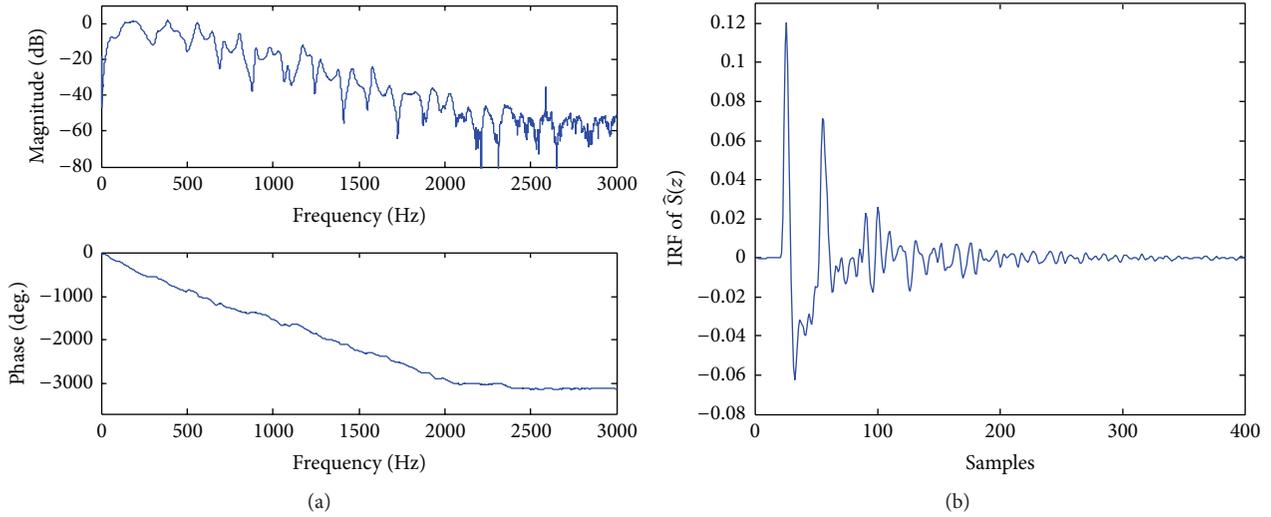


FIGURE 3: Responses of the secondary path. (a) FRF, (b) IRF.

3.2. Plant Modelling. The modeling of the secondary path of the duct system was carried out for the implementation of the control system. Its frequency response function (FRF) and impulse response function (IRF) are plotted in Figures 3(a) and 3(b), respectively.

The FRF indicates that there are many modes in the plant and the phase shows a linear property between 0 and about 2,000 Hz although there are some sudden changes at certain frequencies such as about 500 Hz and 1,100 Hz as illustrated in Figure 3(a).

The length of the IRF reaches about 400 samples until the response is decayed enough when the sampling frequency is 6,000 Hz as plotted in Figure 3(b). However, the IRF length to represent the secondary path $\hat{S}(z)$ was just 50 samples, because of the complexity of the control algorithm applied in this experiment.

3.3. Impulse Input Signals. The impulse inputs signals were designed by combining with half-sine waves and white noises to test the stability and performance of the control algorithm against some sudden impulse noises as described in Section 2. As shown in Figure 4, the three half-sine waves with the frequencies of 77 Hz, 209 Hz, and 431 Hz, respectively, were designed for the control experiment. The sampling frequency of the control experiment was 6,000 Hz, and this indicates that 39 samples were necessary to implement the 77 Hz half-sine wave in Figure 4. And 15 and 7 samples were used for 209 Hz and 431 Hz half-sine waves, respectively.

Actually, the three different patterns of the half-sine waves represent some typical impulse noises. For example, the drivers of passenger cars can experience sudden big impulse noises when they pass over bumps on various roads. There are many different bumps on roads in terms of their heights and lengths. The three half-sine waves of 77, 209, and 431 Hz indicate blunt, medium, and sharp bumps, respectively. White noises were combined in the half-sine waves to make the impulse input signals contain more realistic background disturbances.

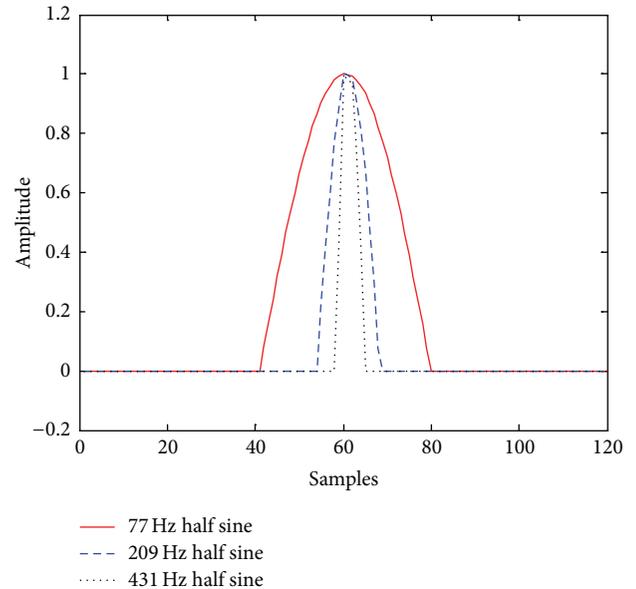


FIGURE 4: Half-sine waves of 77 Hz, 209 Hz, and 431 Hz for impulse input signals.

TABLE 2: The specifications of the components in the duct system.

Impulse input type	Signal combination
A	14 of 77 Hz half-sine waves + white noise
B	14 of 209 Hz half-sine waves + white noise
C	5 of 77 Hz + 5 of 209 Hz + 4 of 431 Hz half-sine waves + white noise

Then, three impulse inputs of types A, B, and C were considered as shown in Table 2. Type A impulse input consists of randomly spaced 14 impulses of 77 Hz half-sine wave with white noise for 2 seconds. Type B input has randomly spaced

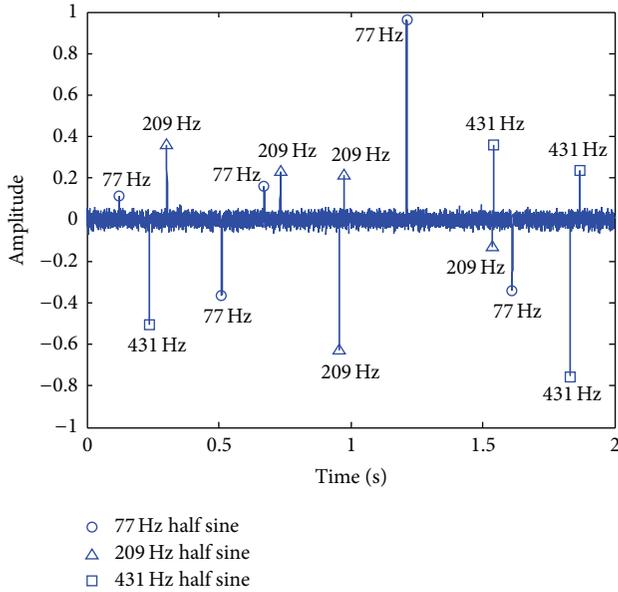


FIGURE 5: Type C impulse input with 14 half-sine waves and white noise.

14 impulses of 209 Hz half-sine wave with white noise for the same duration. Type C input contains randomly spaced 14 impulses of 77 Hz, 209 Hz, and 431 Hz half-sine waves with white noise as plotted in Figure 5.

The number of the half-sine waves was 14 for each impulse input type as illustrated in Figure 5. The magnitudes, intervals, and signs of each half-sine wave were randomly decided for the experiment. Type C impulse input consists of 5 of 77 Hz, 5 of 209 Hz, and 4 of 431 Hz half-sine waves with white noise. The largest magnitude of the 14 impulses in Figure 5 is designed to be 1. Each type impulse input was used as $x(t)$ signal in Figure 1 for the real-time control experiment in this study.

4. Results and Discussions

The real-time control experiments in the duct system against the impulse input noises as the primary sources, which are defined in Table 2 and Figure 4 in the previous section, are shown.

4.1. Control against Type A Impulse Input. As type A impulse input consists of 77 Hz half-sine waves mostly, the spectrogram and PSD plotted in Figures 6(a) and 6(d), respectively, show the dominant frequency components are observed around 80 Hz (the darkest parts) before control. It is also noted that there are some peaky components at about 350 Hz, which was caused by the acoustic modes in the duct.

The spectrogram, PSD, and attenuation in Figures 6(b), 6(d), and 6(e), respectively, after the FxLMS control against the same input type indicate that the reduction of 17~25 dB was achieved below 200 Hz. Particularly the dominant frequency components around 80 Hz were suppressed substantially. The spectrogram image after the FxLMS control

in Figure 6(b) becomes quite brighter than before control in Figure 6(a).

The spectrogram in Figure 6(c) after the FxNLMS control displays even brighter image than after FxLMS control in Figure 6(b). The attenuation plot in Figure 6(e) illustrates that the frequency range of noise suppression was extended to about 800 Hz and the amount of the reduction reached about 15~25 dB.

It is observed that the attenuation by the FxLMS is slightly better than the FxNLMS between 50 and 100 Hz in Figure 6(e). In type A input, the necessary sample number of the 77 Hz half-sine wave is 39 because the sampling frequency was 6000 Hz. As the fixed length of the signal \hat{x} in the denominator of (2) for the FxNLMS was 100 in the control experiment, the FxNLMS could become less effective when the wavelength of the input noise is relatively long.

If the length of the signal \hat{x} becomes longer, then the control effectiveness at low frequency such as 50–100 Hz would be better. However, the effectiveness at high frequency could become worse. Thus, this can be a problem of choosing more important frequency range for control. After intensive experiment, the length of 100 was determined to develop more effective algorithm against impulse input noises in which the frequencies of half-sine waves vary between 77 Hz and 431 Hz.

4.2. Control against Type B Impulse Input. As type B impulse input comprises 209 Hz half-sine waves, these frequency components were cancelled extensively after either the FxLMS or the FxNLMS controls as shown in Figure 7. It is noted that the performances of both control algorithms were quite similar below about 450 Hz with the maximum reduction of about 30 dB as it can be seen from Figures 7(d) and 7(e). The spectrogram images in Figures 7(a), 7(b), and 7(c) demonstrate that the FxNLMS algorithm was better than that of the FxLMS.

4.3. Control against Type C Impulse Input. As type C impulse input has three half-sine waves of 77 Hz, 209 Hz, and 431 Hz randomly in terms of their magnitudes and occurrence and white noise, because of the various magnitudes of the input impulses which cause the different ringing times after each impulse, as plotted in Figure 8(b), the FxLMS control cannot quickly suppress the ringing response followed by the impulse at about 1.2 seconds. It is found that the ringing response continued until the next impulse occurred.

However, the FxNLMS control suppressed the ringing response very quickly as presented in Figure 8(c). This difference of the performances of both control strategies demonstrates that the importance of the FxNLMS approach is very useful against a severe impulse input in those systems like passenger cars. In addition, it is noted that the control performance of the FxNLMS is more efficient against low frequency such as 77 Hz impulse input signals. The overall attenuation difference between the two algorithms in Figure 8(e) is about 5~10 dB at the frequency range of about 70~200 Hz. The spectrogram images in this result clarify the performance differences in the suppression of the mixed impulse input (type C).

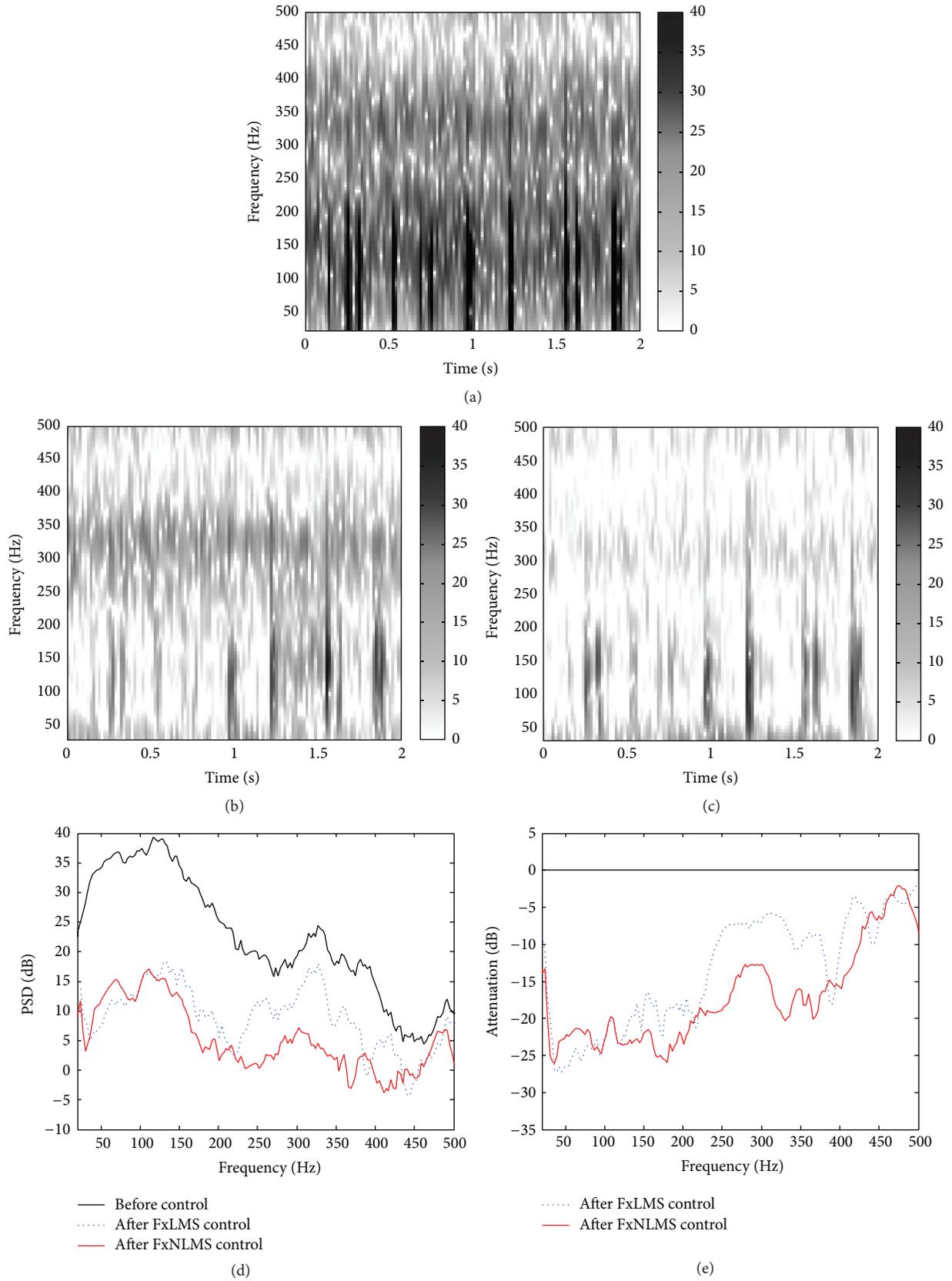


FIGURE 6: Real-time control result against type A impulse input. (a) Before control, (b) after FxLMS control, (c) after FxNLMS control, (d) PSD, and (e) attenuation.

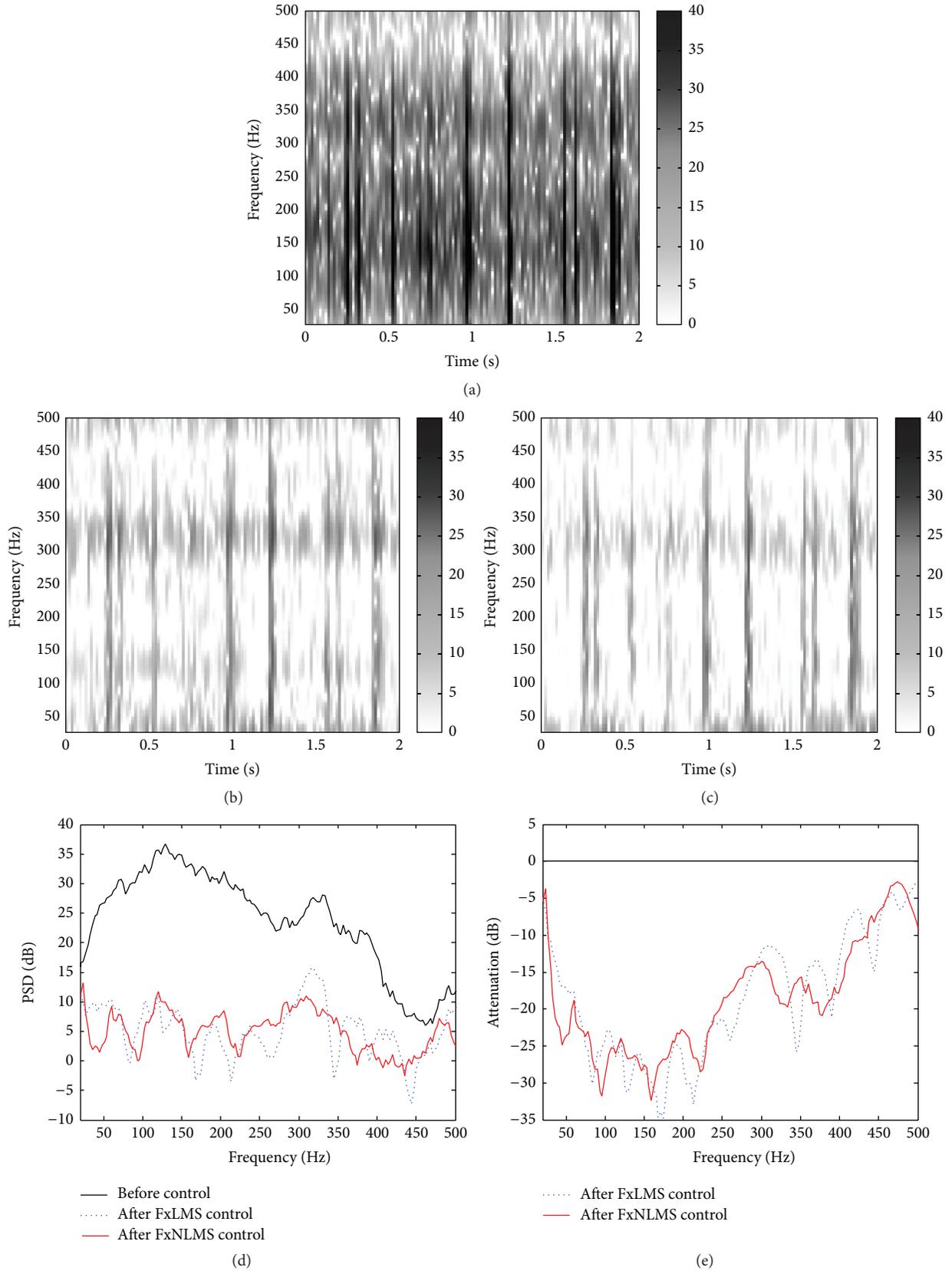


FIGURE 7: Real-time control result against type B impulse input. (a) Before control, (b) after FxLMS control, (c) after FxNLMS control, (d) PSD, and (e) attenuation.

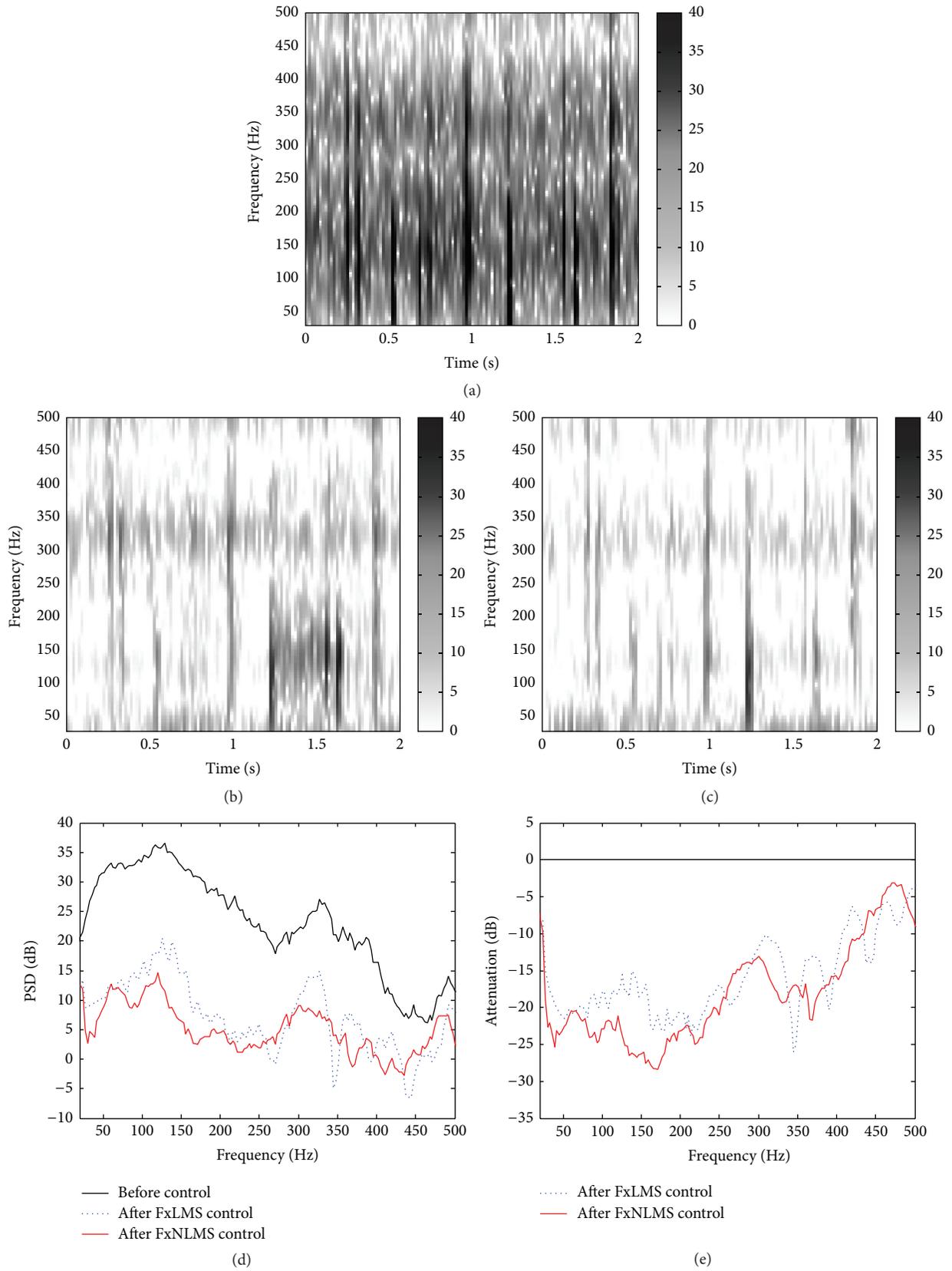


FIGURE 8: Real-time control result against type C impulse input. (a) Before control, (b) after FxLMS control, (c) after FxNLMS control, (d) PSD, and (e) attenuation.

TABLE 3: Overall attenuation in 20~500 Hz.

Impulse input type	Overall attenuation	
	FxLMS	FxNLMS
Type A	-13.44 dB	-16.97 dB
Type B	-18.03 dB	-18.24 dB
Type C	-15.77 dB	-17.80 dB

Table 3 presents the average attenuations in dB against the three impulse input types A, B, or C. As plotted in Figures 6, 7, and 8, the differences in the three spectrogram images are caused by the impulse inputs that are different. The FxNLMS algorithm embedded in the real-time digital controller responds automatically with adaptive algorithms of the FxLMS and the FxNLMS against the input noises. Thus, the spectrogram images must be different because the two algorithms are different. Since the inputs were impulses, the FxNLMS showed better performance than the FxLMS in most of the frequency range as summarized in Table 3. The FxNLMS algorithm performed -16.97 dB, -18.24 dB, and -17.80 dB against the impulse input types A, B, and C, respectively. By the way, the FxLMS algorithm accomplished -13.44 dB, -18.03 dB, and -15.77 dB in the same order. This indicates that the FxNLMS algorithm achieved about 3.5 dB, 0.2 dB, and 2.0 dB more attenuation than the FxLMS.

Comparing the performances of the two algorithms against especially type C impulses input which is expected as more practical, it is noted the FxNLMS provides better stability for suppressing the resonances of the duct system. The two largest resonances at about 125 Hz and 325 Hz before control as plotted in Figure 8(d) were suppressed more about 7~15 dB by the FxNLMS than the FxLMS.

The frequency variation of the impulse input gives important effect to the stability and the performance of the FxNLMS algorithm, since the length of the signal \hat{x} in the denominator of (2) for the FxNLMS is very much related to the input half-sine wave frequency. Thus, if it is necessary to minimize a higher frequency impulse input, then a shorter length of the signal \hat{x} is required. Against low frequency impulses noises, a longer length is better in general. If the variation of the frequency of a half-sine impulse is small, the effect will not be big.

5. Conclusions

This study presents spectrogram images analysis of error signals in an active control system against the impulse input signals in a one-dimensional duct. For the improvement of spectrogram images, a new normalization factor α_{NC} in the update equation was implemented for the real-time control algorithm.

The frequency variation of the impulse input gives important effect to the stability and the performance of the FxNLMS algorithm, since the length of the signal \hat{x} in the denominator for the FxNLMS is very much related to the input half-sine wave frequency.

Control results with the new algorithm showed stable and excellent responses compared to those before control.

As the three impulse input signals types with white noise were generated independently, the control algorithm attenuates the noises from about -17 dB to -18 dB with the new normalization factor in the real-time control. On the other hand, only about -13 dB to -16 dB were reduced with the FxLMS algorithm. The improved algorithm also worked with a nice stability and performance against 3 different types of the impulse input signals. The results in this study can be used for practical active noise control systems.

Conflict of Interests

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

Acknowledgment

This research was supported by a grant (code 15CCTI-B062410-04-000000) from Advanced Water Management Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

- [1] G. Jeon, S. Kang, and Y.-S. Lee, "Noise level estimation for image processing," in *Convergence and Hybrid Information Technology*, vol. 7425 of *Lecture Notes in Computer Science*, pp. 350-356, 2012.
- [2] Y.-S. Lee, P. Gardonio, and S. J. Elliott, "Coupling analysis of a matched piezoelectric sensor and actuator pair for vibration control of a smart beam," *Journal of the Acoustical Society of America*, vol. 111, no. 6, pp. 2715-2726, 2002.
- [3] G. Jeon, M. Anisetti, V. Bellandi, E. Damiani, and J. Jeong, "Fuzzy weighted approach to improve visual quality of edge-based filtering," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1661-1667, 2007.
- [4] P. Gardonio, Y.-S. Lee, S. J. Elliott, and S. Debois, "Analysis and measurement of a matched volume velocity sensor and uniform force actuator for active structural acoustic control," *The Journal of the Acoustical Society of America*, vol. 110, no. 6, pp. 3025-3031, 2001.
- [5] D. Bismor, "LMS algorithm step size adjustment for fast convergence," *Archives of Acoustics*, vol. 37, no. 1, pp. 31-40, 2012.
- [6] S. O. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 5th edition, 2013.
- [7] S. J. Elliott, *Signal Processing for Active Control*, Academic Press, 2001.
- [8] Y.-S. Lee and S. J. Elliott, "Active position control of a flexible smart beam using internal model control," *Journal of Sound and Vibration*, vol. 242, no. 5, pp. 767-791, 2001.
- [9] K. Shin, S. Pyo, and Y.-S. Lee, "Simple feed-forward active control method for suppressing the shock response of a flexible cantilever beam," *Smart Materials and Structures*, vol. 18, no. 2, Article ID 027001, 6 pages, 2009.
- [10] G. Jeon, S. Kang, and Y.-S. Lee, "A filter-based format conversion approach," *Communications in Computer and Information Science*, vol. 310, pp. 559-565, 2012.
- [11] G. Jeon, S. Kang, and Y.-S. Lee, "Weight assignment method for interpolation," *Communications in Computer and Information Service*, vol. 342, pp. 272-278, 2012.

- [12] R. M. Reddy, I. M. S. Panahi, and R. Briggs, "Hybrid FxRLS-FxNLMS adaptive algorithm for active noise control in fMRI application," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 2, pp. 474–480, 2011.
- [13] K. Sivadasan and N. K. Narayanan, "ANC system for noisy speech," *Signal and Image Processing International Journal*, vol. 3, no. 3, pp. 93–101, 2012.
- [14] K. K. Das and J. K. Satapathy, "Frequency-domain block filtered-x NLMS algorithm for multichannel ANC," in *Proceedings of the 1st International Conference on Emerging Trends in Engineering and Technology (ICETET '08)*, pp. 1293–1297, IEEE, Nagpur, India, July 2008.
- [15] Y. Kajikawa and R. Hirayama, "Feedback active noise control system combining linear prediction filter," in *Proceedings of the European Signal Processing Conference*, pp. 31–35, August 2010.

Research Article

Multiscale and Multitopic Sparse Representation for Multisensor Infrared Image Superresolution

Xiaomin Yang,¹ Kai Liu,² Zhongliang Gan,³ and Binyu Yan¹

¹College of Electronics and Information Engineering, University of Sichuan, Chengdu, Sichuan 610064, China

²College of Electrical and Engineering Information, University of Sichuan, Chengdu, Sichuan 610064, China

³College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Correspondence should be addressed to Kai Liu; kailiu@scu.edu.cn

Received 17 March 2015; Revised 4 June 2015; Accepted 9 June 2015

Academic Editor: Marco Anisetti

Copyright © 2016 Xiaomin Yang 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.

Methods based on sparse coding have been successfully used in single-image superresolution (SR) reconstruction. However, the traditional sparse representation-based SR image reconstruction for infrared (IR) images usually suffers from three problems. First, IR images always lack detailed information. Second, a traditional sparse dictionary is learned from patches with a fixed size, which may not capture the exact information of the images and may ignore the fact that images naturally come at different scales in many cases. Finally, traditional sparse dictionary learning methods aim at learning a universal and overcomplete dictionary. However, many different local structural patterns exist. One dictionary is inadequate in capturing all of the different structures. We propose a novel IR image SR method to overcome these problems. First, we combine the information from multisensors to improve the resolution of the IR image. Then, we use multiscale patches to represent the image in a more efficient manner. Finally, we partition the natural images into documents and group such documents to determine the inherent topics and to learn the sparse dictionary of each topic. Extensive experiments validate that using the proposed method yields better results in terms of quantitation and visual perception than many state-of-the-art algorithms.

1. Introduction

High-resolution (HR) infrared (IR) images are desired in various electronic imaging applications, such as medical diagnosis, criminal investigation, surveillance, remote sensing, and aerospace. However, given the inherent limitation of relevant imaging devices or other factors, obtaining images at a desired resolution is difficult. Therefore, many efforts have been devoted to improving the spatial resolution of the IR image. Superresolution (SR) is one of the most promising methods in the research community.

At present, a large number of SR methods have been developed successfully. The existing methods for image SR can be divided into three general categories: interpolation-based methods [1, 2], reconstruction-based methods [3–6], and learning-based methods [7–11].

The interpolation-based [1, 2] scheme applies the correlation of neighboring image pixels to approximate the fundamental HR pixels. These types of methods can be easily implemented at a high speed. However, these methods may lead to the loss of detailed information.

Reconstruction-based approaches utilize additional information from low-resolution (LR) images to synthesize an HR image. These approaches are ill-posed estimation problems and require a priori information on images to regularize the solution. Therefore, various regularization methods have been proposed to improve the performance of SR reconstruction, such as the projection on convex sets [3], maximum a posteriori (MAP) [4, 5], and regularization-based method [6]. Compared with interpolation-based schemes, the reconstruction-based methods deliver better performance with a small desired magnification factor. However,

the most common defect of multiframe SR reconstruction is that, with an increase in the magnification factor, the LR inputs cannot provide sufficient information to maintain a high-quality SR reconstruction result.

Learning-based methods presume that the high-frequency details lost in the LR image can be predicted by learning the cooccurrence relationship between LR training patches and their corresponding HR patches. Freeman et al. [7] first introduced the learning idea for SR reconstruction, which uses a Markov random field model to learn the relationship between local regions of images and their underlying scenes. Various effective tools have been proposed to learn prior information, such as neighbor embedding- (NE-) based methods [8, 12], regression-based methods [9, 10], and sparse coding- (SC-) based methods [11, 13–15]. The NE-based methods estimate each desired HR image patch by linearly combining its neighbor training HR image patches. Chang et al. [12] introduced locally linear embedding from manifold learning to process the image SR task. Zhang et al. [8] proposed a partially supervised NE method. However, given the lack of prior textures and details, NE-based methods are weak in visualizing textures and details. The regression-based methods directly estimate the desired HR pixels using some complicated statistical models. Wang and Tang [9] proposed a principal component analysis-based SR reconstruction method to estimate the desired HR image. Wu et al. [10] used the kernel partial least squares regression model to handle the one-to-many mapping problem. Wu's method requires searching the neighbors in the entire training database and using the same number of principal components to synthesize the desired HR feature patches, which result in high computational costs. The SC-based SR method can better retain the most relevant reconstruction neighbors and can restore more image information than the two learning-based methods discussed above. Yang et al. [16] proposed an approach based on sparse representation, with the assumption that the HR and LR images share the same set of sparse coefficients. Therefore, the HR image can be reconstructed by combining the trained HR dictionary and the sparse coefficients of the corresponding LR image.

The abovementioned SC-based SR methods always suffer from three problems. First, due to the inherent limitation of relevant imaging devices or other factors, IR images always lack detailed information, which leads to unsatisfied IR image reconstruction results. Multiple images acquired by different sensors provide complementary information on the same scene. As such, a reasonable method of improving the resolution of the IR image is the combination of inherently complementary information from the images obtained from different multisensors. Second, a traditional sparse dictionary is learned from patches with a fixed size, which cannot capture the exact information of the images. However, the local structures of an image tend to repeat themselves many times with some similar neighbors across the natural images, not only within the same scale but also across different scales. Details missing in a local structure at a smaller scale can be estimated from similar patches at a larger scale. Different images prefer different patch sizes for optimal representation. Therefore, jointly representing an image at different scales is

important. Considering the above cues, we propose a model of obtaining multiscale patches to learn dictionaries. We use a simple model that generates pyramid images and divides such images into multiscale patches. Finally, given that dictionary learning is a key issue of the sparse representation model, considerable effort in learning dictionaries from example image patches has been exerted, leading to state-of-the-art results in image reconstruction. Many dictionary learning methods aim at learning a universal and overcomplete dictionary that represents various image structures. However, for natural images, a large number of different local structural patterns exist. The contents can vary significantly across different images or different patches in a single image. One dictionary is inadequate in capturing all of the different structures. Multiple dictionaries [15, 17] are more effective in representing various contents in an image and provide better reconstruction results than one universal dictionary [15]. Based on these observations, training patches are categorized into multiple groups based on visual characteristics in our algorithm. A subdictionary is then learned in the respective data groups. Unsuitable training sample groups used in dictionary learning lead to artifacts in example learning-based methods [18]. In this study, we group the patches into several categories. Each category corresponds to a topic. We apply the probabilistic latent semantic analysis (pLSA) model [19] to group the patches and to determine the inherent topics. That means we group the patches into several categories. Each category corresponds to a topic. We then learn the sparse dictionary for each topic. Our framework treats each group individually, thereby leading to more accurate distribution dictionaries. We conduct semantic analysis on a given patch to categorize it to a topic. The given patch can be better represented by the selected topic subdictionary. Thus, the entire image can be more accurately reconstructed using this method than using a universal dictionary, as validated by our experiments.

In summary, this study makes the following three main contributions: (1) IR images always lack detailed information. Meanwhile, VI images contain abundant object edges and details, providing a more perceptual description of a scene for human eyes. This study combines the inherently complementary information from images obtained from different multisensors to improve the resolution of the IR image. (2) To learn the sparse dictionary for representing similar redundancies of local patterns within the same scale and across different scales, this study builds pyramid images downsampled from the images. Then it divides the pyramid images into multiscale patches, thereby representing the image in a more efficient manner and providing a more global look of the image. (3) The pLSA model is applied to group the patches by determining the inherent topics and to group the training patches with similar patterns. Each dictionary is learned from some type of example patches with the same topic, and multiple dictionaries are learned simultaneously. Extensive experimental results show that our proposed method achieves competitive performance compared to state-of-the-art methods.

The remainder of this paper is organized as follows: Section 2 presents the details of the proposed approach. Section 3 reports the experimental results. Section 4 discusses the conclusion.

2. The Proposed SR Scheme

This study proposes a novel sparse representation algorithm, which aims to combine the information of visible images, provide a more global look of the IR image, and simultaneously utilize the inherent topics of IR images in a unified framework. The proposed method can be divided into three steps: (a) combining the information of images from multisensors, (b) obtaining multiscale patches, and (c) learning multitopic sparse dictionaries. In combining the information of visible images, our framework improves the resolution of the IR image when learning the LR sparse dictionary. In obtaining multiscale patches, we build pyramid images and extract multiscale patches from such images, which can provide a more global look of the images. In presenting different structural patterns more accurately, we partition the natural images into documents and group them to determine the inherent topics using the pLSA. A compact subdictionary can then be learned for each topic.

2.1. Combining the Information of Multisensors. Given an observed LR IR image Y , which is a downsampled and blurred version of the HR image X of the same scene, we derive the following equation:

$$Y = HBX, \quad (1)$$

where H denotes a downsampling operator and B is a blurring filter. The goal of a single-image SR is to reconstruct the HR image \widehat{X} from the LR image Y as accurately as possible.

With the LR image Y , y^i is the set of patch features extracted from Y :

$$y^i = \text{Extract}(Y), \quad (2)$$

where $\text{Extract}(\cdot)$ is an operator that extracts the feature of patch i from image Y .

Recent works [16, 20] indicate that derivative features can represent the patch more efficiently than the actual intensities. The derivative features are obtained using four 1D filters:

$$\begin{aligned} f_1 &= [-1, 0, 1], \\ f_2 &= f_1^T, \\ f_3 &= [1, 0, -2, 0, 1], \\ f_4 &= f_3^T. \end{aligned} \quad (3)$$

Images acquired by multisensors provide complementary information on the same scene. IR images always lack detailed information. Meanwhile, VI images contain abundant object edges and details, providing a more perceptual description of a scene for human eyes [21]. As such, combining the detailed information in visible images to improve the resolution of the IR image is reasonable; that is, the information of an LR IR image and the information of the corresponding HR visible image are used for reconstructing an HR IR image.

Applying these four filters, we obtain four description feature vectors for each patch of the LR IR image and its corresponding HR visible image, which are concatenated as one vector in the final gradient representation of the LR patch. The information of the LR IR image and the information of its corresponding HR visible image are combined together to learn the LR sparse dictionary.

With the sparse generative model, each patch feature y^i ($y^i \in \mathfrak{R}^n$) can be projected over the LR dictionary $D_l \in \mathfrak{R}^{n \times k}$, which characterizes the LR patches. This projection produces a sparse representation of y^i via α_l , expressed as follows:

$$y^i = D_l \alpha_l \quad y^i \in \mathfrak{R}^n, \quad \|\alpha_l\|_0 \ll k, \quad (4)$$

where α_l denotes sparse representation atoms. For the HR IR image, high-frequency information is obtained to present the HR patch. The corresponding HR patch feature x^i has $x^i \in \mathfrak{R}^m$ sets of patch features extracted from the HR image X obtained as follows:

$$x^i = \text{Extract}(X). \quad (5)$$

Reapplying the sparse generative model, we have

$$x^i = D_h \alpha_h, \quad (6)$$

where $D_h \in \mathfrak{R}^{m \times k}$ is the HR dictionary that characterizes the HR patches and is coupled with D_l through the relation $D_l = HBD_h$. This relation indicates that each atom in D_h has its corresponding LR version in D_l and vice versa. We assume that the sparse representation of an LR patch in terms of D_l can be directly used to recover the corresponding HR patch from D_h ; namely, $\alpha_l = \alpha_h$. The process of Sparse representation-based SR by combining the information of visible images is described in Figure 1.

As such, the reconstructed HR image \widehat{X} can be built by applying the sparse representation in each y^i and then using the estimated α_l with D_h to obtain each \widehat{x}^i , which together form the image \widehat{X} .

The SC is clearly a bridge between the LR and HR patches. The dictionaries D_l and D_h have a key role in generating such SC. The dictionaries D_l and D_h can be easily generated from a set of samples using algorithms, such as K-SVD [11] and efficient SC [13, 14, 17, 22].

2.2. Obtaining Multiscale Patches. It is observed that different images prefer different patch sizes for optimal performance [13]. Reference [13] even observed the oversmoothing of artifacts when using unsuitable patches. An explanation for this phenomenon is that dictionary learning from patches with a fixed size cannot capture the exact information of the images. One size of the sample patches corresponds to one scale. However, selecting the exact patch size of the image is difficult. As such, having a multiscale dictionary avoids selecting the patch size in advance. A multiscale treatment can help represent the image in a more efficient manner. In our proposed multiscale framework, we focus on simultaneously obtaining the multiscale patches. First, pyramid

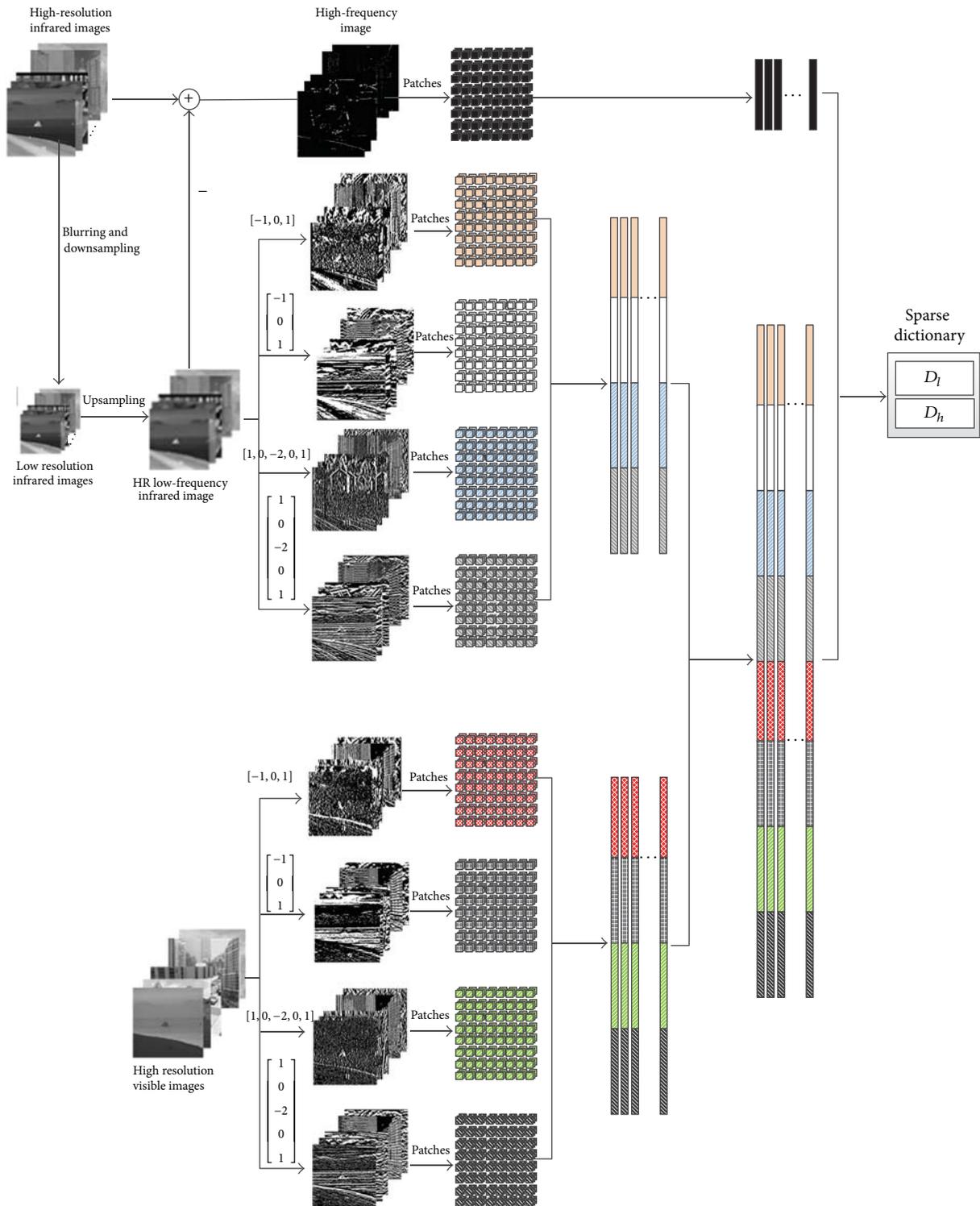


FIGURE 1: Sparse representation-based SR by combining the information of visible images.

images downsampled from the images are built to learn the sparse dictionary for representing similar redundancies of local patterns within the same scale and across different scales. Second, multiscale patches from the pyramid images are then extracted.

Pyramid transform is an effective multiresolution analysis approach. During pyramid transform, each pixel in the low spatial pyramid is obtained by downsampling from its adjacent low-pass filtered HR image. Sequential pyramid images are constructed, as shown in Figure 2. Pyramid images

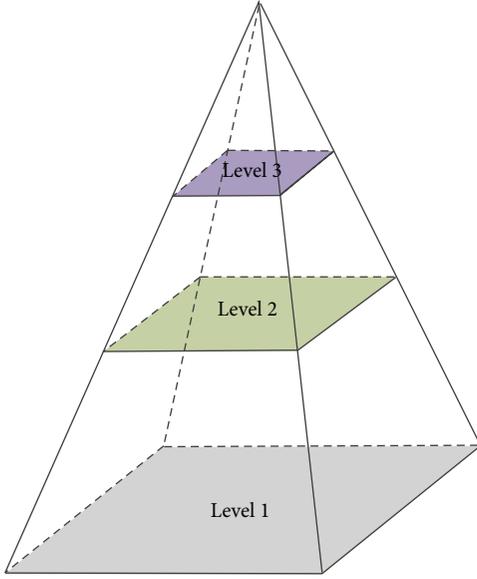


FIGURE 2: A three-level spatial pyramid.

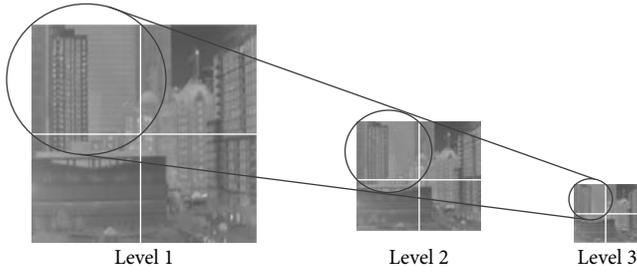


FIGURE 3: Gaussian pyramid images. The original image (level 1) and its Gaussian pyramids are shown from left to right.

can be generated by Gaussian smooth filtering, as shown in Figure 3.

Let $I(x, y)$ denote the original image. The downsampled version $I_p(x, y)$ at the p th level is obtained by convoluting $I(x, y)$ with a Gaussian kernel $G(x, y, p\sigma)$, as follows:

$$G(x, y, p\sigma) = \frac{1}{2\pi p^2 \sigma^2} \exp\left(-\frac{x^2 + y^2}{2p^2 \sigma^2}\right), \quad (7)$$

$$I_p(x, y) = I(x, y) \times G(x, y, p\sigma) \downarrow s_p,$$

where $\downarrow s_p$ denotes the downsampling operator, with the factor $s_p = (0.5)^p$ at the p th level.

After generating the pyramid images, we use the quadtree model [15] to extract multiscale patches from the pyramid images, as shown in Figure 4. We consider a set of large root patches of size $\sqrt{N} \times \sqrt{N}$ extracted from the sequential pyramid images. The root patch is then divided into subpatches of size $N_s = N/4^s$ ($s = 0, \dots, S-1$) along the tree, where S is the depth of the tree. After obtaining multiscale patches, we can learn dictionaries from the patches of different scales. Figure 5 illustrates the process of extracting multiscale patches from the pyramid images.

2.3. Learning the Multitopic Dictionary. We partition the natural images into documents and group them to determine the inherent topics using pLSA and to present the different structural patterns more accurately. Each dictionary is learned from some type of example patches with the same topic, and multiple dictionaries are learned simultaneously. The example image patches are classified into many topics by the pLSA model. Given that each topic consists of many patches with similar patterns, a compact subdictionary can be learned for each topic. For an image patch to be coded, the best subdictionary that is most relevant to the given patch is selected. Considering that the given patch can be better represented by the selected subdictionary, the entire image can be more accurately reconstructed than when a universal dictionary is used, as validated by our experiments. The use of multitopic dictionary learning has two main advantages: (1) the training patches are divided into some topics, which ensure that the subdictionary represents the statistical model of the example patches more accurately and (2) the training patches enhance the speed of dictionary learning on each topic and the final reconstruction accuracy through the transfer of knowledge between topics.

2.3.1. Standard pLSA. The pLSA [19], which is an extension of LSA [23], provides a probabilistic formulation to model documents in a text collection. The pLSA assumes that the words are generated from a mixture of latent aspects, which can be decomposed from a document. The pLSA model has been used successfully in image classification, image retrieval, and image annotation. The pLSA model ignores the orders of words in a document and instead uses the counts of words occurring in a document. We briefly outline the principle of the pLSA in this subsection. More details can be found in [19].

A corpus that contains M documents is denoted by $D = \{d_1, \dots, d_M\}$, and each document d_i is represented with the count of its words from a vocabulary $W = \{w_1, \dots, w_N\}$. The entire corpus is summarized by the $N \times M$ cooccurrence matrix N , where each entry $n(d_i, w_j)$ indicates the count of the word w_j in the document d_i . In the framework of the pLSA, the observed word W is conditionally independent of the document D given a latent variable $Z = \{z_1, \dots, z_K\}$, which is referred to as the ‘‘latent aspect.’’ The graphical model shown in Figure 6(a) illustrates the form of the joint probability of $P(w_j, d_i, z_k) = P(d_i)P(z_k | d_i)P(w_j | z_k)$ in the pLSA model. The joint probability of the observed variables is obtained by marginalizing over the latent aspect z_k :

$$P(d_i, w_j) = P(d_i) \sum_{k=1}^K P(z_k | d_i) P(w_j | z_k). \quad (8)$$

Equation (6) expresses each document as a convex combination of K aspect vectors, which results in matrix decomposition, as shown in Figure 6(b). Each document is essentially modeled as a mixture of aspects, the histogram for a particular document being composed of a mixture of the histograms corresponding to each aspect.

The model parameters of pLSA are the two conditional distributions $P(z_k | d_i)$ and $P(w_j | z_k)$, which are estimated

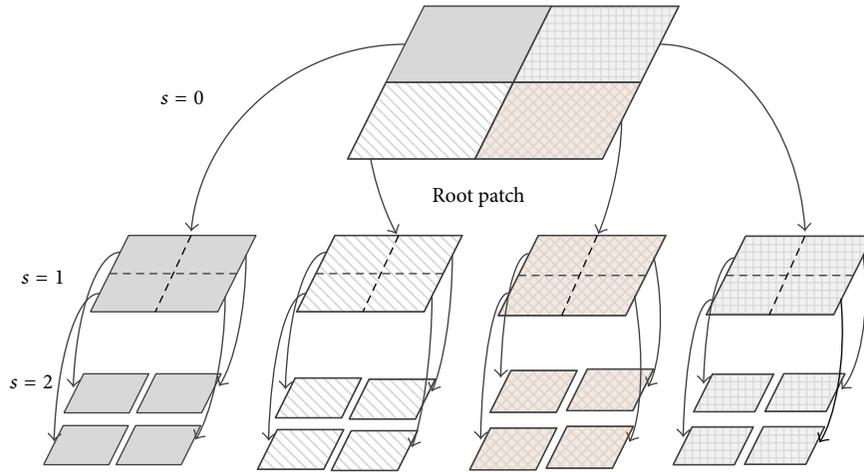


FIGURE 4: Multiscale patches based on the quadtree model.

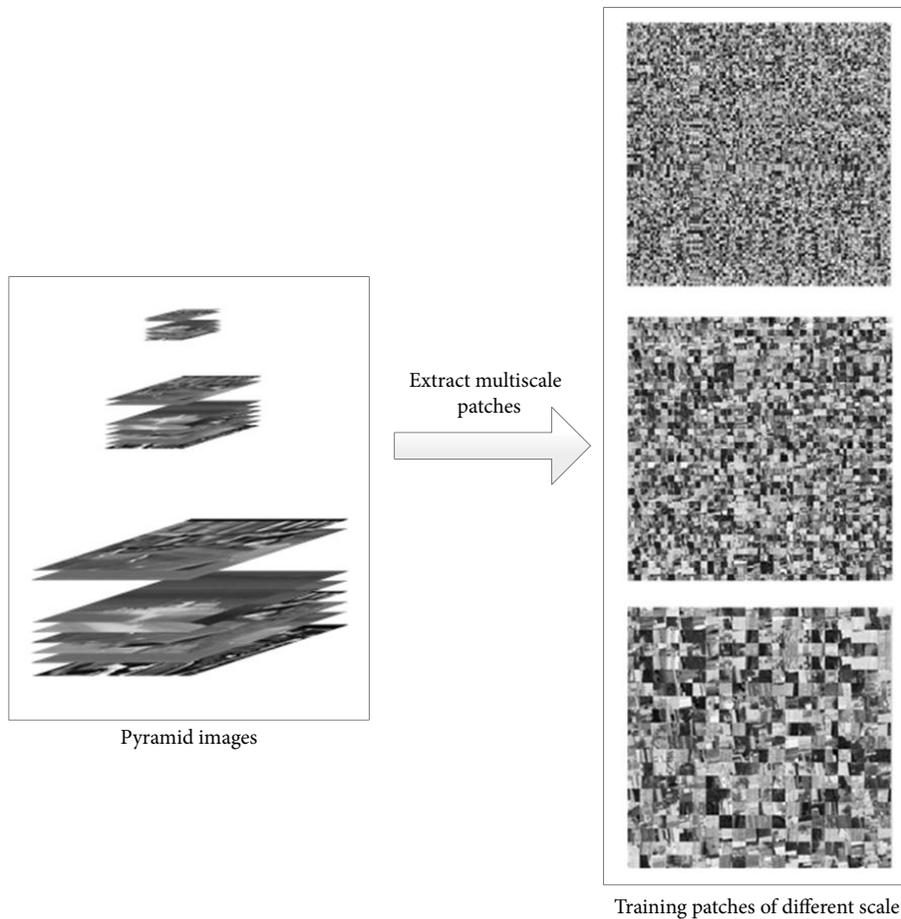


FIGURE 5: Extraction of multiscale patches.

using the expectation-maximization (EM) algorithm on a set of training documents. $P(w_j | z_k)$ characterizes each aspect and remains valid for documents out of the training set. By contrast, $P(z_k | d_i)$ is relative only to the specific documents and cannot carry any prior information to an unseen document.

The EM algorithm is used to compute the parameters $P(w_j | z_k)$ and $P(z_k | d_i)$ by maximizing the log-likelihood of the observed data:

$$L = \sum_{i=1}^N \sum_{j=1}^M n(d_i, w_j) \log P(d_i, w_j). \quad (9)$$

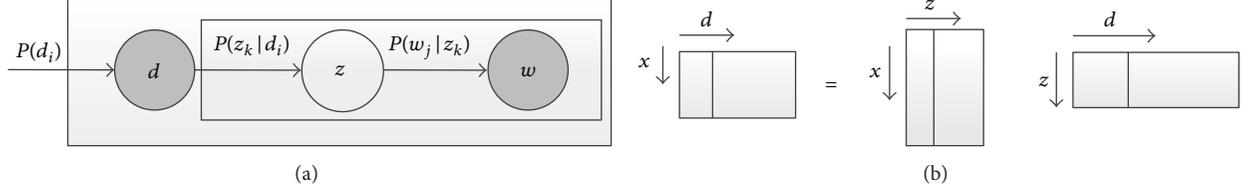


FIGURE 6: (a) Graphical model representation of pLSA; (b) matrix decomposition of conditional distribution.

The steps of the EM algorithm are described as follows:

(1) E-step: the conditional distribution $P(z_k | d_i, w_j)$ is computed from the previous estimate of the parameters:

$$P(z_k | d_i, w_j) = \frac{P(z_k | d_i) P(w_j | z_k)}{\sum_{l=1}^K P(z_l | d_i) P(w_j | z_l)}. \quad (10)$$

(2) M-step: the parameters $P(w_j | z_k)$ and $P(z_k | d_i)$ are updated with the new expected value $P(z_k | d_i, w_j)$:

$$P(w_j | z_k) = \frac{\sum_{i=1}^N n(d_i, w_j) P(z_k | d_i, w_j)}{\sum_{m=1}^M \sum_{i=1}^N n(d_i, w_j) P(z_k | d_i, w_m)}, \quad (11)$$

$$P(z_k | d_i) = \frac{\sum_{j=1}^M n(d_i, w_j) P(z_k | d_i, w_j)}{\sum_{j=1}^M n(d_i, w_j)}.$$

2.3.2. Our Method. Given a collection of IR images, we intend to determine the inherent topics of the images. We use general terms [24], such as topics, documents, and words, which are mostly used in the text of the literature. In our application, we define the atoms of the sparse dictionary as the “words” of the vocabulary and the sliding window of the sparse dictionary as the “document.” The sliding window consists of patches. Figure 7 shows the $L \times L$ sliding window (large blue square) and one patch (small red square) in it. All of the documents are grouped by “topic” based on the cooccurrences of different words within and across the documents. Our method has the following five steps: (1) vocabulary formulation, (2) document representation, (3) topic learning, (4) subdictionary construction, and (5) superresolution image reconstruction (SRIR). Our method is illustrated in Figure 8.

Vocabulary Formulation. We need to represent each document by a collection of words from a vocabulary. A general sparse dictionary D_G with M atoms α_m ($m = 1, 2, \dots, M$) is learned over all of the patches to construct the vocabulary. Each atom in D_G is defined as a word of the vocabulary. All of the atoms of D_G produce the vocabulary for the pLSA model.

Document Representation. We assume that document d_i has L patches p_l ($l = 1, 2, \dots, L$). We represent each patch in the document using a linear combination of atom α_m ($m = 1, 2, \dots, M$) from the general dictionary. We denote the atoms representing patch p_l as $\alpha_{l,m}$. We denote the count of



FIGURE 7: Sliding window and patches.

vocabulary α_j in document d_i as $n(w_j, d_i)$, where $n(w_j, d_i) = \sum_{l=1}^L \alpha_{l,j}$. We then use the pLSA model to learn the latent topic of the documents.

Topic Learning. All of the documents can be summarized by the $M \times N$ cooccurrence matrix, where each entry $n(d_i, w_j)$ indicates the count of the word α_j in document d_i and N is the total number of documents. The EM algorithm is used to compute the parameters $P(w_j | z_k)$ and $P(z_k | d_i)$ by maximizing the log-likelihood of the observed data. After learning, $P(z_k | d_i)$ represents the mixture proportions of each document. The maximum value for each of the document can be assumed as the document topic assignment.

Subdictionary Construction. We assume K determined topics. All of the documents are then classified into K group Documents = [documents₁, documents₂, ..., documents_K]. For one document group documents_i, we collect all of the patches that belong to these documents and denote these patches as patches_i. As such, we can obtain K group Patches = [patches₁, patches₂, ..., patches_K]. We aim to learn K compact subdictionaries D_k ($k = 1, 2, \dots, K$) from Patches. Each of the patches_k ($k = 1, 2, \dots, K$) is apparently expected to have the same distinctive patterns. We use the SRIR for each group's patches_k to learn the subdictionary for each topic, such that the most suitable subdictionary for each given local image patch can be selected using the pLSA model.

SRIR. We divide the LR image into overlapping documents and the documents into overlapping patches. Then, we represent each document in the same manner as that conducted

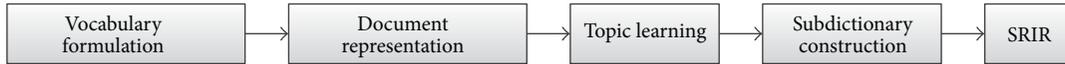


FIGURE 8: Illustration of our method.

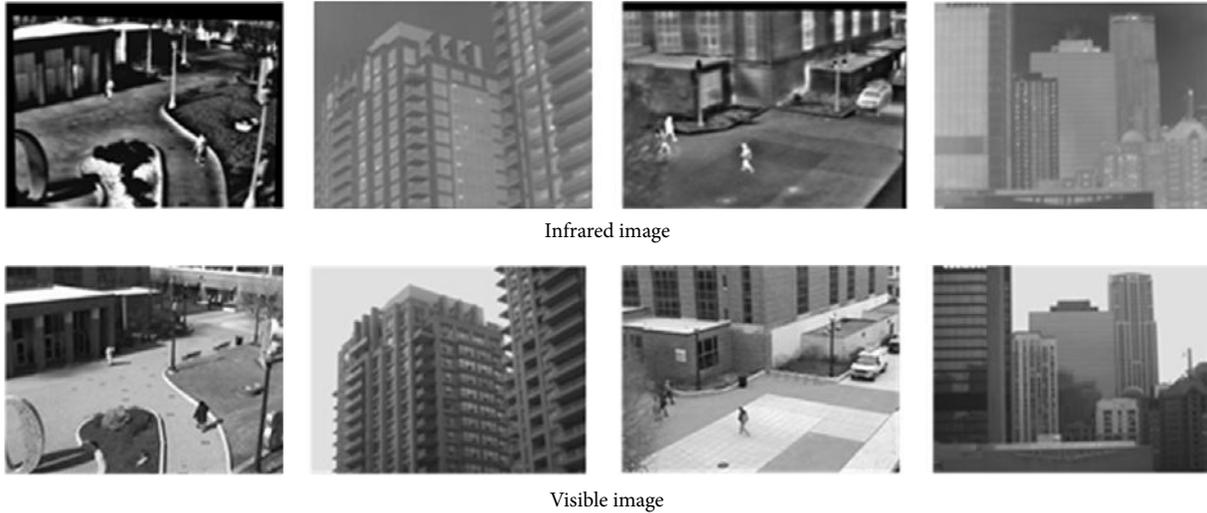


FIGURE 9: Some examples of the infrared images and visible images used in our experiments.

during topic discovery. Each document is analyzed by using the EM algorithm to determine its topic assignment. Each patch of a document is reconstructed by using the topic corresponding to the subdictionary. We do this for all of the documents in the test image and then take the average of all overlapping portions to obtain the reconstructed HR image.

3. Experimental Results

3.1. Samples and Settings. In our experiments, the IR images and corresponding visible images were obtained from [25] <http://www.dgp.toronto.edu/~nmorris/data/IRData/>. Samples of the training images are shown in Figure 9. The LR images used in all the experiments were downsampled from the HR images. In our experiments, the LR images were generated by shrinking the corresponding HR images with the scale factor of 3.

We employed the peak signal-to-noise ratio (PSNR) and the structural similarity measurement (SSIM) to evaluate the superresolved image and assess the performance of the proposed method. The mean values of the PSNR and SSIM of all of the test images were used as the quality index. The PSNR evaluates the reconstruction quality based on the pixel intensity. The SSIM measures the similarity between two images based on their structural information. The SSIM metric needs a “perfect” reference image for comparison and provides a normalized value between $[0, 1]$, where “0” indicates that the two images are totally different, whereas “1” confirms that the two images are the same. Thus, higher values of PSNR and SSIM indicate a result with better quality.

3.2. Reconstruction Results. In this section, we conduct several experiments to evaluate the effectiveness of the proposed method.

Experiment 1 (comparison with the state-of-the-art algorithms). The proposed method was tested using some IR images to validate the effectiveness of the proposed resolution enhancement method in terms of visual fidelity and objective criterion. We compare our algorithm with some well-known image SR algorithms, such as the nearest neighbor, cubic B-spline interpolation method, and Yang’s method [16], to validate the efficiency of our method. In our method, the root patch size is 16×16 , the depth of the tree is 3, and the number of training patches in the training process is 100,000. For the multitopic dictionary, the number of atoms in the general dictionary is 1,000. The number of atoms is the same in the multitopic dictionaries. We assume K determined topics ($K = 500$). For Yang’s method, the number of atoms is 1,000. We present the SR results of images (with a scale factor of 3) obtained using different methods in Figure 10. We extract the region after magnification within the red box to show the details after SR. We observe that the bicubic interpolation method blurs the sharpness of the edges and misses some fine details in the reconstructed images. Yang’s method [16] recovers a significant number of details but produces many jagged and ringing artifacts, along with edges or details. The proposed method obtains better visual quality than all of the other three competing methods.

Moreover, the PSNR and SSIM values of the SR results on LR images using various algorithms are listed in Table 1.

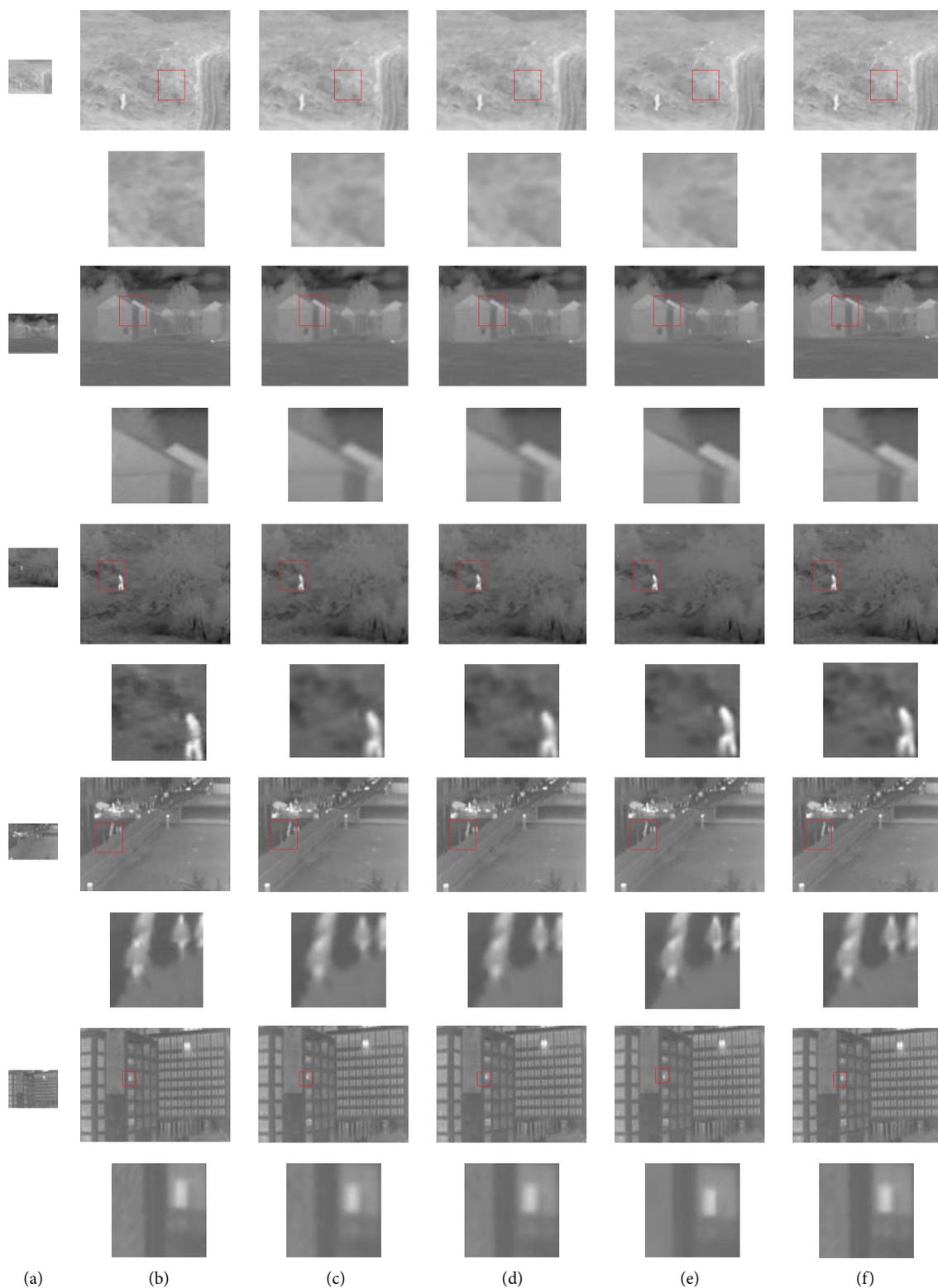


FIGURE 10: Visual comparison of four test images: (a) LR image, (b) original HR image, (c) results obtained using the nearest neighbor interpolation, (d) results obtained using the cubic B-spline interpolation, (e) results obtained using the sparse representation-based method, and (f) results obtained using the proposed method.

TABLE 1: Numerical results of Figure 10.

Method	Nearest neighbor		Cubic B-spline		Yang's method		Our method	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	36.77	0.4440	36.80	0.4458	36.98	0.5394	37.32	0.5857
	38.06	0.7210	38.07	0.7279	38.17	0.7691	38.20	0.8107
	36.19	0.4652	36.19	0.4713	36.63	0.6005	36.68	0.6405
	36.49	0.4933	36.48	0.5007	36.73	0.5477	36.93	0.5964
	32.91	0.3621	32.93	0.3646	34.02	0.3702	34.10	0.3741

We observe that the average PSNR and SSIM gains of the proposed method over Yang's method [16] and the bicubic interpolation method are in dB, which show that the SR results from the proposed method have better objective quality in terms of PSNR and SSIM.

Experiment 2 (effect of multisensor). To validate the effectiveness of multisensor by combining the information of visible images, we compared multisensor SRIR with traditional SRIR algorithm as Yang's method [16]. The number of training patches in the training process is 100,000. For the dictionary learning step, the number of atoms in the dictionary is 1,000. Figure 11 shows the SR results of the IR image. Figure 11(c) shows the results of the traditional SRIR algorithm as Yang's method [16], where severely jagged artifacts along the edges and annoying details are produced. The SR result is limited. Figure 11(d) shows the results of combining the information of visible images. We observe that the result is significantly improved qualitatively and quantitatively. The PSNR and SSIM values of the SR results on LR images using various algorithms are listed in Table 2.

Experiment 3 (effect of multiscale patches). We compare the SR results obtained from the dictionaries using multiscale patches and one fixed-scale patch. In the multiscale patches-based method, the root patch size is 16×16 . In the fixed-scale patch, fixed patches with three different patch sizes 4×4 , 8×8 , and 16×16 are analyzed. The number of training patches in the training process is 100,000. For the dictionary learning step, the number of atoms in the dictionary is 1,000. The reconstruction results are shown in Figure 12. We have observed that different images prefer different patch sizes for

TABLE 2: Numerical results of Figure 11.

Method	Yang's method		Multisensor	
	PSNR	SSIM	PSNR	SSIM
	34.28	0.4123	35.53	0.4237
	36.63	0.6005	36.65	0.6247
	36.73	0.5477	36.87	0.5751
	33.07	0.3731	34.57	0.3769

optimal performance. The multiscale treatment can help represent the image in a more efficient manner, thereby allowing applications to provide a more global look of the image. We observe that the reconstructed HR images obtained from the multiscale patches-based method, as shown in Figure 12(f), are better in terms of quantitation and visual perception than those obtained from the single-scale patches-based methods, as shown in Figures 12(c) to 10(e). The PSNR and SSIM values of the SR results on LR images using various algorithms are listed in Table 3.

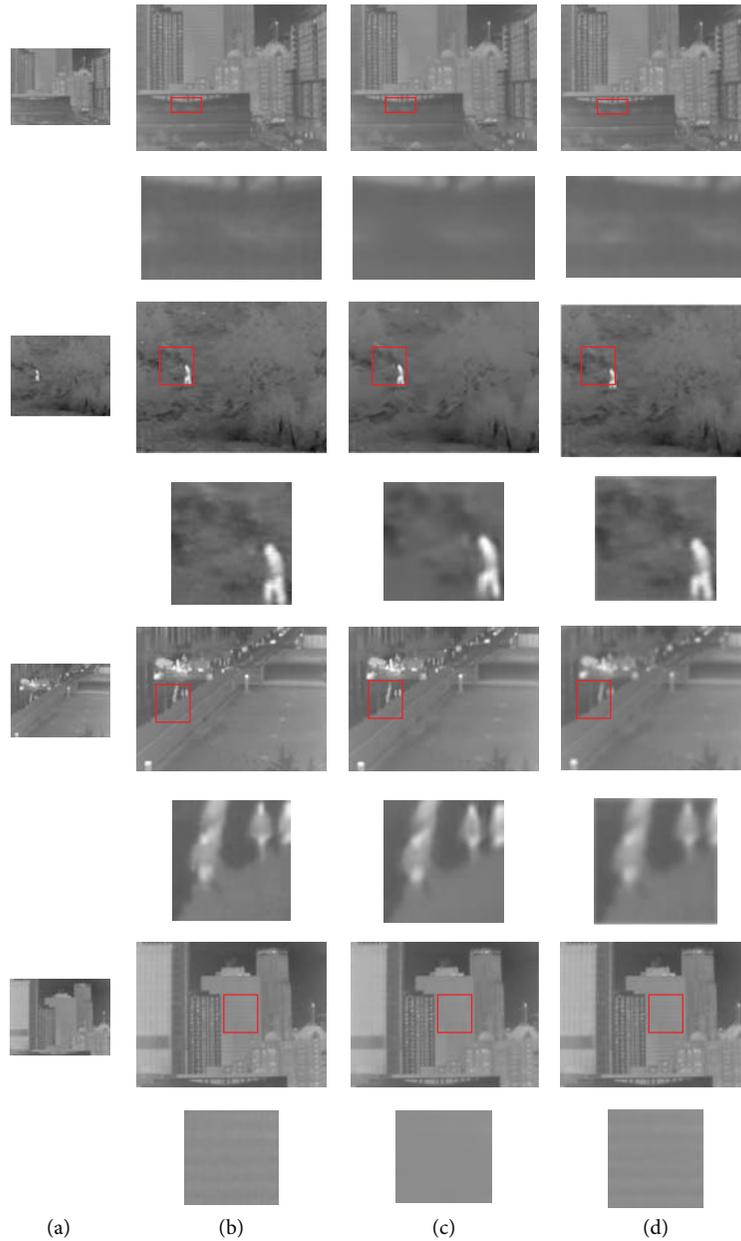


FIGURE 11: Visual comparison of four test images: (a) LR image, (b) original HR image, (c) results obtained using the sparse representation-based method, and (d) results obtained using the multisensor based method.

4. Conclusion

We proposed a novel sparse representation-based image SR method. The algorithm combines detailed information in visible images to improve the resolution of the IR image. Given the complementary nature of these types of information, the proposed method can generate state-of-the-art results in SR tasks. Considering the fact that the optimal sparse domains of natural images can vary significantly across different images and different image patches in a single image, the proposed method uses a simple model that generates pyramid images and divides the pyramid images into multiscale patches to

represent the image in a more efficient manner. We also partition the natural images into documents and group the documents to determine the inherent topics using pLSA and to learn the sparse dictionary of each topic using the sparse dictionary learning technique. Extensive experimental results show that our proposed method can achieve competitive performance compared to state-of-the-art methods.

Conflict of Interests

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

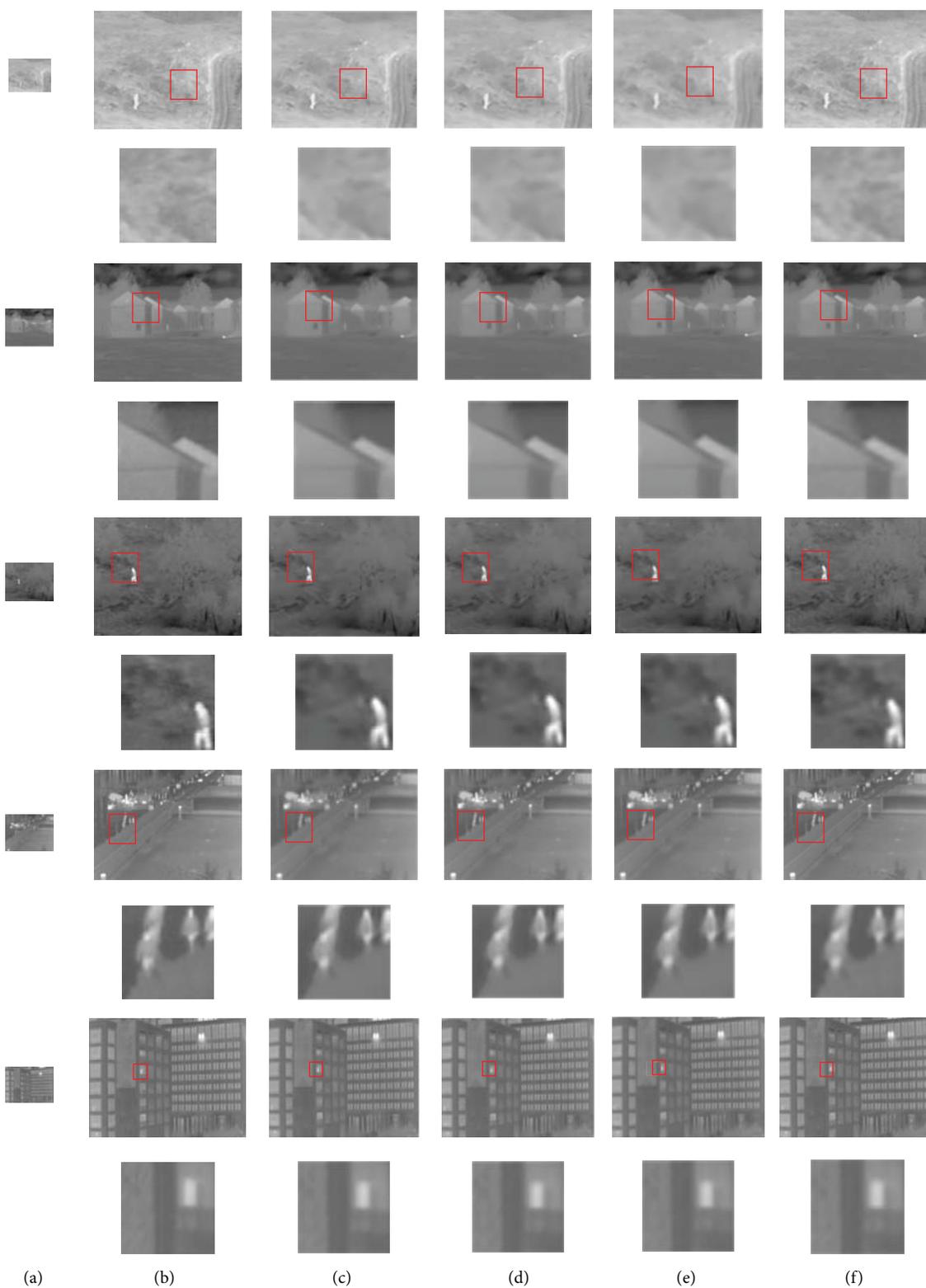
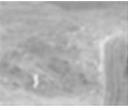


FIGURE 12: Visual comparison of four test images: (a) LR image, (b) original HR image, (c) results obtained with patch size 4×4 , (d) results obtained with patch size 8×8 , (e) results obtained with patch size 16×16 , and (f) results obtained using the proposed method.

TABLE 3: Numerical results of Figure 12.

Method	Patch size 4×4		Patch size 8×8		Patch size 16×16		Multiscale patches	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	36.96	0.5378	36.93	0.5327	36.92	0.5298	37.27	0.5679
	38.07	0.6247	38.15	0.6538	38.18	0.7721	38.19	0.7894
	36.64	0.6109	36.24	0.5371	36.17	0.4928	36.65	0.6362
	36.51	0.4959	36.72	0.5468	36.57	0.5179	36.89	0.5683
	34.04	0.3723	34.01	0.3703	33.72	0.3674	34.06	0.3726

Acknowledgments

The research is sponsored by the National Natural Science Foundation of China (nos. 61271330 and 61411140248), the Research Fund for the Doctoral Program of Higher Education (no. 20130181120005), the National Science Foundation for Postdoctoral Scientists of China (no. 2014M552357), the Science and Technology Plan of Sichuan Province (no. 2014GZ0005), and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References

- [1] R. Kimmel, "Demosaiicing: image reconstruction from color CCD samples," *IEEE Transactions on Image Processing*, vol. 8, no. 9, pp. 1221–1228, 1999.
- [2] X. Li, "Demosaiicing by successive approximation," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 370–379, 2005.
- [3] T. S. Huang and R. Y. Tsai, "Multi-frame image restoration and registration," *Advances in Computer Vision and Image Processing*, vol. 1, pp. 317–339, 1984.
- [4] H. F. Shen, L. P. Zhang, B. Huang, and P. X. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 479–490, 2007.
- [5] L. P. Zhang, H. Y. Zhang, H. F. Shen, and P. X. Li, "A super-resolution reconstruction algorithm for surveillance images," *Signal Processing*, vol. 90, no. 3, pp. 848–859, 2010.
- [6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.
- [7] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.
- [8] K. Zhang, X. Gao, X. Li, and D. Tao, "Partially supervised neighbor embedding for example-based image super-resolution," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 2, pp. 230–239, 2011.
- [9] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 35, no. 3, pp. 425–434, 2005.
- [10] W. Wu, Z. Liu, and X. He, "Learning-based super resolution using kernel partial least squares," *Image and Vision Computing*, vol. 29, no. 6, pp. 394–406, 2011.
- [11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [12] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 1, pp. 275–282, Washington, DC, USA, June–July 2004.
- [13] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.
- [14] G. Monaci and P. Vanderqheynst, "Learning structured dictionaries for image representation," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '04)*, vol. 4, pp. 2351–2354, IEEE, Singapore, October 2004.
- [15] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, vol. 6920 of *Lecture Notes in Computer Science*, pp. 711–730, 2012.
- [16] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

- [17] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.
- [18] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4544–4556, 2012.
- [19] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning*, vol. 42, no. 1-2, pp. 177–196, 2001.
- [20] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.
- [21] W. Wu, X. Yang, Y. Pang, J. Peng, and G. Jeon, "A multifocus image fusion method by using hidden Markov model," *Optics Communications*, vol. 287, no. 1, pp. 63–72, 2013.
- [22] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS '08)*, pp. 1033–1040, December 2008.
- [23] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990.
- [24] P. Purkait and B. Chanda, "Image upscaling using multiple dictionaries of natural image patches," in *Computer Vision—ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5–9, 2012, Revised Selected Papers, Part III*, vol. 7726 of *Lecture Notes in Computer Science*, pp. 284–295, Springer, Berlin, Germany, 2013.
- [25] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister, "Statistics of infrared images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–7, June 2007.

Research Article

Active Suppression of Narrowband Noise by Multiple Secondary Sources

Seokhoon Ryu, Yun Jung Park, and Young-Sup Lee

Department of Embedded Systems Engineering, University of Incheon, 119 Academy-ro, Yeonsu-gu, Incheon 406-772, Republic of Korea

Correspondence should be addressed to Young-Sup Lee; ysl@incheon.ac.kr

Received 19 March 2015; Revised 23 June 2015; Accepted 28 June 2015

Academic Editor: Marco Anisetti

Copyright © 2016 Seokhoon Ryu 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.

This study presents theoretical and experimental investigation on the active suppression of narrowband noise with C1, C1.5, and C2 components by using multichannel secondary sources in a duct. The quality manipulation in the duct was controlled by changing quality factors which were incorporated into a multichannel FxLMS algorithm. The algorithm is extensively investigated in both theory and real-time control experiment. After analysing the primary and secondary paths of the duct system, an acoustic narrowband signal was chosen as a primary noise and the impulse responses were implemented as the secondary path models. Control results show that the quality factors in the algorithm that was implemented in a dSPACE 1104 provide a stable and excellent response compared to before control. It is obvious that the lower quality factor cancels the more primary noise as defined in the theory although the attenuation levels are not exactly and inversely proportional to the quality factor. The results in this study can be used for practical active sound quality control systems.

1. Introduction

The active noise control (ANC) systems seek usually to maximize the attenuation of a primary noise by cancelling the unwanted noise based on the principle of superposition [1, 2]. It takes advantage of the biologically inspired adaptive feedforward learning algorithms such as the filtered-reference least mean square (FxLMS) to compensate for the effect of the secondary path in order to ensure convergence [1, 2]. This ANC has been widely applied successfully to many applications such as airplanes, cars, headsets, mobile devices, and other consumer electronics.

However, in contrast to an ANC system, in some applications it is necessary to retain residual noise with a specified target profile over frequency [3, 4] because some intentional residual noise could provide better natural feeling rather than simply minimizing the residual noise. This approach is known as an active sound quality control (ASQC) and it can be implemented by incorporating some factors in its algorithm. A typical ASQC system, therefore, needs to have a capability of attenuating unwanted noise to a certain level and, at the same time, of enhancing wanted noise to

a predefined target level. An adaptive noise equalization (ANE) is one of the concepts and has been extending to control narrowband and broadband noises [5–8]. Although a narrowband ANC system can reduce narrowband noise components at maximum, a narrowband ANE system can be independently manipulated.

The duct systems are widely applied to a number of different buildings, factories, and facilities. In some applications, such as clean rooms for precise manufacturing, they need to keep a comfortable noise level for workers and this leads to product quality. Also excessive reduction of noise can cause uncomfortable awkward silence. Thus the quality factor allows operating the noise reduction level with the ANC systems.

In this study, thus, an active sound quality control (ASQC) system using this ANE algorithm based the FxLMS is investigated in depth in order to control narrowband noise, which are dominant in a duct, separately or altogether in terms of sound quality control. For the selective quality control, a quality factor or parameter which is predetermined is applied in the real-time FxLMS adaptive algorithm.

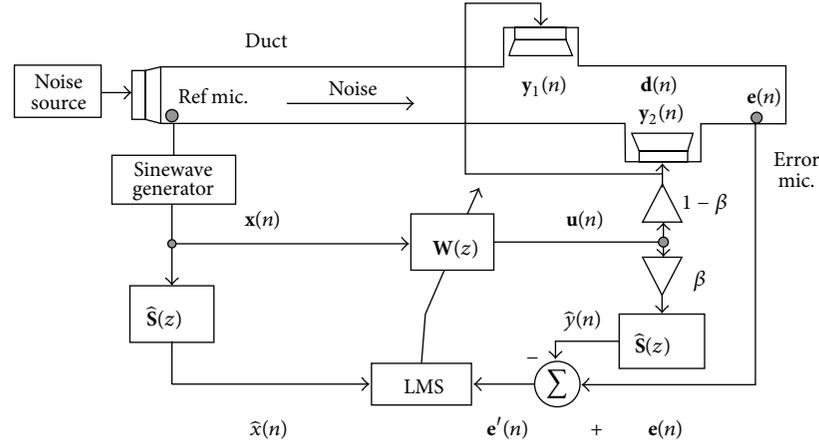


FIGURE 1: Active sound quality control for narrowband noise using a factor β in a duct.

The rest of the paper is organized as follows. In Section 2, theoretical considerations are described to derive proper equations for sound quality control based on the multichannel FxLMS algorithm with a quality factor matrix in a duct. Section 3 presents the experimental arrangement including the test duct and control board for the real-time control. In the same section, the primary and secondary paths are discussed and the modeling of the secondary path to implement in the control algorithm is included. Intensive analysis and discussions from the measured results from the real-time sound quality control in terms of time and frequency domain are written in Section 4. Finally, conclusions are summarized in Section 5.

2. Theoretical Considerations

2.1. Multichannel Active Sound Quality Control for Narrowband Noise. An ANC system with an adaptive feedforward approach is known as an effective method to suppress unwanted noise [1, 2]. As shown in Figure 1, a duct with a noise source (primary source), a secondary source, and an error microphone are considered to suppress the primary noise as much as necessary. A real-time algorithm which is embedded in a digital signal processor is applied for sound quality control as illustrated schematically in Figure 1 as well. The summation of the primary noise $d(n)$ and the secondary noise $y(n)$ at the error microphone is the error signal $e(n)$ as $e(n) = d(n) - y(n)$.

Assuming that there are K reference signals, M secondary loudspeakers, and L error microphones, the practical form of the multichannel FxLMS update equation based on the steepest descent algorithm in the time domain without the sound quality factor can be given as [1, 2]

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \alpha \widehat{\mathbf{X}}^T(n) \mathbf{e}(n), \quad (1)$$

where \mathbf{w} is the vector of filter coefficient, α is a convergence coefficient, and $\mathbf{e}(n)$ is the error signal vector.

$\widehat{\mathbf{X}}(n)$ is the filtered-reference signal matrix which can be expressed as

$$\widehat{\mathbf{X}}(n) = \begin{bmatrix} \widehat{\mathbf{x}}_1^T(n) & \widehat{\mathbf{x}}_1^T(n-1) & \cdots & \widehat{\mathbf{x}}_1^T(n-I+1) \\ \widehat{\mathbf{x}}_2^T(n) & \widehat{\mathbf{x}}_2^T(n-1) & \cdots & \widehat{\mathbf{x}}_2^T(n-I+1) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\mathbf{x}}_L^T(n) & \widehat{\mathbf{x}}_L^T(n-1) & \cdots & \widehat{\mathbf{x}}_L^T(n-I+1) \end{bmatrix}, \quad (2)$$

where the filtered-reference signal vector can be written by

$$\begin{aligned} \widehat{\mathbf{x}}_\ell(n) &= [\widehat{x}_{\ell 11}(n) \ \widehat{x}_{\ell 12}(n) \ \cdots \ \widehat{x}_{\ell 1K}(n) \ \widehat{x}_{\ell 21}(n) \ \cdots \ \widehat{x}_{\ell MK}(n)]^T. \end{aligned} \quad (3)$$

In (1), the practical update quantity in the multichannel system is given as $\partial J / \partial \mathbf{w}$, where J is the cost function and is identical to the gradient of the instantaneous sum of squared outputs of the error signals with respect to the filter weights. Thus it can be presented as follows [1, 2]:

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{w}(n)} &= \frac{\partial \mathbf{e}^T(n) \mathbf{e}(n)}{\partial \mathbf{w}(n)} \\ &= 2 \left[\widehat{\mathbf{X}}^T(n) \mathbf{X}(n) \mathbf{w}(n) + \widehat{\mathbf{X}}^T(n) \mathbf{d}(n) \right] \\ &= 2 \widehat{\mathbf{X}}^T(n) \mathbf{e}(n). \end{aligned} \quad (4)$$

In this control system, sound quality can be controlled using a digital equalizer [2]. The quality factor matrix \mathbf{B} , which consists of each quality factor β_i to each secondary loudspeaker as presented in Figure 1, can be considered in the multichannel FxLMS algorithm and can allow actively controlling sound quality in a duct system. The output vector $\mathbf{u}(n)$ from the adaptive filter matrix \mathbf{W} is designed to separate and to multiply with $\mathbf{I} - \mathbf{B}$ for the actual secondary path matrix \mathbf{S} in one branch and \mathbf{B} for the modelled one $\widehat{\mathbf{S}}$ in the other

branch. This makes the error signal vector capable of being written by

$$\mathbf{e}(n) = \mathbf{d}(n) - (\mathbf{I} - \mathbf{B}) \sum_{p=1}^{P-1} \mathbf{S}_p \mathbf{u}(n-p). \quad (5)$$

This error signal vector now indicates that the residual noise after control can be kept by the quality factor matrix \mathbf{B} [5]. Thus, the practical update equation of the multichannel FxLMS algorithm for ASQC with the quality factor matrix \mathbf{B} is given by

$$\begin{aligned} \mathbf{w}_{n+1} &= \mathbf{w}_n + \alpha \widehat{\mathbf{X}}^T(n) \left[\mathbf{e}(n) - \mathbf{B} \sum_{p=1}^{P-1} \widehat{\mathbf{S}}_p \mathbf{u}(n-p) \right] \\ &= \mathbf{w}_n + \alpha \widehat{\mathbf{X}}^T(n) \left[\left\{ \mathbf{d}(n) - (\mathbf{I} - \mathbf{B}) \sum_{p=1}^{P-1} \mathbf{S}_p \mathbf{u}(n-p) \right\} \right. \\ &\quad \left. - \mathbf{B} \sum_{p=1}^{P-1} \widehat{\mathbf{S}}_p \mathbf{u}(n-p) \right] = \mathbf{w}_n + \alpha \widehat{\mathbf{X}}^T(n) \mathbf{e}'(n). \end{aligned} \quad (6)$$

The new update equation in (6) operates automatically to minimize the signal vector $\mathbf{e}'(n)$. So if the disturbance signal vector $\mathbf{d}(n) \approx \sum_{p=1}^{P-1} \mathbf{S}_p \mathbf{u}(n-p)$, then (5) can be expressed as

$$\mathbf{e}(n) \approx \mathbf{B} \mathbf{d}(n). \quad (7)$$

Equation (7) represents the quality factor matrix governing the residual noise level at the error microphones positions in the duct. The separation of the signal y_i introduces four different cases of results as follows: (1) if $\beta_i = 0$, it acts as a normal ANC system which is on (this is equivalent to (1)) and makes $e_i(n) \approx 0$, (2) if $\beta_i = 1$, it acts as the ANC system which is off and makes $e_i(n) \approx d_i(n)$, (3) if $0 < \beta_i < 1$, it acts to control the noise to a certain level dependent upon the value β_i and makes $e_i(n) \approx \beta_i d_i(n)$, and (4) if $\beta_i > 1$, it acts to enhance the noise level and makes $e_i(n) > d_i(n)$.

3. Implementation and Experiment

3.1. Experimental Set-Up and Primary Path. The experimental set-up for the real-time narrowband active sound quality control consists of an acrylic duct (length = 1800 mm), a primary loudspeaker at the left end, two secondary sources (control loudspeakers), an error microphone (1/2" PCB 377B0E) at the right end, a power amplifier (B&K 2716C), a PCB signal conditioner, and low-pass filters. In addition, it involves the real-time control unit of a dSPACE 1104 for the implementation of control algorithm. So there are 1 reference signal ($K = 1$), 2 secondary loudspeakers ($M = 2$), and 1 error microphone ($L = 1$) in this control system.

The sampling frequency was $f_s = 6.000$ kHz and the cut-off frequencies of the antialiasing low-pass filters before the two ADCs (for the reference signal and the error signal) and the reconstruction low-pass filter after the DAC (for the control signal) were both 500 Hz. The control signal vector $\mathbf{u}(n)$, which was generated in the control algorithm, was

transferred to the control loudspeakers through the DAC. The physical acoustic path lengths between the secondary sources and the error microphone are 1140 mm for the source 1 and 600 mm for source 2, respectively.

In a duct system, fan noise is the most dominant one in general. Thus a narrowband signal is highly likely as the input. Other different signals pure tones or broadband signals can be the input. Control against pure tones is very easy but the usefulness in practical applications is very low. Control against broadband signals is useful but the signals require very long control filters for successful control. However, many practical applications such as fan, motors, engines, and other rotating devices generate nonstationary narrowband signals. The narrowband control based on the adaptive notch filter needs two filter coefficients per order, and this allows less control filter lengths. By the way, this study investigates the narrowband signal which consists of three different orders.

Thus, a narrowband signal with three components of C1 (mode 1), C1.5 (mode 2), and C2 (mode 3) is considered for the primary disturbance noise as shown in Figure 2(a).

The spectrogram of the disturbance signal in Figure 2(a) indicates that three components are involved as described previously and it stays stationary for the first and the last 2 seconds, respectively, but sweeps nonstationarily for 5 seconds between the two stationary signals. During the sweep, C1, C1.5, and C2 order signals vary from 100 Hz to 200 Hz, 150 Hz to 300 Hz, and 200 Hz to 400 Hz, respectively.

In Figure 2(b), the block diagram of the multichannel narrowband FxLMS which was implemented in dSPACE 1104 is displayed. The block diagram shows adaptive notch filters to cancel actively the narrowband disturbances. The block diagram in Figure 2(b) is the extended version of Figure 1 to generate proper control signals to operate the two secondary loudspeakers. Each dashed rectangle in Figure 2(b) represents an adaptive notch filter against one of the three orders to drive one of the two loudspeakers.

3.2. Secondary Path Modelling. The plants of the active control system which is known as secondary paths between each control loudspeaker, LS1 or LS2, and the error microphone were measured with the dSPACE 1104 as plotted in Figure 1. Both the plant models $\widehat{S}_1(z)$ and $\widehat{S}_2(z)$ were obtained using the offline identification method from the measured data. As it can be seen from Figure 3, the frequency response function (FRF) and the impulse response functions (IRF) are plotted.

The FIR filters $\widehat{s}_1(n)$ and $\widehat{s}_2(n)$ were implemented for the plant models in the control algorithm as shown in Figure 2(b). The lengths of the FIR filters were 50 samples, respectively, although the IRFs in Figure 3 show 200 samples. The length of $\widehat{s}_1(n)$ depends upon the performance of the control processor for the real-time control and the complexity of the algorithm used.

The lengths of the secondary path models in FIR filters were chosen after analyzing the mean square of the error signal which is the difference of the outputs between the actual secondary path and the secondary path model. The analysis showed that the length of 50 samples is reasonable to choose as it offers nice control performance in terms of the

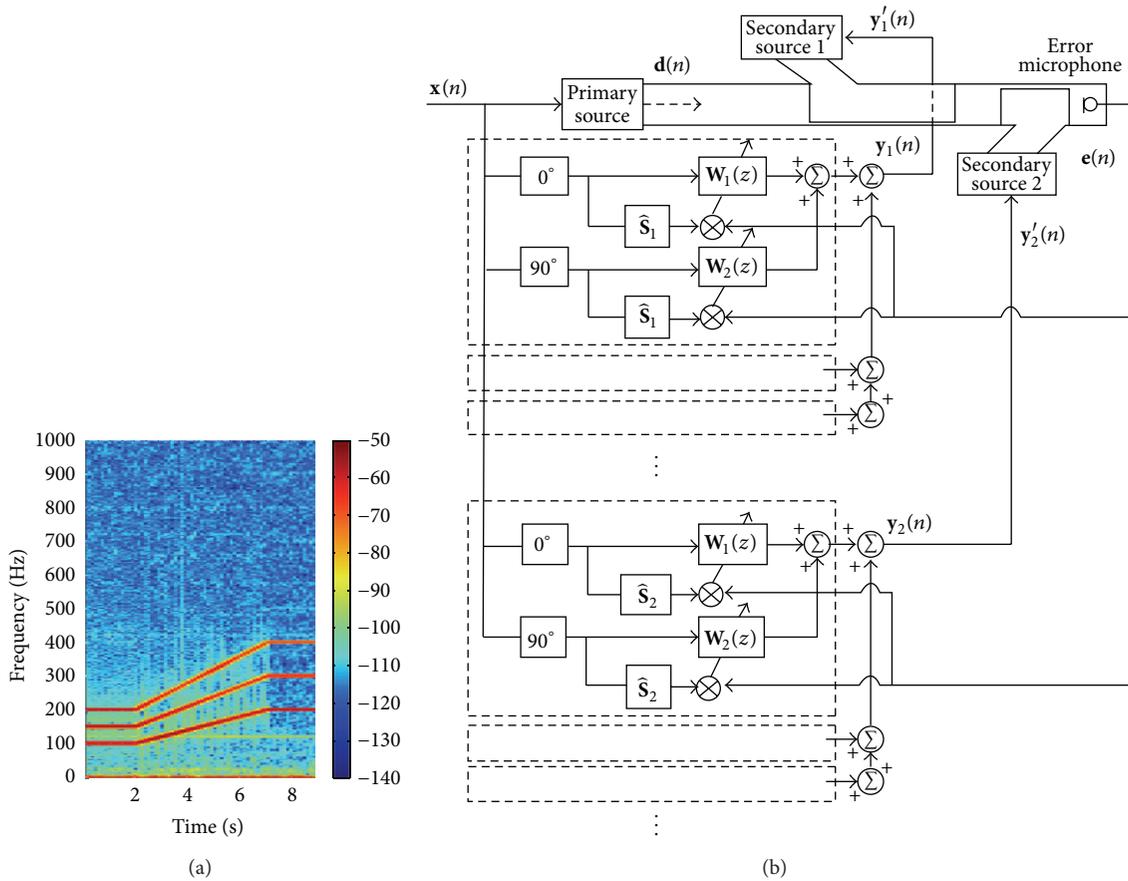


FIGURE 2: Disturbance and the control algorithm. (a) Spectrogram of the disturbance signal. (b) Implementation of the multichannel FxLMS algorithm for narrowband noise control.

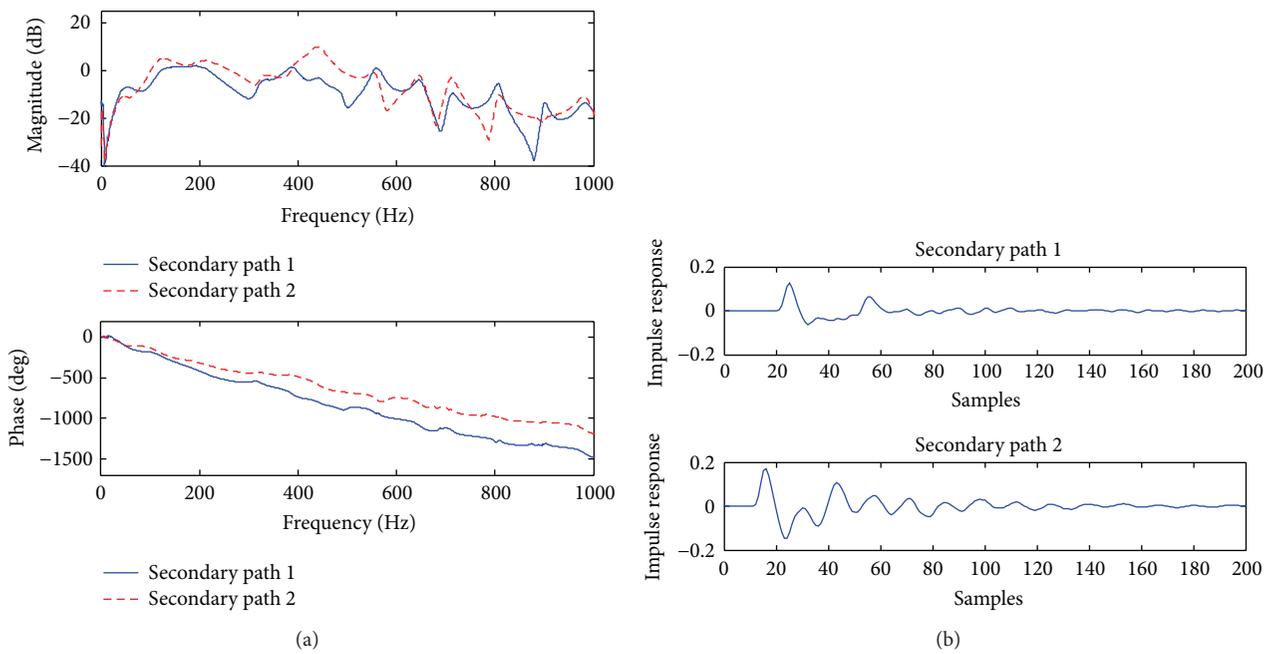


FIGURE 3: Modelling of the two secondary paths. (a) Frequency response function (b) impulse response function of the modelled plants $\hat{S}_1(z)$ and $\hat{S}_2(z)$.

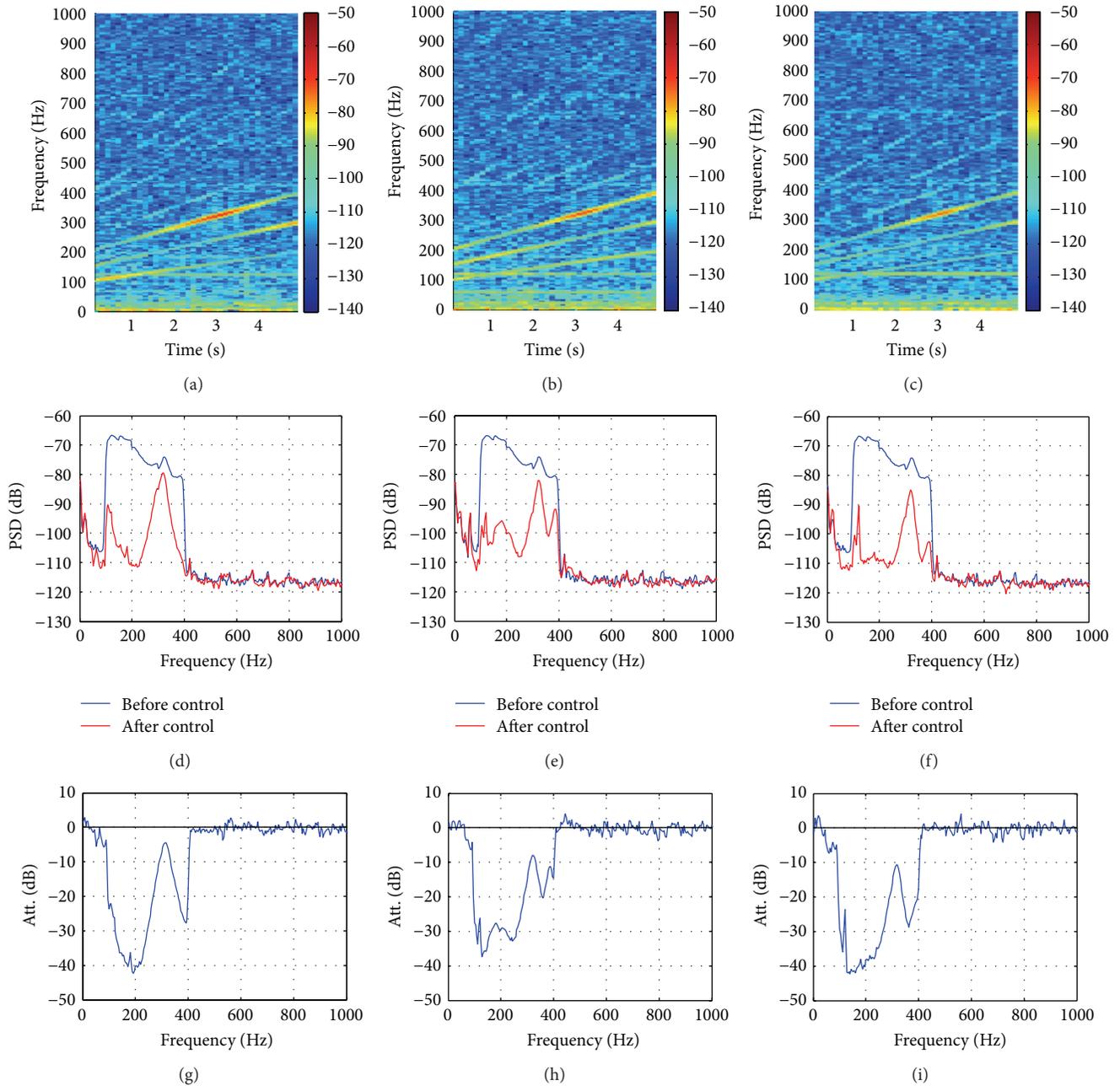


FIGURE 4: Real-time narrowband control results without the quality factor. ((a), (b), (c)) Spectrograms after control when LS1 only, LS2 only, and LS1 + LS2 are operated, respectively. ((d), (e), (f)) Comparison of error spectra between “before control” and “after control.” ((g), (h), (i)) Comparison of attenuations.

MSE (mean square error) and real-time computation time. The longer filter length gives the further reduction in control but it requires further real-time processing capability.

4. Results and Discussions

In this experiment, either LS1 or LS2 or both (LS1 + LS2) the two secondary loudspeakers were used for the control.

Also the control results present without ($\beta = 0$) and with the quality factor ($0 \leq \beta \leq 1$).

4.1. Active Control Results without the Quality Factor ($\beta = 0$).

When the quality factor was not applied ($\beta = 0$) to the control implementation, the real-time control experiment results show the full performance with the largest attenuations as shown in Figure 4. Figures 4(a), 4(d), and 4(g) indicate the spectrogram of the error signals after control, comparison of

the error spectra $S_{ee}(f)$ between before and after control, and the attenuation $\text{Att}(f)$, respectively, when *only LS1 is used*. The attenuation is defined as follows:

$$\begin{aligned} \text{Att}(f) &= 10 \log_{10} \left[\frac{S_{ee,\text{after}}(f)}{S_{ee,\text{before}}(f)} \right] \\ &= 10 \log_{10} \left[\frac{S_{ee,\text{after}}(f)}{S_{dd}(f)} \right], \end{aligned} \quad (8)$$

where the subscripts after and before represent “after control” and “before control,” respectively, and $S_{dd}(f)$ is the disturbance spectrum.

When only LS1 is used, the spectrogram of the error signals after control, compared with Figure 2(a), shows that three sweeping components (C1, C1.5, and C2) are attenuated very well. The average attenuation level over the frequency range of 100–400 Hz in this case is about 25.62 dB as summarized in Table 1.

Figures 4(b), 4(e), and 4(h) show that the average attenuation level of the error signals over the same frequency range is about 24.09 dB as *only LS2 is used*. Also Figures 4(c), 4(f), and 4(i) demonstrate that the average attenuation level is about 29.74 dB when *both LS1 and LS2 are used*.

It is noted that the case with both LS1 and LS2 shows the best performance compared with the other two cases. It is eminent from the results in Figure 4 that the frequency region especially around 320 Hz (C3) is not controlled effectively. This is caused by the fact that the eigenvalue spread is inherently wider than other regions in this duct control system.

The less controlled peak around 300 Hz in Figure 4 is caused by the sudden phase change in the secondary path around that frequency as plotted in Figure 3(a), which is determined by the physical dimension of the duct and the physical and electrical properties between the secondary loudspeakers and the error microphone. In active control, the sudden phase change at a certain frequency especially limits the convergence coefficient value in order to maintain the stability of the control system. So the sudden phase change determines the control performance. Once the convergence coefficient value increases, the peak around that frequency can be increased abruptly and threatens the stability. Thus even if the duct dimension cannot be redesigned, one can reduce the amount of the phase change at a certain frequency by modifying the physical and electrical properties of the secondary paths. Then the change can lead to better control performance.

From Table 1, the result differences between LS1 only and LS2 are observed although they are not big. This is mainly caused by the differences in terms of the physical and electrical properties between the two secondary paths. The convergence coefficient values are different slightly in the two paths and bring about the result difference.

In addition, although it might look so natural this justifies the fact that the use of more control loudspeakers can introduce better attenuation in active noise control provided that the complexity of the algorithm is not beyond the

TABLE 1: Average attenuation in dB of the error signals without/with the quality factor.

	LS1 only (dB)	LS2 only (dB)	LS1 + LS2 (dB)
C1 ($\beta = 0.0$)	5.28	5.97	—
C1.5 ($\beta = 0.0$)	1.40	1.22	—
C2 ($\beta = 0.0$)	6.03	5.82	—
C1, C1.5, C2 ($\beta = 1.0$) <i>Control off</i>	0.00	0.00	0.00
C1, C1.5, C2 ($\beta = 0.5$) <i>Quality control on</i>	—	—	5.84
C1, C1.5, C2 ($\beta = 0.2$) <i>Quality control on</i>	—	—	13.71
C1, C1.5, C2 ($\beta = 0.0$) <i>Control on (full)</i>	25.62	24.09	29.74

computation power of the real-time processor and the control is stable.

4.2. *Active Control Results for Each Component without the Quality Factor ($\beta = 0$).* Again if the quality factor was not applied ($\beta = 0$) to the narrowband control implementation, the spectrograms of the error signals after control on each component by using either LS1 only or LS2 only are plotted in Figure 5.

Figures 5(a), 5(b), and 5(c) show that C1, C1.5, and C2 components are cancelled out by using LS1 only. The noise at about 320 Hz in C2 is still not suppressed even if the control is dedicated only to one component (C2) by LS1. The average attenuation levels in this case are summarized in Table 1.

In Figures 5(d), 5(e), and 5(f), the spectrograms of the error signals after control are quite similar to Figures 5(a), 5(b), and 5(c). It is noted that the average attenuation at C1.5 with LS2 is a bit better than LS1 from Figures 5(b) and 5(e), although the average attenuation level in Table 1 shows differently. Because the C1.5 which extends from 150 Hz to 300 Hz overlaps with C1 at 150–200 Hz and C2 at 200–300 Hz, the average attenuation levels for C1.5 cannot represent actual results in this case.

4.3. *Active Control Results with the Quality Factor ($0 \leq \beta \leq 1$).* In this section, three different quality factors of $\beta = 0$, $\beta = 0.2$, and $\beta = 0.5$ were applied to the actual control implementation; the real-time control experiment results in terms of their attenuations when both LS1 and LS2 are used are shown in Figure 6. As Figures 6(a), 6(b), and 6(c) indicate the attenuation at $\beta = 0$ (full control), $\beta = 0.2$, and $\beta = 0.5$, respectively, the largest attenuation was observed at $\beta = 0$.

The control results including the error spectrum $S_{ee}(f)$ and the attenuation $\text{Att}(f)$ in the frequency domain are presented in Figures 7(a) and 7(b) with 4 different β of 1.0 (before control, thin lines), 0.5 (quality control, dashed lines), 0.2 (quality control, thick lines), and 0.0 (full control, thickest lines).

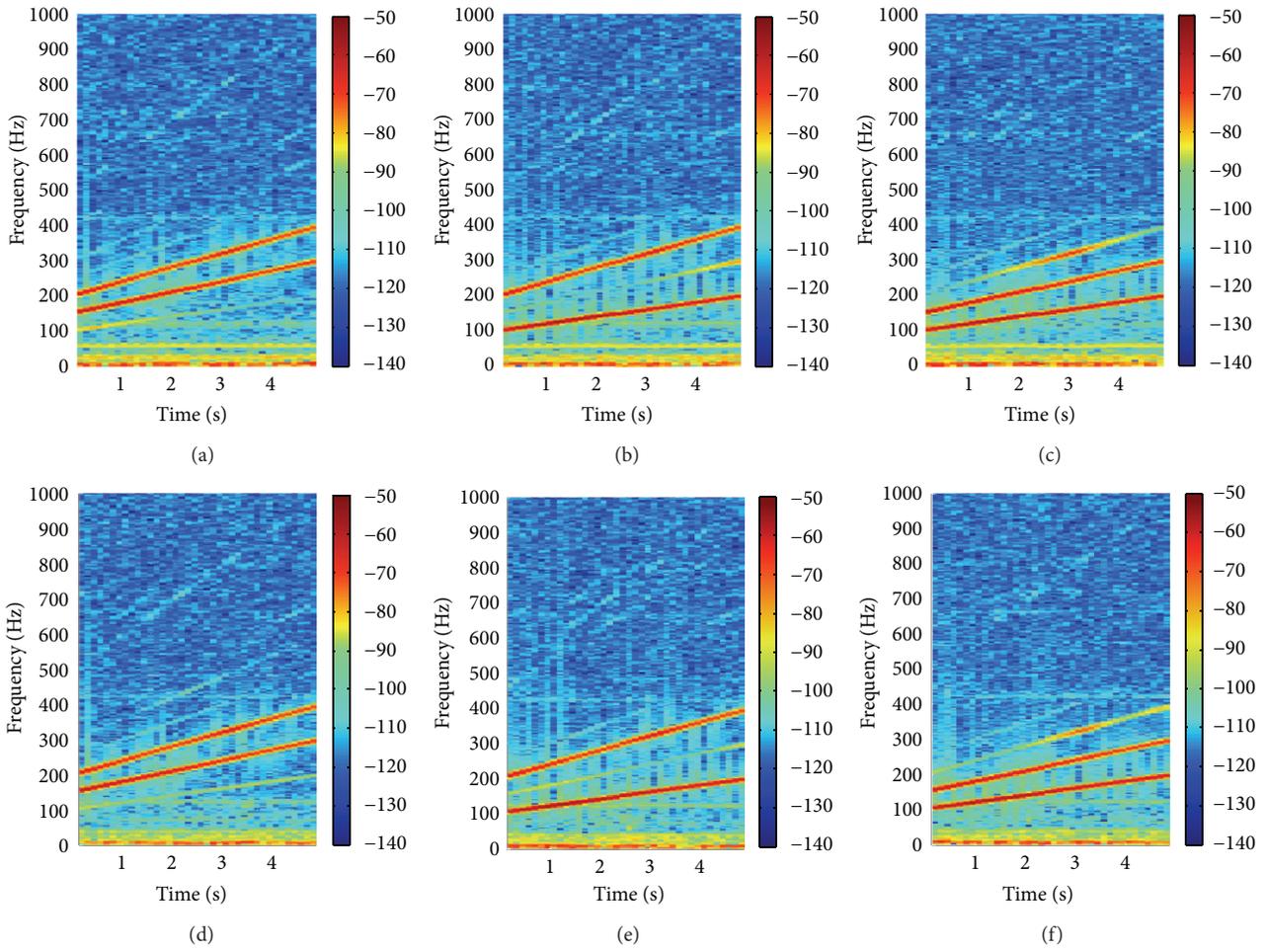


FIGURE 5: Spectrograms of the error signals from real-time narrowband control results for each component without the quality factor. (a) C1 control with LS1. (b) C1.5 control with LS1. (c) C2 control with LS1. (d) C1 control with LS2. (e) C1.5 control with LS2. (f) C2 control with LS2.

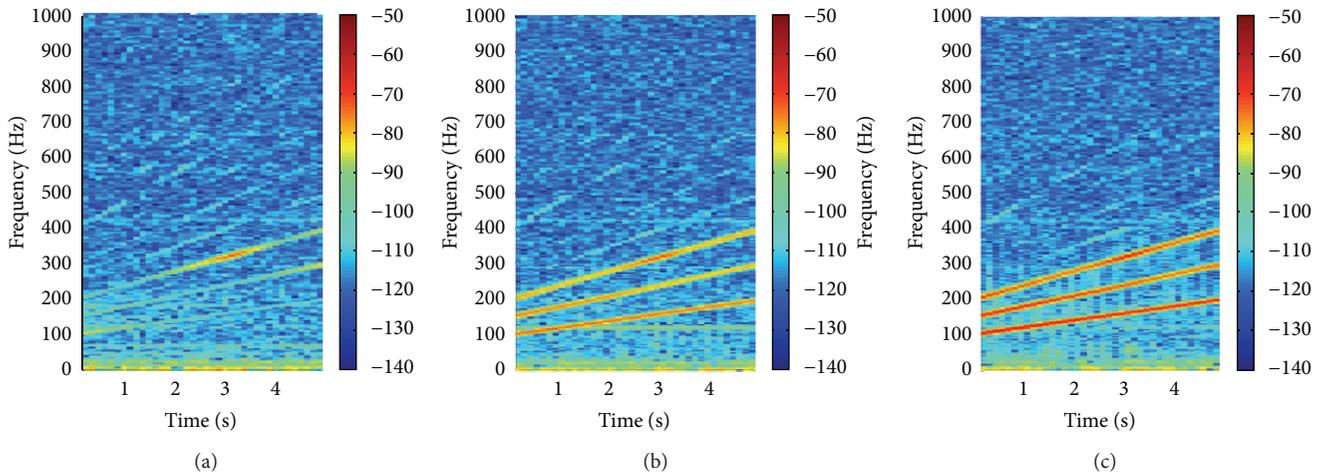


FIGURE 6: Spectrograms of the error signals from real-time narrowband control results with the quality factor when both LS1 and LS2 are operated. (a) $\beta = 0$. (b) $\beta = 0.2$. (c) $\beta = 0.5$.

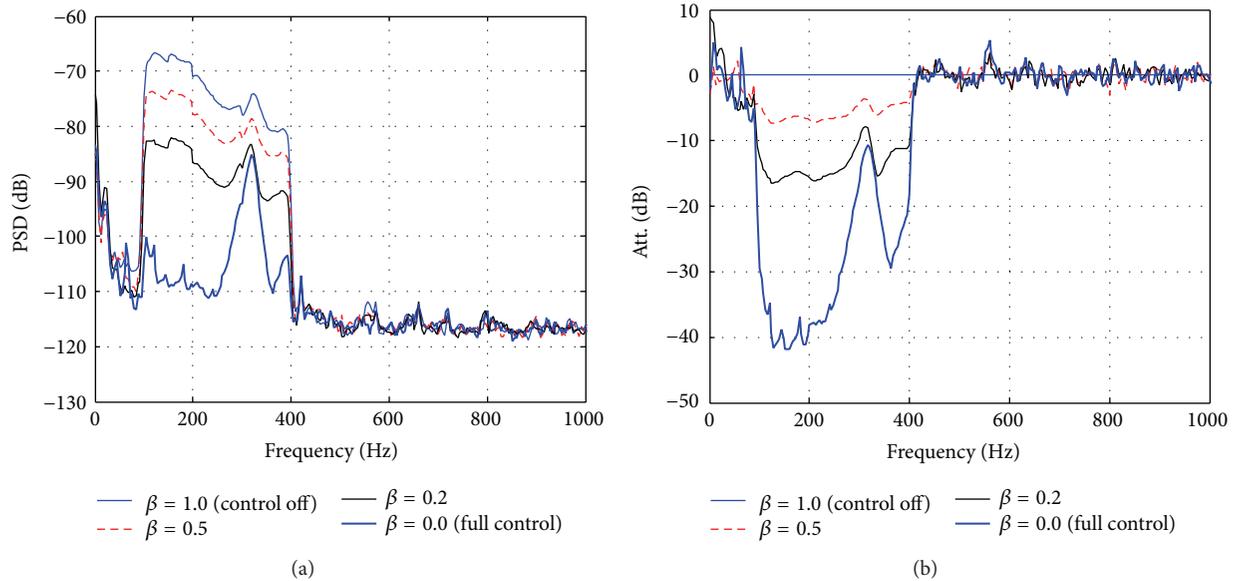


FIGURE 7: Comparison of control performances in terms of error spectra and attenuation between before control ($\beta = 1.0$: thin lines) and after control ($\beta = 0.5$: dashed lines, $\beta = 0.2$: thick lines, and $\beta = 0.0$, thickest lines) against narrowband noise. (a) Error spectra. (b) Attenuations.

Figure 7 shows the effect of the sound quality factor variation dramatically in both the error spectra and the attenuation. As summarized in Table 1, the average attenuations over the frequency range of 100 Hz–400 Hz for $\beta = 0.0$, $\beta = 0.2$, $\beta = 0.5$, and $\beta = 1.0$ are 29.74 dB, 13.71 dB, 5.84 dB, and 0.00 dB, respectively.

It is obvious that the lower β cancels the more primary noise as defined in the theory although the attenuation levels are not exactly and inversely proportional to the quality factor β . It is also observed that there are no spillover phenomena after control beyond the frequency of 400 Hz in the low-pass filters and this indicates that this approach operates with high stability and very nice performance over a wide frequency range.

The control results shown in Figures 6 and 7 demonstrate that any specific acoustic modes can be suppressed as much as it is necessary by using the quality factor β . This can be extended to even more complicated sound quality control systems such as vehicles or other products.

By the way, the optimal quality factor can be calculated when a target profile is given before control in this approach. The quality factor can be adapted automatically during control when the target profile over frequency is predefined before control. For the implementation of those systems, it is necessary to insert the command input (target profile) signal into the block diagram in Figure 1; then the difference between the command input and the error signal is feedback to the LMS algorithm.

As an ASQC system requires controlling the primary noise to a certain target profile over frequency, it is expected that this approach could provide a practical solution.

In addition, if the primary input signal is broadband, the quality factor will work equally over the frequency range. This will lead to attenuation of the noise level but the amount of the

attenuation might be less than that for the narrowband input signal.

5. Conclusions

This study presents active sound quality control of some acoustic modes (C1, C1.5, and C2 components) in a 1-dimensional duct by means of the quality factor β manipulation, which was implemented in the real-time multichannel FxLMS algorithm based on biologically inspired learning. Control results show that the algorithm with or without the quality factor provided stable and excellent responses in experiments compared to before control. As either or both the two control loudspeakers in the duct can be used, the three components were controlled at the same time or separately with large attenuations.

For the individual control with either LS1 or LS2, the average attenuations are quite similar with or without the quality factor. LS1 especially showed 25.62 dB reduction in overall when beta = 0, but LS2 gave 24.09 dB reduction.

The quality factor value influences tremendously control results; as the values decrease with beta = 1.0, 0.5, 0.2, and 0.0, the attenuation levels increase to 0.00, 5.84, 13.71, and 29.74 dB when LS1 and LS2 are used.

It is obvious that the lower β cancels the more primary noise as defined in the theory although the attenuation levels are not exactly and inversely proportional to the quality factor. The multichannel sound quality algorithm worked with nice stability and performance against a narrowband noise. The results in this study can be used for practical active sound quality control systems.

For the future, a target profile based control will be investigated to extend this study for actual application

to a large duct system. The new control scheme will contain the command input (target profile over frequency) and the input will be compared with the error signal to drive the adaptive FxLMS algorithm. The quality factor will be automatically updated at each sample to meet the target profile.

Conflict of Interests

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

Acknowledgment

This work was supported by the Incheon National University Research Grant in 2012.

References

- [1] Y. Kajikawa, W. S. Gan, and S. M. Kuo, "Recent advances on active noise control: open issues and innovative applications," *APSIPA Transactions on Signal and Information Processing*, vol. 1, article e3, 21 pages, 2012.
- [2] L. V. Wang, W. S. Gan, A. W. Khong, and S. M. Kuo, "Convergence analysis of narrowband feedback active noise control system with imperfect secondary-path estimation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2403–2411, 2013.
- [3] Y.-S. Lee and S. J. Elliott, "Active position control of a flexible smart beam using internal model control," *Journal of Sound and Vibration*, vol. 242, no. 5, pp. 767–791, 2001.
- [4] D. Bismor, "LMS algorithm step size adjustment for fast convergence," *Archives of Acoustics*, vol. 37, no. 1, pp. 31–40, 2012.
- [5] S. J. Elliott, *Signal Processing for Active Control*, Academic Press, 2001.
- [6] S. O. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 5th edition, 2013.
- [7] J. C. Burgess, "Active adaptive sound control in a duct: a computer simulation," *The Journal of the Acoustical Society of America*, vol. 70, no. 3, pp. 715–725, 1981.
- [8] K. Lee, W. S. Gan, and S. M. Kuo, *Subband Adaptive Filtering: Theory and Implementation*, Wiley, 2009.

Research Article

A Tile-Based EGPU with a Fused Universal Processing Engine and Graphics Coprocessor Cluster

Yang Wang,¹ Li Zhou,¹ Tao Sun,² Yanhu Chen,¹ Lei Wang,² and Shaotao Sun³

¹*School of Information Science and Engineering, Shandong University, No. 27, South Shanda Road, Jinan 250100, China*

²*Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, No. 336, West Nan Xinzhuang Road, Jinan 250022, China*

³*Administration Center, Shandong Academy of Information and Communication Technology, Jinan 250101, China*

Correspondence should be addressed to Li Zhou; zhou.li@sdu.edu.cn

Received 16 March 2015; Revised 15 May 2015; Accepted 18 May 2015

Academic Editor: Gwanggil Jeon

Copyright © 2016 Yang 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.

As various applied sensors have been integrated into embedded devices, the Embedded Graphics Processing Unit (EGPU) has assumed more processing tasks, which requires an EGPU with higher performance. A tile-based EGPU is proposed that can be used in both general-purpose computing and 3D graphics rendering. With fused, scalable, and hierarchical parallelism architecture, the EGPU has the ability to address nearly 100 million vertices or fragments and achieves 1 GFLOPS per second at a clock frequency of 200 MHz. A fused and scalable architecture, constituted by Universal Processing Engine (UPE) and Graphics Coprocessor Cluster (GCC), ensures that the EGPU can adapt to various graphic processing scenes and situations, achieving more efficient rendering. Moreover, hierarchical parallelism is implemented via the UPE. Additionally, tiling brings a significant reduction in both system memory bandwidth and power consumption. A 0.18 μm technology library is used for timing and power analysis. The area of the proposed EGPU is 6.5 mm * 6.5 mm, and its power consumption is approximately 349.318 mW. Experimental results demonstrate that the proposed EGPU can be used in a System on Chip (SoC) configuration connected to sensors to accelerate its processing and create a proper balance between performance and cost.

1. Introduction

With the development of embedded applications, various embedded platforms and devices have become an essential part of people's daily lives [1]. Mobile phones, hand-held electronics, and automobile electronics have greatly changed the way people live. As the most significant part of an embedded platform, the Graphics Processing Unit (GPU) plays an important role in an embedded system [2].

The concept of GPU was first put forward by NVIDIA in 1999 [3]. Its powerful performance in both 3D graphic processing acceleration and general-purpose computing has attracted considerable attention from researchers in different fields [4, 5]. The hardware design of GPUs has made dramatic progress during the last decade. The modern GPU has evolved from a traditional fixed-function graphics pipeline to a programmable parallel processor. Traditional graphics processing pipelines consisted of fixed-function stages

without programmability. With the advent of vertex and fragment shaders, vertices, and pixels could be processed, respectively. However, when these shaders were applied in different processing cases, typical workloads of vertex and fragment shaders were usually not well balanced, which led to inefficiency. Unified shader architecture was then introduced to enable dynamic load balancing of mutative vertex- and pixel-processing workloads [6].

Mobile applications have achieved great success in recent years [7]. The performance of the Embedded GPU (EGPU) has become one of the most crucial factors in evaluating embedded platforms. Compared with a desktop GPU, an EGPU requires equivalent processing performance, reduced energy consumption, better portable APIs, low cost, and more efficient use of memory bandwidth. These critical factors relate to each other dependently and tightly, which determines the optimization strategies in EGPU hardware design. The Immediate Mode Renderer (IMR) and Tile-Based

Renderer (TBR) are two of the most popular renderers in modern EGPU hardware design. Whereas a traditional IMR renders all objects within the screen's boundaries, a TBR renders each screen tile one after the other until the full image is rendered. Furthermore, an IMR relies on the depth values in the Z-buffer to sort the final results. As a result, obscured fragments are still processed, and the amount of overdraw has increased by between 4 and 5 times as modern games have become more realistic [8], causing a great waste in memory bandwidth. On the contrary, TBR determines what is visible and only renders what is necessary to avoid overdraw.

Tile-based methods split the display screen into tiles, which are independent of each other. Because each tile region occupies a small subset of the entire scene, reduced memory access can easily be implemented based on on-chip memory. There are several on-chip memories to support tile-based graphic processing, including the Z-buffer, color buffer, and tile frame buffer. A Z-test is performed based on the on-chip Z-buffer. Pixel processing and blending use the color buffer and the on-chip "tile frame buffer" is used as a local storage area. Compared with a conventional 3D EGPU, a tile-based EGPU reduces memory bandwidth costs and enhances the system performance. All the on-chip processing is performed at high depth and pixel accuracy at the full clock rate without external memory access latency. This approach greatly saves memory bandwidth and, thus, enables modern games and other graphics applications to run with optimized performance [9].

In this paper, a fused EGPU architecture is proposed. The Universal Processing Engine (UPE) cooperates with the Graphics Coprocessor Cluster (GCC), completing graphics processing tasks and general-purpose computation efficiently. This design is a flexible combination that can operate in both tile mode and blending mode to handle most scenes. The proposed tile-based EGPU platform renders opaque objects based on a fused and parallel architecture. Considering that tiles are independent from each other, more processing elements mean more powerful performance and higher parallelism. Two Universal Processors (UPs) are designed in this paper to address tiles rendering under the management of the Universal Processor Controller (UPC). Based on a SIMT (Single Instruction Multiple Threads) [10] architecture, different threads are concurrently distributed to Streaming Processors (SP) in the UPs. Both specific graphics processing and General-Purpose GPU (GPGPU) tasks can be executed and accelerated.

In this work, the following contributions have been made:

- (1) The UPE and GCC constitute fused and scalable architecture, allowing the whole platform to operate in different modes to meet various scenes.
- (2) Different tiles are assigned to different UPs in the rendering stage, making the system more efficient. SIMT architecture is applied in the UP, whose processing units are allocated with different threads. The entire UPE is implemented with hierarchical parallelism.
- (3) A flexible tiling mechanism is designed to finish the Z-test earlier, reducing overdraw and improving the efficiency.

- (4) A Memory Hub (MH) is introduced to solve the conflicts in memory access.

The rest of the paper is organized as follows. Section 2 describes the architecture of the proposed EGPU. The experimental results are provided in Section 3. Section 4 presents the conclusion and direction of further consideration for future works.

2. Tile-Based EGPU

In this section, a tile-based EGPU is presented in detail, as shown in Figure 1. The proposed EGPU architecture contains a UPE and GCC for processing acceleration. The UPE's processing operations are based on a unified programmable model, which can be used in both graphics processing and general-purpose computation. The GCC, including tiling, blending, texturing, and interpolating coprocessors, cooperates with the UPE to achieve enhanced graphics processing performance. The UPE and GCC are fused together in the EGPU to constitute a fused hardware platform for different processing situations and application cases. The other components, including command processing, data preparation, data preprocessing, and the raster operations, cooperate with the UPE and GCC to complete the entire graphics processing flow. All components communicate with each other via an internal network and buffers, working as a whole.

A command processor (CP) responds to commands from the host CPU and coordinates with other EGPU components to work harmoniously. The CP analyzes instructions, sends control signals to the data preparation (DP) unit, dispatches instructions to the UPE, and changes the rendering mode according to the current application scenes. The DP unit collects geometric primitives, such as points, lines, triangles, and fetches associated vertex attribute data from the MH. The fetched data are stored in an input buffer so that the UPE can access them directly. After the UPE completes the processing of vertex shading programs, the result data are written to the on-chip internal buffer and are then further processed by the Primitive Assembler (PA)/clip/viewport/setup/raster unit under the CP's control until realization of the final pixel fragments.

The PA assembles related vertices into triangles to build up basic geometry primitives. Then, the viewport and clip units clip the primitives into the standard view frustum. They transform the postclipping vertices into screen (pixel) space and reject primitives outside the view volume as well as back-facing primitives. Surviving primitives are then processed by the setup unit to generate edge equations for the rasterizer. Attribute plane equations are also generated for the linear attribute interpolation of pixels in the pixel shading stage. A coarse rasterization stage generates all pixel tiles that are at least partially covered by the primitive. The UPE reads the relative pixel fragment data from the internal buffer to complete pixel-fragment shading. Shaded pixel-fragments are sent across the interconnection network to the Raster Operation Processor (ROP) unit. Data in the output buffer are finally written to system memory via the MH.

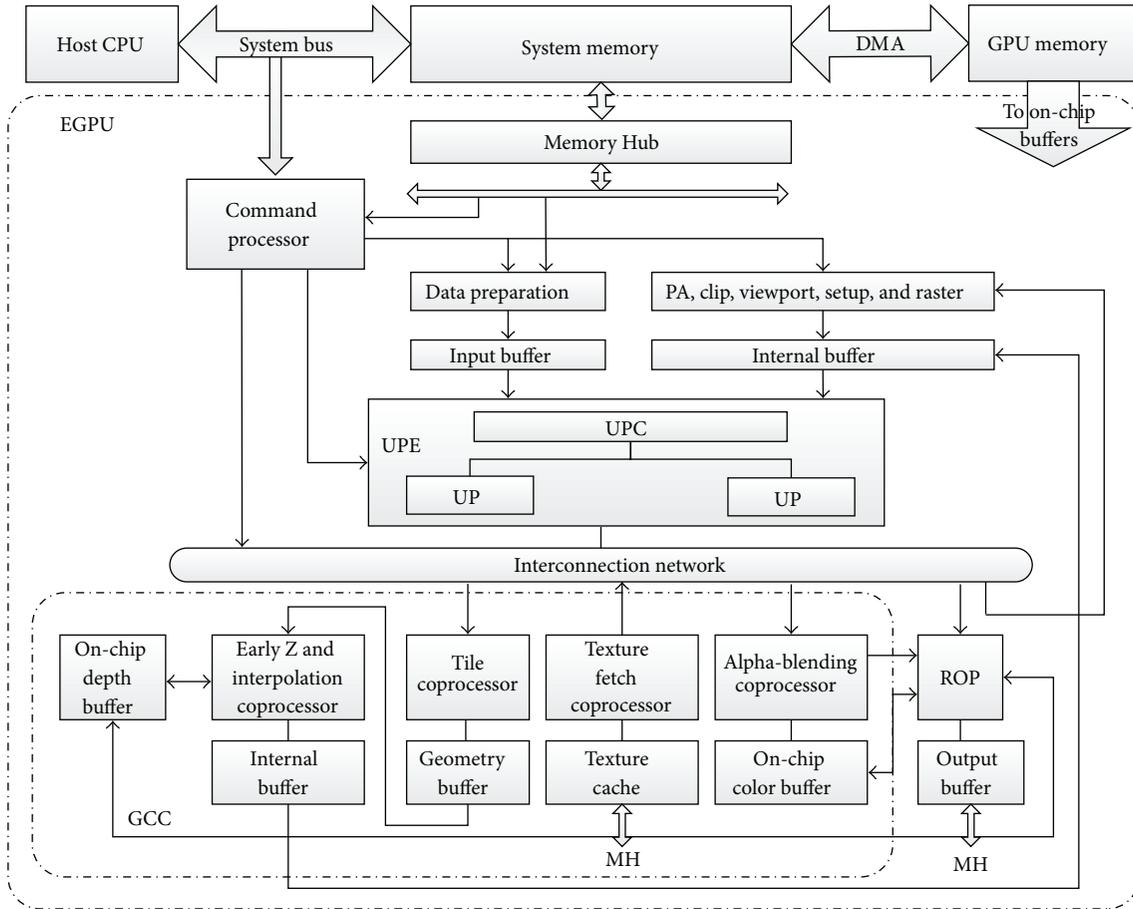


FIGURE 1: Architecture overview of the EGPU.

2.1. *Universal Processing Engine.* The UPE, as shown in Figure 2, is the processing core of the entire EGPU, undertaking the majority of the workload in both graphics processing and scientific computation. The UPE is designed as a parallel and hierarchical architecture, containing 2 unified UPs that work together under the management of the UPC. The UPC sends instructions and corresponding data from the input buffer or internal buffer to each UP, which executes vertex, geometry, and pixel shading programs and general computing programs. As the central control unit in the UPE, the UPC distributes various types of tasks to the UPs and balances the workload of each UP dynamically. Two independent UPs can work in parallel and improve the efficiency of the EGPU. In particular, in a tile-based EGPU architecture, each UP can process different tiles simultaneously in the fragment shading stage.

UP, based on a hierarchical architecture, is the basic and vital processing unit of the UPE. A UP is composed of an instruction pool (IP), a thread distributor (TD), a Hierarchical Processing Group (HPG), and internal memory. The TD reads instructions from the IP, which stores 32 instructions from the CP, and allocates them to the HPG. The HPG, including SPs and a Special Function Unit (SFU), is the execution unit of the UP. Once all data have been processed, they are written to the output buffer. General algorithms,

logic, and memory access can be processed in the HPG, which supports 16 specific operations for each type. Most EGPU operations are covered.

The memory unit in each UP consists of a constant buffer, shared memory, and relative memory. The constant buffer stores constants used in the EGPU shading stage. The shared memory belongs to the HPG and is divided into 5 banks for the 4 SPs and 1 SFU. Intermediate results are written into shared memory for data exchange in the HPG. Relative memory stores data corresponding to the current shading stage. An anticonflict mechanism is proposed to ensure ordered memory access. SPs and the SFU write data to their own bank to avoid data conflicts. Their reading and writing access is also given specific priority to ensure that the correct data are accessed in order.

Instruction coissue is implemented in the UP with the help of the TD. Instructions are tagged from 1D to 4D according to their dimension. All processors in the HPG, SPs, and SFU are 1D scalar ones that can only address 32-bit floating operations. The dimension reflects the width of the data that are waiting to be processed. Thus, the operations on 64-bit, 96-bit, and 128-bit data are defined as 2D, 3D, and 4D operations, respectively. While 128-bit operations can fully utilize the 4 SP cores, other operations must be combined to occupy all SP cores simultaneously. The

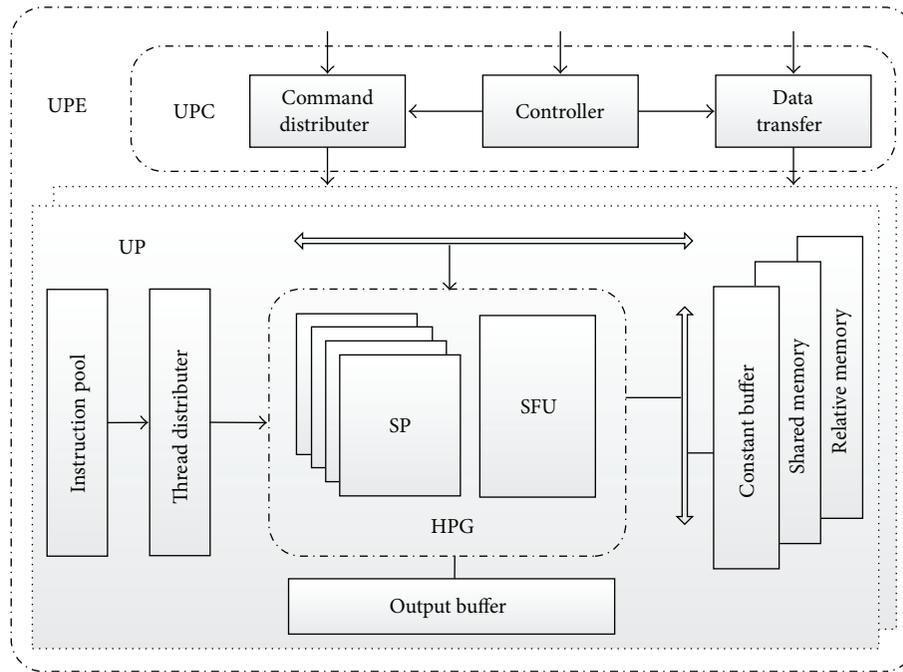


FIGURE 2: Architecture of the UPE.

combinations of 3D and 1D operations, 2D and 2D operations are used in this paper. For the most intricate occasion, the TD launches 4 independent 1D instructions simultaneously. With the coissue architecture, all the SP cores are occupied to make the instruction execution more efficient. Moreover, the TD distinguishes between basic operations and complex ones and distributes them to SPs and the SFU as different threads. By communicating with the HPG, the TD balances the workload dynamically. Additionally, a flexible memory access mechanism is designed to solve the memory access conflicts that arise with instruction coissue. Two AXI buses are valid for SPs, and another one is reserved for the SFU. Consequently, 4D, 3D and 1D, and 2D and 2D instructions can read corresponding data in one cycle. Four 1D instructions spend 2 cycles to finish a memory access.

Processing element (PE) and computing unit (CU) are the basic parallel processing units of the SP and SFU, respectively. The HPG is a hierarchical processing system. SPs and the SFU compose a scalable architecture to address basic and complex operations, respectively, which constitutes top-level parallelism. Additionally, the parallel processing among PEs and CUs is an underlying parallelism. Hierarchical parallelism makes it efficient to address those threads allocated by the TD.

The architecture of an SP and an SFU is shown in Figure 3. The SP and the SFU have similar architectures, and the SFU is more complicated to address those complex operations. The basic operations, such as adding, multiplying, and comparison, can be executed in one cycle. However, complex ones, such as trigonometric functions and logarithm, may need multiple cycles because of their complexities in accessing lookup tables (LUTs). Recognizing the fact that most operations are basic ones, the SFU is separated and

shared by the SPs to reduce the consumption of area on the chip [14].

An LUT is designed to yield a complex calculation result by indexing a predefined array to reduce processing time because retrieving a value from memory is usually faster than undergoing a complex computation. The width of the input operand influences the hardware area and accuracy of the LUT. The input of an LUT in this paper is 6 bits in width, with a 3-bit function control. Functions implemented in an LUT include reciprocal, sine, cosine, exponent, and binary logarithm.

Considering that each SP or SFU can only obtain one memory bus, PEs and CUs work in a pipeline to ensure that only one PE or CU occupies the memory bus in each clock cycle. In each PE, operations are divided into 5 substages: instruction fetch (IF), instruction decode (ID), read data (RD), execution (EXE), and write data (WD). An instruction is first fetched from the instruction FIFO, and then it is decoded in the ID stage. After data are prepared in the RD stage, the instruction is executed in the EXE stage. Finally, corresponding data are written to memory in the WD stage. The general-purpose registers (GPR) array is the interface for PEs and CUs to interact with shared memory. A CU is designed with a similar pipeline. Three extra cycles are added to address the requests of complex operations, and these cycles are defined as EXE0, EXE1, EXE2, and EXE3. Compared with PE cores, which can only address basic operations, CU cores contain more algorithm and logic resources for complex operations.

2.2. Graphics Coprocessor Cluster. Graphics coprocessors are integrated into the EGPU system to accelerate

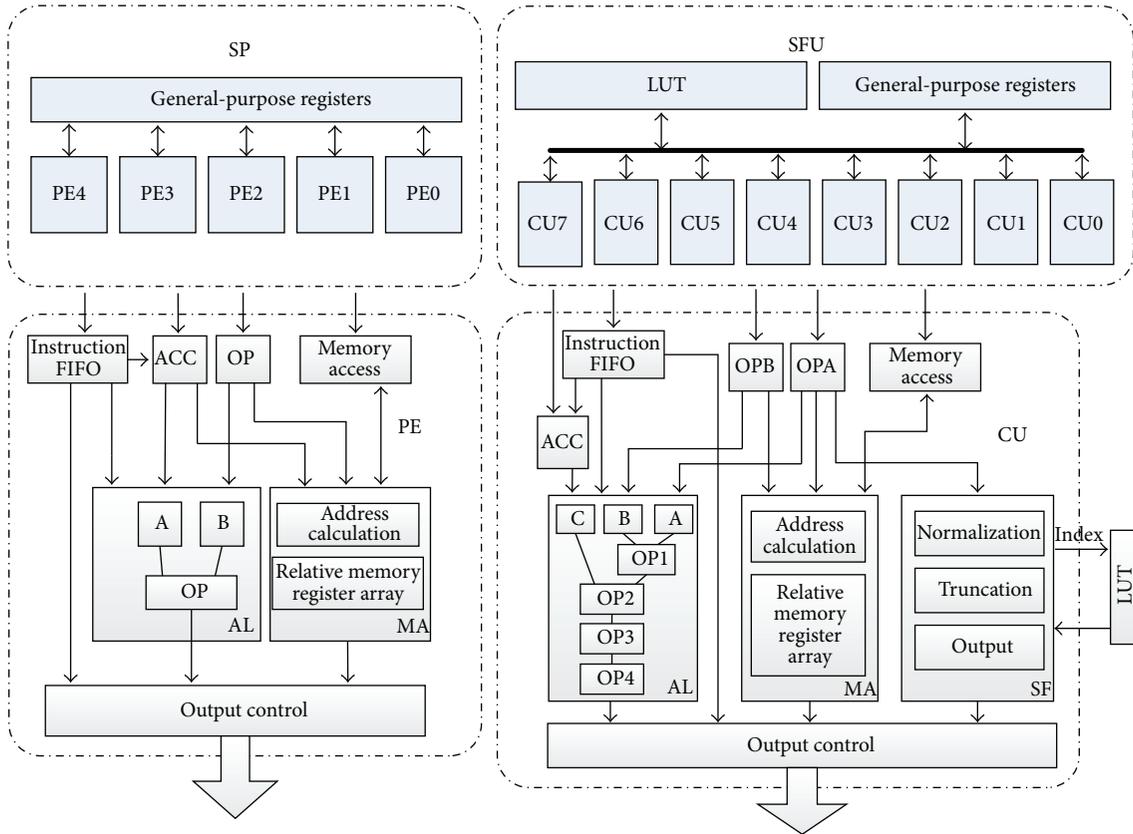


FIGURE 3: Architecture of an SP and SFU.

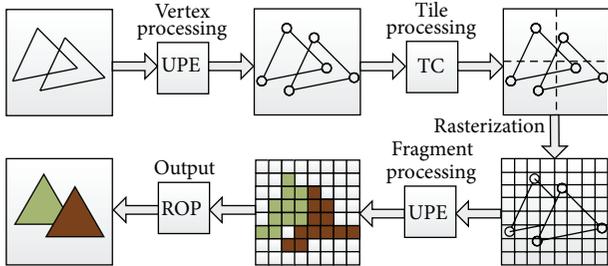


FIGURE 4: Processing flow with the tile architecture.

graphics-related calculations. In this design, four graphics coprocessors, including a tile coprocessor (TC), early Z and interpolation coprocessor (EZIC), texture fetch coprocessor (TFC), and alpha-blending coprocessor (ABC), are integrated to undertake specific graphics tasks. Various coprocessors are used in different modes under the control of the CP. Two modes are provided in this design, tile mode and blending mode. The first one is used for opaque scenes, as shown in Figure 4. The latter is applicable for transparent or blended scenes.

The TC, as shown in Figure 5, performs clip, project, and overlap test operations for geometric data that have been transformed by the UPE. In a unified shader-based system, the UPE executes vertex shading programs on geometric

data for coordinates transformation and other per-vertex operations, such as lighting. Result data are then given to the TC. The TC divides the screen into several tiles, judges the relationships between triangles within the screen and tiles, and writes relative information to the tile list. A Bounding Box Test (BBT) is used to decide the overlap between triangles and tiles. As shown in Figure 5 [8], if the triangle is judged as belonging to a specific tile, information of that triangle is written into the corresponding position of the tile list. After all triangles have passed the BBT, the TC updates all the tiles covered by objects and writes out the transformed data to the geometry buffer.

The number of tiles required to complete the render is determined by the resolution of the tiles. Larger tile size does improve performance, leading to fewer tiles to process and fewer tile lists to update. However, it also causes an increase of the on-chip memory requirements in the graphics core. The choice of tile size is a balance between graphics processing performance and the cost of additional resources. In this paper, the size of each tile is fixed at 32×32 to handle different scenes.

After tiling, object data are tested and processed by the EZIC. In general, the TC and EZIC are enabled simultaneously. The EZIC, as shown in Figure 6, compares the calculated depth information of each fragment with the values stored in the on-chip tile depth buffer to determine if the current fragment is visible. Only those visible fragments

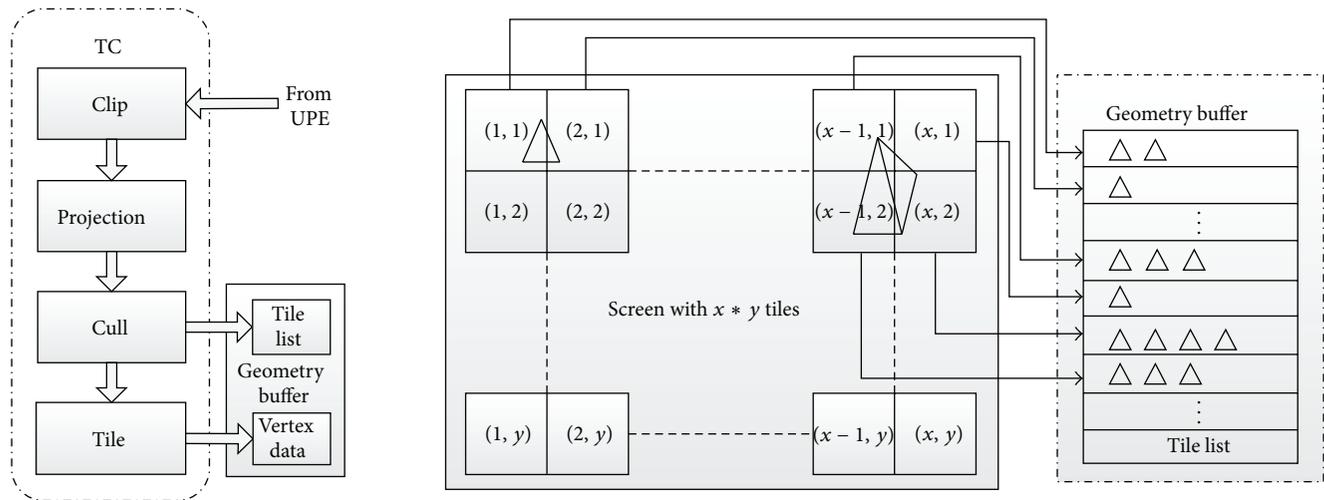


FIGURE 5: Tile mechanism. Left: architecture of the TC; right: TC tiling.

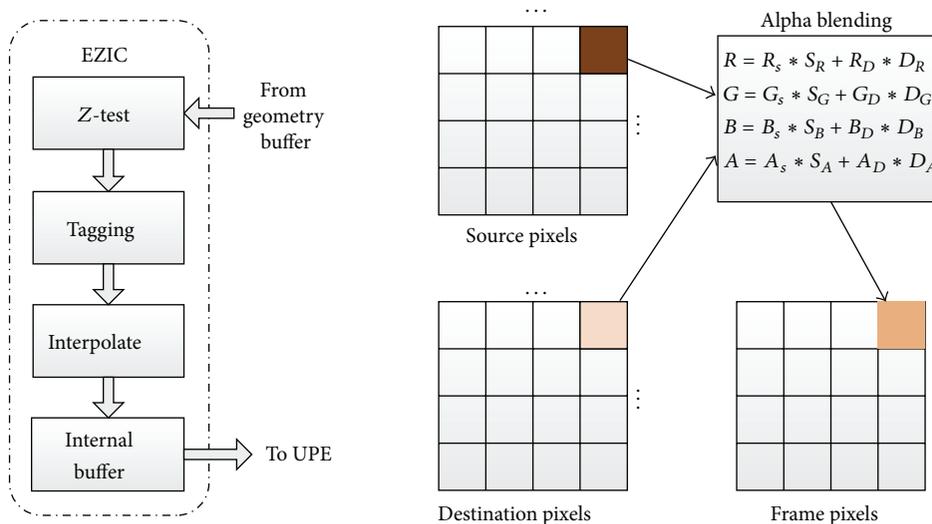


FIGURE 6: Architecture of the EZIC and alpha blending in the ABC.

are then interpolated and stored in the internal buffer. Fragments of different primitives are distinguished by tags to identify the primitive properties that should be used to texture the fragments later on. When all of the primitives in the tile list have been processed, the EZIC submits the remaining fragments back to the UPE for texturing and other operations. Additionally, the fragments belonging to the same primitive are organized in the same group with the help of their tags, to improve efficiency in memory access and fragment processing.

The TFC prefetches texture data based on calculated texture coordinates before the UPE begins to process fragments. The TFC calculates the corresponding address according to the texture coordinates and then sends reading requests to the MH. The prepared texture data are stored in the on-chip buffer for ease of access by the UPE.

The ABC is designed to draw transparent objects and implement graphics blending effects. To render blended

objects accurately, the hardware has to process each object individually as they may all contribute to the frame buffer's color. Unfortunately, with the tile architecture, the depth test is performed earlier, and the obscured objects are all discarded. As a result, alpha and blending cannot be simultaneously realized. To overcome this drawback, the ABC method is proposed. When the current scene requires transparency or a blending effect, the CP halts the TC and EZIC. The entire system exits tile mode and enters blending mode. Then, the CP dispatches the rendering task to the UPE and ABC. The UPE completes the vertex and fragment stages without tiling acceleration. At the end of rendering, the ABC calculates the new color according to the blending coefficients and the original color in the color buffer and then replaces the original color in the frame buffer, as shown in Figure 6. The ABC is designed for transparent or translucent scenes, which require blending operations. All blending operations are performed by accessing the on-chip color buffer so that

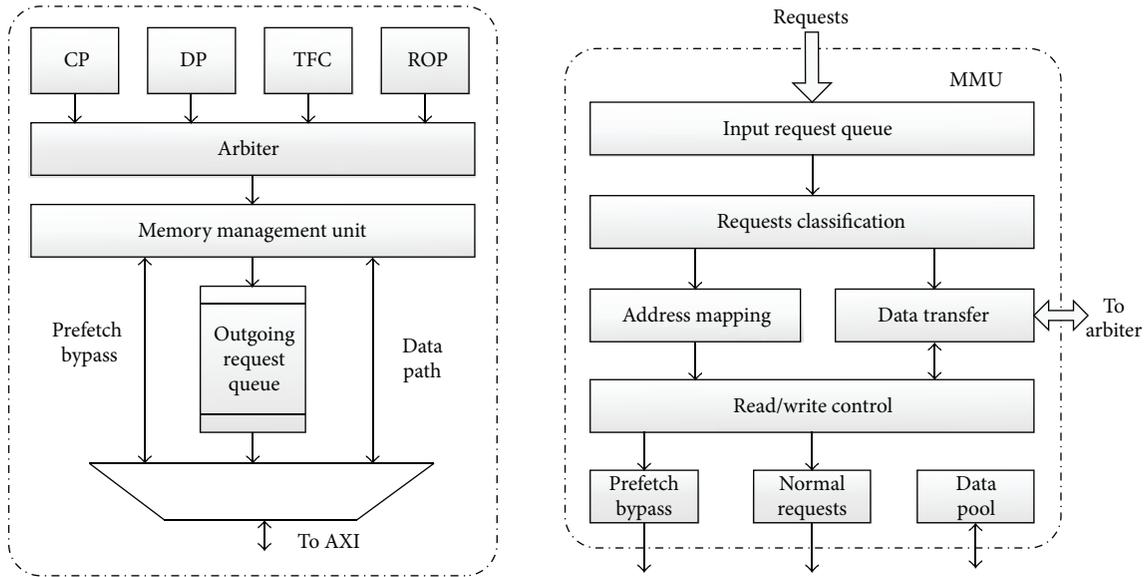


FIGURE 7: Architecture of the MH. Left: arbitration in the MH; right: architecture of the MMU.

they can be executed very quickly and not waste additional system memory bandwidth. The architecture of the EGPU with an ABC can adapt to increased types of scenes and obtain further flexibility.

2.3. Memory Hub. The MH provides arbitration for outgoing requests to system memory and also buffers data simultaneously. The MH functions as an arbiter to handle conflicts among related parts during memory access. As shown in Figure 7, the MH interacts with the following EGPU modules: the CP, DP, TFC, and ROP. The CP fetches instructions from system memory via the MH. The DP reads vertex data from the MH and then sends them to the input buffer, which consists of a list of buffers, including the vertex buffer, color buffer, and normal buffer. Additionally, all of these buffers are used to store the corresponding data. The TFC fetches textures via the MH, and the ROP writes data back to the system with the help of the MH. Moreover, the CP and DP support multiple read targets, which are identified with different master IDs. The MH supports a total of 4 unique read masters and 2 write masters. The MH responds to those requests and gives them permission to read or write. When several memory requests are sent to the MH simultaneously, the MH gives turn-around access to one of these blocks based on a predefined priority. Then, the memory bus will be occupied and busy. The memory management unit (MMU) will map the reading or writing address to the input address and number, and the following requests will enter the queue to wait for a response. Only when the memory bus is released by the current client will the MH continue to address the remaining outstanding requests. The fixed priority is designed according to the following order of graphics processing for increased efficiency: CP → DP → TFC → ROP.

The MMU handles the reading and writing requests of the MH, as shown in Figure 7. Because different requests have been arbitrated, the MMU responds to the current request, calculates the memory address, and establishes the data path. Requests are firstly classified. Then, their access addresses are calculated according to their index and number. A prefetching bypass mechanism is designed in this paper, ensuring that high priority or emergency requests can be responded to in time. The order of reading and writing requests is controlled to avoid potential conflicts. Data are transferred through the data transfer block.

3. Experimental Results

In this section, several experimental results are provided to evaluate the performance of the EGPU system.

A system verification platform is established to verify the performance of the EGPU design proposed in this paper. A 3D graphics scene has been developed by transferring related data to a test case manually to meet the requirements of the EGPU's input. This graphics scene includes approximately 1000 vertices and 800 triangles. No point and line primitives are included. As a result, the EGPU processes the data successfully and produces a 2D image in the frame buffer.

The speed of the processor is regarded as one of the most important evaluation factors. The clock cycles used by each instruction can reflect the speed directly. Moreover, the hardware design cost is also considered by means of the logic design area and power consumption. The balance between the performance and hardware cost has been studied during the design of this work. For the proposed UP, it can enhance the speed of processing at a low hardware resource cost. With a frequency of 200 MHz, an average of approximately 50 million vertices or fragments can be processed per second

TABLE 1: Performance analysis of a UP.

Case	Instruction package type: Number		Clock cycles used	Execution unit
1	1D * 4:	X	X + 5	SPs
2	1D + 3D:	X	X + 4	SPs
3	2D + 2D:	X	X + 4	SPs
4	4D:	X	X + 4	SPs
5	Complex operation:	X	X + 7	SFU

TABLE 2: Performance analysis of other modules.

Target: number	Processing time (Clock cycles)	Processing module
Triangles: X	[X, 7X]	Clipper
Primitive: X	[X, 3X]	PA
Primitive: X	Depends on the size of the primitives	TC

TABLE 3: Technology condition and results.

Processing technology	Voltage	Temperature
0.18 μm CMOS	1.68 V	125°C
Area	Power consumption	Frequency
6.5 mm * 6.5 mm	349.318 mW	200 MHz

on each UP. Table 1 presents the processing ability of a UP. Take case 1 as an example. For four 1D instructions packed together, the SPs take $X + 5$ cycles to process X packages.

In Table 2, a series of design indicators of corresponding modules are presented. Taking triangles as an example, the processing time is between X and $7X$ clock cycles to process X triangles. For some operations, the clock cycle consumption depends on objects' sizes, positions, and also attributes in the 3D application. For example, triangles are usually of different sizes, which indicate the different numbers of fragments they contain, and this will eventually lead to difficulty in evaluating the clock cycle consumption. Moreover, in the Clipper, primitives will firstly be classified into three types, and operations will be different according to these types. The first type is the primitive that is totally inside the view frustum, and such a primitive will not be clipped. The second type is the primitive that is totally outside the view frustum, and such a primitive will be discarded. The final type is the primitive that is partially inside the view frustum, and such a primitive will be clipped. Different types have different associated clock cycle consumptions.

Now that the clock cycles used in tiling are dependent on the size of primitives and the overlap conditions between the primitives and tiles, the performance of the TC is measured by the number of triangles tiled per second. With a frequency of 200 MHz, the TC can process 5.7 million triangles per second in the worst case and 18.4 million in the best case.

Moreover, a 0.18 μm CMOS technology library is applied to analyze the timing and power consumption in the worst case. The results of the PVT corner, area, and power are demonstrated in Table 3.

Comparisons have been made as shown in Table 4. The performance of the UPE in this work is compared with that of other designs. The area and energy performance have been improved by varying degrees as compared with conventional works. The UPE also provides a floating-point capacity of 1 GFLOPS.

4. Conclusion

In this paper, a tile-based EGPU is proposed with fused, scalable, and hierarchical parallelism architecture. A UPE is proposed to improve the computing efficiency via SIMT architecture, a coissue mechanism, and high parallelism. Different modes are provided to meet the demands of different scenes with the help of the GCC, leading to a wider range of adaptability. Additionally, the tile-based mechanism dramatically reduces the consumption of memory bandwidth by tiling the screen and using on-chip buffers. The design of the MH avoids memory access conflicts efficiently. With the innovation of the architecture and processing mechanism, the proposed EGPU has achieved a proper balance between performance and hardware costs. The UPE has the ability to address nearly 100 million vertices or fragments per second at a 200 MHz clock frequency. Additionally, the TC can process 18.4 million triangles in its best case in the tiling mode. Additionally, a 0.18 μm technology library is used for timing and power analysis. The area of the EGPU is approximately 6.5 mm * 6.5 mm, and the power consumption is approximately 349.318 mW. The entire EGPU can complete graphics processing tasks successfully. Moreover, it is possible to further improve the current design. For example, the TC cannot operate at full speed because of the speed limitations of the Clipper and the PA. Simultaneously, the GCC can also be improved to match the UPE in performance and processing quality. The next stage of research will be focused on ways to enhance the efficiency of the UPE via a more advanced architecture.

Conflict of Interests

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

Acknowledgments

This work was supported by the Natural Science Foundation of Shandong Province (ZR2013FQ006), the Key Innovation Project of Shandong Province (2013CXB3020),

TABLE 4: Compared results and conditions.

Architecture	[11]	[12]	[13]	This work
Feature	Vertex processor	Unified shader	Graphics processor	UPE
Acceleration range	Geometry	Geometry, rendering	Geometry, rendering	GP-GPU, Geometry, rendering
Performance	186 M vertices/s	9.1 M vertices/s 100 M fragments/s	50 M vertices/s 50 M fragments/s	1 GFLOPS 100 M vertices/s 100 M fragments/s
Processing technology	0.18 μm CMOS	0.13 μm CMOS	0.18 μm CMOS	0.18 μm CMOS
Frequency	110 MHz	100 MHz	200 MHz	200 MHz
Power consumption	160.7 mW	195 mW	155 mW	198.228 mW
Area	4 mm * 4.8 mm	3.3 mm * 3 mm	22 mm ²	~27 mm ²

the State Key Laboratory of Digital Multimedia Technology (2013-1-2569), Shandong Post-Doctor Innovation Foundation Grant (201002029), Grant of the China Post-Doctor Innovation Foundation (20110491601), and National Natural Science Foundation of China (NSFC) Grant (61302063). The authors would like to thank all research partners for their significant contributions to this work. The authors also thank University of Jinan for their support of the hardware platform.

References

- [1] L. Garber, "GPUs go mobile," *Computer*, vol. 46, no. 2, Article ID 6457381, pp. 16–19, 2013.
- [2] Imagination Technologies, *PowerVR MBX Technology Overview*, Revision 1, Imagination Technologies, 2009.
- [3] C. M. Wittenbrink, E. Kilgariff, and A. Prabhu, "Fermi GF100 GPU architecture," *IEEE Micro*, vol. 31, no. 2, pp. 50–59, 2011.
- [4] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, and J. C. Phillips, "GPU computing," *Proceedings of the IEEE*, vol. 96, no. 5, pp. 879–899, 2008.
- [5] S. W. Keckler, W. J. Dally, B. Khailany, M. Garland, and D. Glasco, "GPUs and the future of parallel computing," *IEEE Micro*, vol. 31, no. 5, pp. 7–17, 2011.
- [6] E. Lindholm, J. Nickolls, S. Oberman, and J. Montrym, "NVIDIA Tesla: a unified graphics and computing architecture," *IEEE Micro*, vol. 28, no. 2, pp. 39–55, 2008.
- [7] B.-G. Nam, H. Kim, and H.-J. Yoo, "A low-power unified arithmetic unit for programmable handheld 3-D graphics systems," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 8, pp. 1767–1778, 2007.
- [8] Imagination Technologies, *POWERVR Series5 Graphics SGX Architecture Guide for Developers*, Version 1.0.8, Imagination Technologies, 2011.
- [9] M. Pharr and R. Fernando, *GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation*, Addison-Wesley, New York, NY, USA, 2005.
- [10] H.-Y. Kim, Y.-J. Kim, J.-H. Oh, and L.-S. Kim, "A reconfigurable SIMT processor for mobile ray tracing with contention reduction in shared memory," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 4, pp. 938–950, 2013.
- [11] C.-H. Yu, K. Chung, D. Kim, S.-H. Kim, and L.-S. Kim, "A 186-Mvertices/s 161-mW floating-point vertex processor with optimized datapath and vertex caches," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 10, pp. 1369–1382, 2009.
- [12] J. H. Woo, J. H. Sohn, H. Kim, and H. J. Yoo, "A 195 mW/152 mW mobile multimedia SoC with fully programmable 3-D graphics and MPEG4/H.264/JPEG," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 9, pp. 2047–2056, 2008.
- [13] J.-H. Sohn, J.-H. Woo, M.-W. Lee, H.-J. Kim, R. Woo, and H.-J. Yoo, "A 155-mW 50-m vertices/s graphics processor with fixed-point programmable vertex shader for mobile applications," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 5, pp. 1081–1091, 2006.
- [14] Y.-J. Kim, H.-E. Kim, S.-H. Kim, J.-S. Park, S. Paek, and L.-S. Kim, "Homogeneous stream processors with embedded special function units for high-utilization programmable shaders," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 9, pp. 1691–1704, 2012.

Research Article

Facial Feature Extraction Using Frequency Map Series in PCNN

Rencan Nie, Dongming Zhou, Min He, Xin Jin, and Jiefu Yu

College of Information, Yunnan University, Kunming 650091, China

Correspondence should be addressed to Dongming Zhou; zhoudm@ynu.edu.cn

Received 20 March 2015; Accepted 14 May 2015

Academic Editor: Gwanggil Jeon

Copyright © 2016 Rencan Nie 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.

Pulse coupled neural network (PCNN) has been widely used in image processing. The 3D binary map series (BMS) generated by PCNN effectively describes image feature information such as edges and regional distribution, so BMS can be treated as the basis of extracting 1D oscillation time series (OTS) for an image. However, the traditional methods using BMS did not consider the correlation of the binary sequence in BMS and the space structure for every map. By further processing for BMS, a novel facial feature extraction method is proposed. Firstly, consider the correlation among maps in BMS; a method is put forward to transform BMS into frequency map series (FMS), and the method lessens the influence of noncontinuous feature regions in binary images on OTS-BMS. Then, by computing the 2D entropy for every map in FMS, the 3D FMS is transformed into 1D OTS (OTS-FMS), which has good geometry invariance for the facial image, and contains the space structure information of the image. Finally, by analyzing the OTS-FMS, the standard Euclidean distance is used to measure the distances for OTS-FMS. Experimental results verify the effectiveness of OTS-FMS in facial recognition, and it shows better recognition performance than other feature extraction methods.

1. Introduction

Face recognition is an important research field of pattern recognition; it has good potential applications in biological recognition technology, security system, and so on. At present, there are many face images through sensors, so we need a good algorithm to deal with these images. In the process, because of face data space caused by the problem of dimension disaster, facial feature extraction method with spatial dimension reduction effect is becoming the key technology of face recognition. In the past several decades, many researchers proposed a lot of methods to extract facial features such as geometric characteristics, subspace analysis [1–5], and neural network method [6].

The method of geometric characteristics [7] uses the calculation of geometric parameters as the face features; it has good adaptability to illumination changes, but poor adaptability to the more obvious facial expression, posture, and rotation changes, and so forth. At present, the subspace analytical method is a popular face recognition method, it employs a transform method of linear or nonlinear to make the data in the cast shadow space embodying explicit feature pattern, so as to extract the key features such as

the method of PCA [1], LDA [2], and ICA [3] based on linear transformation and the method of KPCA [4], KFDA [5], and KICA based on nonlinear transformation. However, these methods have poor adaptability to the changes of rotation and distortion in the image. Neural network is based on the nonlinear transform ability of the network structure and uses the learning of the training sample to get nonlinear transforming space of the data and then to obtain the facial features according to the nonlinear transforming space, for instance, using SOM neural network and fuzzy RBF neural network, and so forth. But neural network method will likely cause over-fitting in the learning process for the samples of empirical risk minimization principle. Besides, the subspace analysis method and the face feature extraction method of traditional neural network need face samples to learn; if the training samples are changed, the projection transformation space of the data also wants to change; thus, the calculated amount in large-scale human face feature extraction is too large; its application in the real-time demand higher occasion is limited.

In 1993, Johnson and Ritter proposed pulse coupled neural network (PCNN) [8] based on Eckhorn research in cat's visual cortex. It has widely been used in image

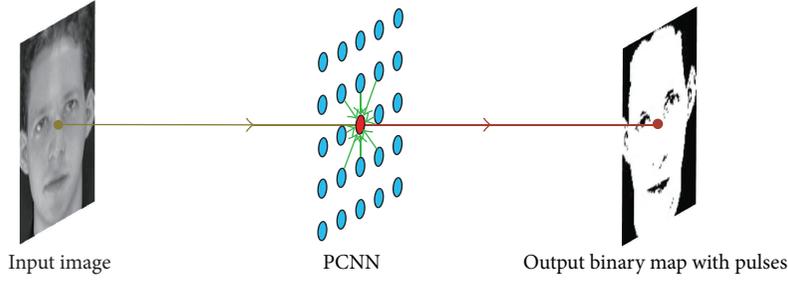


FIGURE 1: The structure of image processing using PCNN.

segmentation [9, 10], image fusion [11–13], image retrieval [14], and so forth. In PCNN, the similarity group neurons will issue synchronous pulses under the effect of mutual coupling pulses. These pulses information constitute a 3D binary map series (BMS), which effectively describes the information of the edge and regional distribution of the image. But the data size of the BMS is larger and cannot be directly used as the image features; for this reason, by calculating the area of the binary image, Johnson [15] transform BMS into 1D oscillation time sequences (OTS), and the OTS has a good invariance in geometric transforming such as rotation, translation, and zoom. Based on the pulse mechanism of the neuron in PCNN, the pulses are divided into the capture pulses and self-excitation pulses, and the OTS is divided into C-OTS and S-OTS, and serve as the facial features extraction in face recognition [16]. Reference [14] extracted 1D OTS of the BMS by calculating the normalization rotational inertia (NRI) of the binary image and applying it to the image retrieval. However, the OTS feature extraction method of the image based on BMS did not fully consider the correlation between BMS in the binary images; those discontinuous features regions will cause influence for the pattern classification capability of OTS features. In addition, the OTS of Johnson's form is statistical characteristics in the sense of whole situation of the binary images; these did not consider the spatial structure of the image, and the spatial structure information often plays an important role in pattern classification.

In view of the above analysis, this paper proposed a novel face feature extraction method of OTS based on the BMS of PCNN output, and, compared with the traditional subspace analysis and neural network method, the results will not change with the sample space change.

2. Pulse Coupled Neural Network

The PCNN model consists of the receptive field, the modulation field, and the pulse generator. In the receptive field, the neuron, respectively, receives the coupling pulse input Y and external stimulus input S of neighboring neurons and consists of L and F channels, which is described by (1). In L and F channels of the neuron, the neuron links with its neighborhood neurons via the synaptic linking weights W and M , respectively; the two channels accumulate input and exponential decay changes at the same time; the decay

exponentials of L and F channels are α^L and α^F , while the channel amplitudes are V^L and V^F :

$$\begin{aligned} F_{ij}(n) &= V^F \sum_{kl} M_{ijkl} Y_{kl}(n-1) + e^{-\alpha^F} F_{ij}(n-1) + S_{ij}, \\ L_{ij}(n) &= V^L \sum_{kl} W_{ijkl} Y_{kl}(n-1) + L_{ij}(n-1) e^{-\alpha^L}. \end{aligned} \quad (1)$$

In the modulation field, the linking input $L_{ij}(n)$ is added a constant positive bias; then, it is multiplied by the feeding input; the bias is unitary, β is the linking strength, and the total internal activity U is the result of modulation, which is described by

$$U_{ij}(n) = F_{ij}(n) [1 + \beta L_{ij}(n)]. \quad (2)$$

Pulse generator consists of a threshold adjuster, a comparison organ, and a pulse generator. Its function is to generate the pulse output Y , and it adjusts threshold value θ_{ij} ; V_{ij}^θ is threshold range coefficient, which is described by (3). When the internal state U is larger than the threshold θ , the neuron generates a pulse, which is described by

$$\theta_{ij}(n) = e^{-\alpha^\theta} \theta_{ij}(n-1) + V_{ij}^\theta Y_{ij}(n-1), \quad (3)$$

$$Y_{ij}(n) = \begin{cases} 1, & U_{ij}(n) > \theta_{ij}(n) \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

In the above equations, the subscripts i and j denote the neuron location in a PCNN and n denotes the current iteration (discrete time step), where n varies from 1 to N (N is the total number of iterations).

The PCNN used for image processing is a single layer two-dimensional array of laterally linked pulse coupled neurons as shown in Figure 1, and all neurons are identical. The number of neurons in the network is equal to the number of pixels in the input image. There exists one-to-one correspondence between the image pixels and network neurons, and the gray value of a pixel is taken as the external input stimulus of the neuron in F channel; namely, $S_{ij} = I(i, j)$. The output of each neuron results in two states, namely, pulse (1 state) and nonpulse (0 state), so the output states of neurons comprise a binary map.

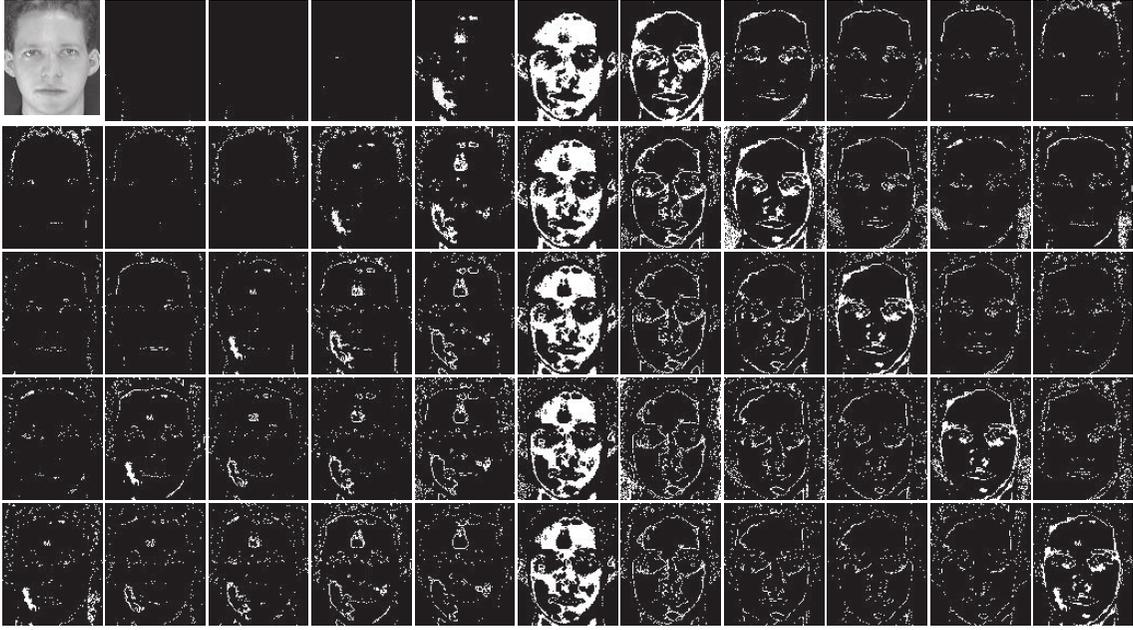


FIGURE 2: An example of binary map series (BMS).

3. Face Feature Extraction Using PCNN

In PCNN, each neuron is connected with neighboring neurons in linking range, so a neuron will receive some pulse inputs from its adjacent neurons. If adjacent neurons have a similar external stimulus with the current neuron, the neuron will issue pulse because of pulse coupled action; that is to say, the neuron and its similar adjacent neurons will emit synchronous pulses. So the similar group of the PCNN neurons possesses the pulse synchronization characteristic and the characteristic benefit features clustering for an image.

3.1. Feature Extraction Based on Binary Map Series of PCNN.

The PCNN will output a binary map in each iteration, so a binary map series (BMS) which contains N binary maps will be generated after N iterations, which is recorded $Y = \langle Y(1), Y(2), \dots, Y(N) \rangle$ and is shown in Figure 2.

Figure 2 shows a face image in the upper left corner, and others are the face image of BMS generated by PCNN. The results in Figure 2 show that the BMS effectively reflects the edge details and regional features distribution, and on the time axis of the image sequence, it well demonstrates that the process produces changes of the image features by neighborhood neurons pulse coupling. In addition, some characteristics of the regional component cycle are repeated with certain cycle in BMS, but some feature components are contracting on the time axis, while others are expanding. The phenomenon actually shows that the PCNN operates the features region clustering in smooth form.

In PCNN, an image of size $m \times n$ will generate a BMS with the size $m \times n \times N$; this means the data of the BMS is N times greater than that of the original image, so the BMS cannot be directly applied to pattern classification. If each 2D binary map in BMS could be translated into 0D data points by some

means or another, then the 3D BMS could be translated into a 1D oscillation time sequence (OTS) of the size $1 \times N$; thus, the OTS realizes the feature extraction and data dimension reduction of the image. Because this kind of OTS is generated by BMS, it can be denoted as OTS-BMS. Johnson [15] put forward a method of translating a binary image into 0D data points by calculating the area of the binary map; thus, this kind of OTS-BMS can be described by (5); an example is shown in Figure 3:

$$BS(n) = \sum Y(n). \quad (5)$$

3.2. Frequency Map Series and Feature Extraction. OTS-BMS defined by (5) is a global statistical result of each binary image, so in a binary map, those pulses (1 state) at different location and not in same feature region will play the same role in the feature description with OTS-BMS. Thus, this kind of OTS-BMS not only failed to properly describe the correlation among binary maps in chronological order but also cannot well describe the spatial structure information of a binary map, but the information is very important for improving the effectiveness of pattern classification using BMS. In view of this, we define a frequency map series (FMS) based on the above BMS, which is described by (6), and an example is shown in Figure 4:

$$FS = \langle F(1), F(2), \dots, F(N) \rangle, \quad (6)$$

$$F(n) = \sum_{k=1}^n Y(k). \quad (7)$$

In FMS, if each 2D frequency map can be transformed into 0D feature points, then 1D time sequence signature OTS-FMS can be extracted from 3D FMS. In addition, to

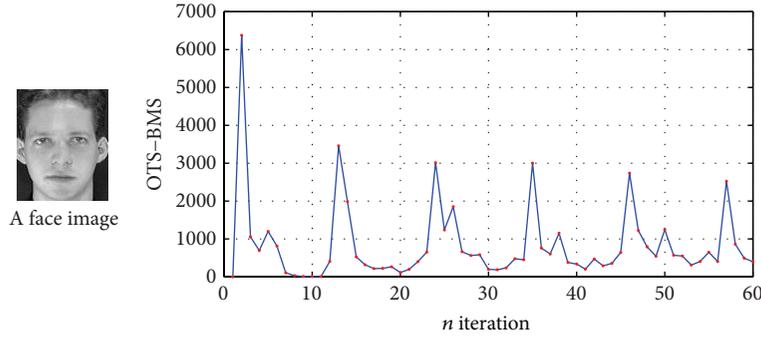


FIGURE 3: An example of OTS-BMS for a face image.



FIGURE 4: An example of FMS for a face image.

make the 0D feature points as possible as including the space structure information of a frequency map, here we use two-dimensional entropy of an image to extract 0D feature information for a frequency map. The two-dimensional entropy of an image is defined by (8):

$$H = - \sum_{i=0}^Q \sum_{j=0}^Q P_{ij} \log P_{ij}, \quad (8)$$

$$P_{ij} = \frac{f(i, j)}{m \times n}. \quad (9)$$

In (8), Q denotes gray levels of the image. In (9), i represents pixel gray value, and j represents k neighbor mean value of this pixel. Obviously, two-dimensional entropy not only transforms the image of the size $m \times n$ into 0D data points, because of P_{ij} reflecting the relation between pixel gray value and neighbor mean value, but also contains the spatial structure information of the image. Through calculating

the two-dimensional entropy of the image, the FMS of the size $m \times n \times (N - n_s + 1)$ can be transformed into 1D OTS-FMS of the size $1 \times (N - n_s + 1)$; that is, $FS = \langle H(1), H(2), \dots, H(N - N_s + 1) \rangle$; as a result, this method greatly reduces the feature dimension. In addition, OTS-FMS also has good ability of pattern classification for the face image; as shown in Figure 5, OTS-FMS is supremely similar with the same kind of face S1_1 and S1_7 and is different for S3_9.

Because the calculation of the two-dimensional entropy of an image is independent of the rotation and translation of the image and has good adaptability to image scaling, therefore, this kind of OTS-FMS has good invariance for the image rotation and translation and has invariance for the image scaling with small error. As shown in Figure 6, the face image is rotated 30 degrees, zoomed 0.6, and moved 15 pixels vertically and horizontally, respectively. Obviously, the image occurs geometric changes, but its OTS-FMS characteristic curves have little difference.

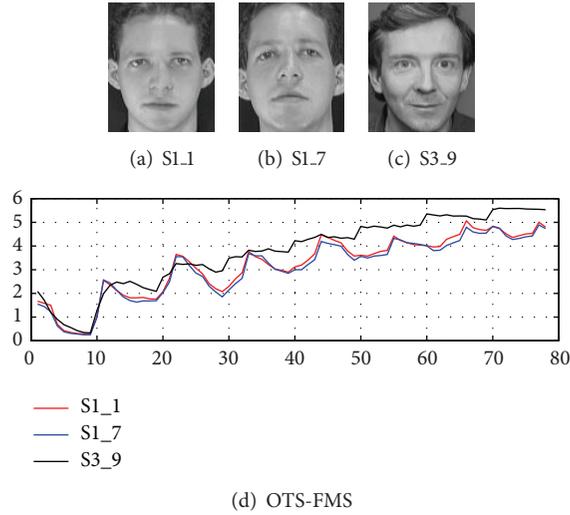


FIGURE 5: An example of pattern classification for OTS-FMS.

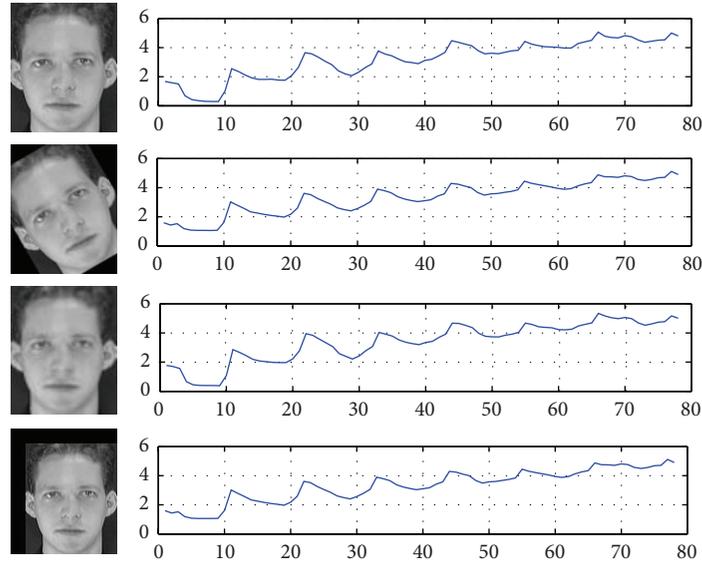


FIGURE 6: An example of geometrical invariability for OTS-FMS.

4. Distance Measure Selection of OTS-FMS

After feature extraction and dimensionality reduction of the face image, we can use common Euclidean distance (ED) as distance measure in pattern classification for the face recognition. But ED requires that the distribution of each dimension component of data should be consistent. For the OTS-BMS feature data of an image with the size of $m \times n$, because it is global pulse statistics in $[0, 1]$ data space, the mapping space of each dimension component is $[0, m \times n]$, so ED can be used to measure the distance among OTS-BMS.

However, for OTS-FMS feature data of the image, because it is monotone increasing totally, as shown in Figure 5, namely, the frequency rank of each dimension component of data also is increasing in the overall. Therefore, even if we use the same entropy method to calculate the frequency map,

mapping spaces of the frequency maps are inconsistent; that is to say, their operation results of two-dimensional entropy are not consistent in their dimension. So the distance measure of OTS-FMS feature data is inappropriate to select ED, and it should select a distance measure which is independent of dimension. Here, standardized Euclidean distance (SED) is a suitable choice. For eigenvectors X and Y , V denotes the standard variance, and then SED is defined by

$$D_{XY} = (X - Y) V^{-1} (X - Y)^T. \quad (10)$$

5. Experimental Results and Analysis

To evaluate the performance of the proposed facial feature extraction in face recognition, experimental studies are carried out on the ORL face database or MIT-CBCL face

TABLE 1: Average recognition rate (%) of OTS-FMS or OTS-BMS.

n	OTS-FMS		OTS-BMS	
	ED	SED	ED	SED
1	52.26	63.40	57.38	57.1
2	70.03	79.83	74.78	74.34
3	78.85	87.04	83.51	82.76
4	85.24	92.67	90.17	89.42
5	88.72	95.27	93.05	92.3
6	92.33	97.46	95.69	95.42
7	93.25	98.19	96.97	96.42
8	94.62	98.75	97.79	97.46
9	94.75	99.08	98.50	98.00

database. The ORL database contains 400 images of 40 individuals (4 females and 36 males). Each individual has 10 images varying in position, rotation, zoom, and expression. The MIT-CBCL database comprises 2000 images from 200 people; each person has 10 images with different illumination, pose, and rotation. The parameters of PCNN are set as $W = [0.5, 10.5; 1, 0, 1; 0.5, 10.5]$, $V^L = 1$, $\alpha^L = 1.4$, $V^F = 0.001$, $\alpha^F = 0.1$, $\beta = 0.02$, $\alpha^\theta = 0.1$, $V^\theta = 12.5$, $\theta(0) = 1.2$, the default iterations $N = 200$, and the default statistics starting point $n_s = 10$.

First of all, to test the classification performance of the OTS-BMS or OTS-FMS when using ED and SED, we constructed experiments based on ORL database by randomly leaving out n image per person each time for testing, the remainder being $10 - n$ images per person for training. This was repeated 30 times by leaving out n images per person each time. The experimental results listed in Table 1 are the average of 30 times' results each time.

It can be seen from Table 1, under different training sets, that the recognition rate of SED measure is significantly higher than ED measure for OTS-FMS, and the recognition rate of ED measure is also slightly higher than SED measure for OTS-BMS. This shows that the OTS-FMS by calculating two-dimensional entropy of each frequency map in FMS is suitable for using SED measure, while the OTS-BMS by calculating the area of each binary map in BMS is suitable for selecting ED measure. In addition, the recognition rate of OTS-FMS under SED measure is significantly higher than OTS-BMS under ED measure. That shows the proposed OTS-FMS in this paper can effectively extract facial image features.

Then, to investigate the influences on the facial recognition performance for different iterations N in PCNN or the different statistical starting point n_s in FMS, experiments were carried out on ORL face database, by randomly choosing 6 images per person each time for testing, the remainder being 4 images per person for training. This was repeated 30 times with different N and n_s each time. The average face recognition rates in these experiments are shown in Table 2.

In Table 2, we can see that the recognition rate firstly increases with the increase of iterations N under definite n_s , but soon afterwards it is gradually stable; hence, N can be set as $N \in [80, 150]$ to decrease the amount of calculation for PCNN. Under definite N , the recognition rate decreases with

TABLE 2: Average recognition rate (%) with different N and n_s .

n_s/N	30	50	80	120	150	200	300	450
1	95.60	97.02	96.38	97.08	96.75	96.63	96.56	96.52
3	95.50	96.63	95.88	97.13	97.52	97.13	96.04	97.06
5	95.52	96.88	97.52	97.08	97.02	97.25	96.90	96.77
8	95.29	97.31	96.35	95.90	97.00	97.21	97.02	96.85
10	95.75	96.83	97.04	96.48	96.85	96.46	97.33	96.90
20	77.21	93.02	95.31	94.63	96.73	96.27	95.77	95.81
40	0	69.71	91.69	93.40	94.67	94.19	93.50	93.98
70	0	0	64.21	89.06	91.38	91.94	92.31	92.90

TABLE 3: Average recognition rate (%) under different methods.

n	PCA	2DPCA	KPCA	PCNN [15]	In this paper
4	88.33 (61)	89.58 (14)	87.50 (62)	78.33	80.42
5	90 (70)	91.00 (14)	89.0 (70)	92.00	95.50
6	93.75 (80)	95.00 (14)	96.25 (80)	97.50	98.75
7	95.00 (88)	94.17 (14)	95.83 (89)	96.67	99.17
8	95.00 (96)	96.25 (14)	96.25 (96)	96.25	98.75
9	95.00 (103)	92.50 (14)	95.00 (103)	100	100

the increase of n_s , and to increase the recognition rate, n_s is set as $n_s \in [3, 5]$.

On the ORL database, selecting the frontal n images of each person as training set, and the rest of images as test set, we make a comparison of the recognition rate of our method (OTS-FMS + SED) with other methods such as common subspace face recognition algorithms (PCA, 2DPCA, and KPCA) and PCNN [16], as shown in Table 3. Obviously, under the condition $n \geq 5$, the recognition rate of the proposed method is significantly higher than other methods, but under the condition $n = 4$, the recognition rate in this paper is lower than other subspace methods, but it is higher than PCNN method [16]. Combined with the experimental results of Table 1 ($n = 4$, the average recognition rate is 92.67%), we can know that there are a few influential samples in the training sets.

In order to verify the validity of this method for face recognition in the bigger data space, on MIT-CBCL, we randomly select face images of each person at rate p to compose training set, the remaining images as test set; then, 30 times experiments were carried out, and the average recognition rate is shown in Table 4.

As shown in Table 4, under the same training set, the recognition rate of OTS-FMS in this paper is higher than [16] based on BMS, which illustrates the validity of this paper. In addition, the experimental results in MIT-CBCL Library are the same as ORL Library; the recognition rate using SED measure is higher than that using ED measure. This again shows the correctness of the analysis of the characteristics of the OTS-FMS in this paper.

On ORL or MIT-CBCL face database above, the face has been tailored manually from original image, but on a practical face recognition system, obtaining the face from

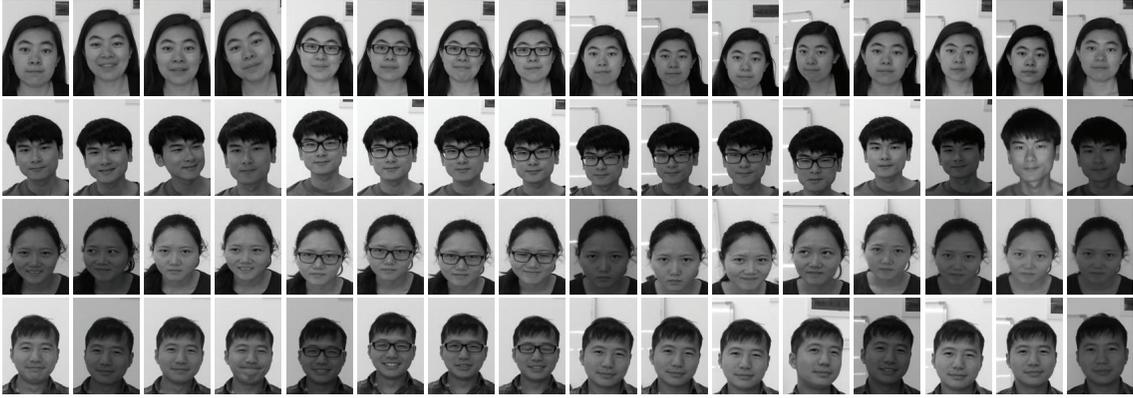


FIGURE 7: Part of face images on the custom face database.

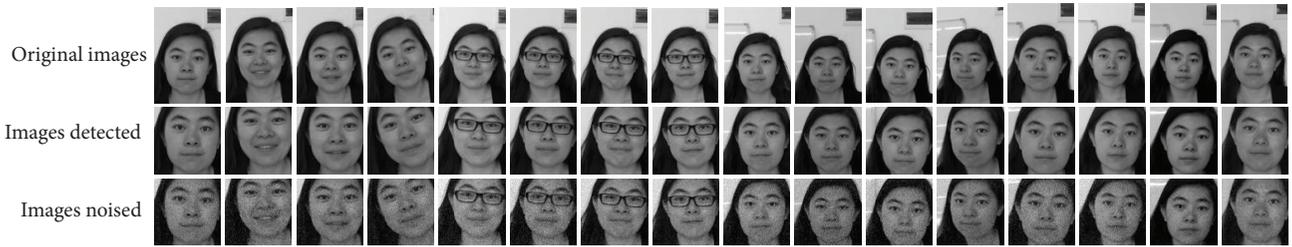


FIGURE 8: Face detection and noise sample.

TABLE 4: Average recognition rate (%) of different methods on MT-CBCL.

p	OTS-FMS		PCNN [15]
	ED measure	SED measure	
0.1	91.61	92.48	91.30
0.2	95.65	96.16	95.14
0.3	96.98	97.40	96.70
0.4	97.72	98.01	97.57
0.5	98.36	98.65	98.05
0.6	98.49	98.70	98.39
0.7	98.86	99.01	98.60
0.8	98.84	99.10	98.76
0.9	98.92	99.15	98.82

an image needs to use a certain face detection algorithm such as the method in [17], and the facial recognition performance is easily affected by the detection accuracy. To test the adaptability of both OTS-FMS and OTS-BMS for face detection precision, we set up a custom face database, which includes 208 images of 13 individuals; each individual has 16 images varying in facial expression, illumination, pose, background, and facial details (with/without glasses), and the part of the database was shown in Figure 7.

The face has a larger proportion of background on the custom database; first, we use the face detection algorithm [17] for processing (as shown in Figure 8); then, we compare the face recognition performance before and after face detection for both OTS-FMS and OTS-BMS, and the results are

TABLE 5: OTS feature average recognition rate (%) before and after the face detection on custom face database.

p	OTS-FMS		OTS-BMS	
	Without face detection	Face detection	Without face detection	Face detection
0.4	80.54	88.46	78.18	84.38
0.5	88.69	82.05	86.57	89.23
0.6	92.18	94.87	89.74	91.31
0.7	95.08	98.10	92.41	96.31
0.8	97.26	99.34	93.5	98.38

shown in Table 5. Here, experimental repetitions are 30, $n_s = 3$. As we can see in Table 5, after face detection, the average recognition rates are improved for OTS-FMS and OTS-BMS; this suggests that the efficient face detection method has important significance for improving identification performance in the proposed method. In addition, both before and after the face detection, the average recognition rate for the OTS-FMS is better than that for the OTS-BMS; it also suggests that the improvement based on OTS-BMS for the facial feature extraction method (i.e., OTS-FMS) is effective.

In addition, after face detection on custom face database, we add Gaussian noise to each kind of faces detected as shown in Figure 8 with the mean value 0 and the variance $\delta \in [0.0005, 0.003]$; here, the variance satisfies with uniform distribution, and then we investigate the adaptation of both OTS-FMS and OTS-BMS for noise; the results are shown in Table 6. Here, experimental repetitions are 30, $n_s = 3$.

TABLE 6: Average recognition rate (%) after the face detection on custom face database.

p	OTS-FMS	OTS-BMS
0.4	56.03	80.82
0.5	62.08	87.95
0.6	64.62	88.12
0.7	67.77	90.38
0.8	71.92	92.35

As we can see in Tables 5 and 6, after the faces are polluted by noise, both for OTS-FMS and OTS-BMS, the average recognition rate is decreased, but the influence of the noise for the OTS-FMS identification is more serious than OTS-BMS. This is mainly due to the OTS-BMS which is a global statistics on BMS of the facial image, but the OTS-FMS is more emphasized on the facial local detail information. Therefore, when a face is polluted by Gaussian noise, the noise damage to the facial local details is fatal, but to the global statistics it is lighter. The fact shows that the face recognition algorithm (OTS-FMS + SED) is sensitive to noise.

6. Conclusions

In this paper, a novel method was proposed to extract the facial feature based on PCNN. Through the analysis of the limitations of BMS extracted one-dimensional oscillation time sequence (OTS) and considering the correlation between binary images in BMS, we proposed a method to transform BMS into frequency map series (FMS), which reduces the influences of the discontinuous value of binary image on the effectiveness of OTS. Then, the paper considers the method that 2D frequency map is transformed into 0D data points; in order to reflect the spatial structure information in the frequency map, the paper employs the computing method of two-dimensional entropy of the image; it will transform 3D FMS into 1D OTS-FMS feature, and the feature has a good invariance for face image geometric changes. Finally, based on the analysis of characteristics of OTS-FMS data features, the paper we proposed uses the standard Euclidean distance measure as the distance measure of OTS-FMS features. The experimental results show that the recognition rate of OTS-FMS is significantly higher than PCA and KPCA, and so forth, and also better than method [16] based on BMS extracted OTS-BMS features. And compared to traditional subspace analysis and neural network method, face features extracted by this method do not change with the sample space change.

Conflict of Interests

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

Acknowledgments

This research was financially supported by the National Natural Science Foundation of China (nos. 61365001 and 61463052), the Natural Science Foundation of Yunnan Province (no. 2012FD003), and the Science and Technology Plan of Yunnan Province (no. 2014AB016).

References

- [1] G.-C. Luh and C.-Y. Lin, "PCA based immune networks for human face recognition," *Applied Soft Computing*, vol. 11, no. 2, pp. 1743–1752, 2011.
- [2] G.-F. Lu, J. Zou, and Y. Wang, "Incremental complete LDA for face recognition," *Pattern Recognition*, vol. 45, no. 7, pp. 2510–2521, 2012.
- [3] I. Marques and M. Graña, "Face recognition with lattice independent component analysis and extreme learning machines," *Soft Computing*, vol. 16, no. 9, pp. 1525–1537, 2012.
- [4] R. K. Yadav and R. Upadhyay, "Applications expansion in kernel PCA approach for face recognition," *International Journal of Artificial Intelligence and Knowledge Discovery*, vol. 3, no. 4, pp. 1–4, 2013.
- [5] Z. Sun, J. Li, and C. Sun, "Kernel inverse Fisher discriminant analysis for face recognition," *Neurocomputing*, vol. 134, pp. 46–52, 2014.
- [6] J. Haddadnia, K. Faez, and M. Ahmadi, "A fuzzy hybrid learning algorithm for radial basis function neural network with application in human face recognition," *Pattern Recognition*, vol. 36, no. 5, pp. 1187–1202, 2003.
- [7] F. Beverina, G. Palmas, M. Anisetti, and V. Bellandi, "Tracking based face identification: a way to manage occlusions, and illumination, posture and expression changes," in *Proceedings of the 2nd IET International Conference on Intelligent Environments (IE '06)*, pp. 161–166, July 2006.
- [8] J. L. Johnson and D. Ritter, "Observation of periodic waves in a pulse-coupled neural network," *Optics Letters*, vol. 18, no. 15, pp. 1253–1255, 1993.
- [9] C. Gao, D. Zhou, and Y. Guo, "Automatic iterative algorithm for image segmentation using a modified pulse-coupled neural network," *Neurocomputing*, vol. 119, pp. 332–338, 2013.
- [10] J. Li, B. Zou, L. Ding, and X. Gao, "Image segmentation with PCNN model and immune algorithm," *Journal of Computers*, vol. 8, no. 9, pp. 2429–2436, 2013.
- [11] I. De and B. Chanda, "Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure," *Information Fusion*, vol. 14, no. 2, pp. 136–146, 2013.
- [12] J. Lang and Z. Hao, "Novel image fusion method based on adaptive pulse coupled neural network and discrete multi-parameter fractional random transform," *Optics and Lasers in Engineering*, vol. 52, no. 1, pp. 91–98, 2014.
- [13] N. Wang, Y. Ma, K. Zhan, and M. Yuan, "Multimodal medical image fusion framework based on simplified PCNN in non-sampled contourlet transform domain," *Journal of Multimedia*, vol. 8, no. 3, pp. 270–276, 2013.
- [14] Q. Liu, L.-P. Xu, Y.-D. Ma, and Y. Wang, "Image NMI feature extraction and retrieval method based on pulse coupled neural networks," *Acta Automatica Sinica*, vol. 36, no. 7, pp. 931–938, 2010.

- [15] J. L. Johnson, "Pulse-coupled neural nets: translation, rotation, scale, distortion, and intensity signal invariance for images," *Applied Optics*, vol. 33, no. 26, pp. 6239–6253, 1994.
- [16] R. C. Nie, S. W. Yao, and D. M. Zhou, "Face recognition using simplified pulse coupled neural network," *Computer Science*, vol. 41, no. 2, pp. 297–301, 2014.
- [17] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

Research Article

Visibility Enhancement of Scene Images Degraded by Foggy Weather Conditions with Deep Neural Networks

Farhan Hussain and Jechang Jeong

Department of Electronics & Computer Engineering, Hanyang University, Seoul 133-791, Republic of Korea

Correspondence should be addressed to Jechang Jeong; jjeong@hanyang.ac.kr

Received 19 March 2015; Revised 16 May 2015; Accepted 18 May 2015

Academic Editor: Wei Wu

Copyright © 2016 F. Hussain and J. Jeong. 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.

Nowadays many camera-based advanced driver assistance systems (ADAS) have been introduced to assist the drivers and ensure their safety under various driving conditions. One of the problems faced by drivers is the faded scene visibility and lower contrast while driving in foggy conditions. In this paper, we present a novel approach to provide a solution to this problem by employing deep neural networks. We assume that the fog in an image can be mathematically modeled by an unknown complex function and we utilize the deep neural network to approximate the corresponding mathematical model for the fog. The advantages of our technique are as follows: (i) its real-time operation and (ii) being based on minimal input, that is, a single image, and exhibiting robustness/generalization for various unseen image data. Experiments carried out on various synthetic images indicate that our proposed technique has the abilities to approximate the corresponding fog function reasonably and remove it for better visibility and safety.

1. Introduction

Degraded visibility and lack of luminance in foggy weather pose a serious threat to the safety of drivers. These conditions increase the danger of vehicle collision and are a major cause of injuries and fatalities on roads covered with fog. Suspension of very fine droplets in the fog causes blocking and scattering of the light. This leads to less light reaching the driver's eye, lower contrast, and hence reduced visibility (Figure 1). To enhance visibility in bad weather is an area of high interest for the researchers. Various studies have been carried out to observe and model the effects of wide range of weather conditions on vision systems. The authors in [1–4] present such weather models. They analyze how scenes are affected by various weather conditions. These studies attempt to restore various attributes of a scene in an image degraded by bad weather conditions with the help of these weather models. A few researches, for example, [5, 6], suggest the use of polarizing filters to remove the effect of haze from the images. To realize the techniques described in [5, 6], two or more independent images are required to comply with the condition that the air light

produces some measurable partial polarization in them. Weather removal algorithms proposed in [7] make use of a single image and some supplementary information provided by the user. A deep photo system is proposed in [8] based on the georegistration of the photographs. This georegistration system (GIS) data such as 3D models for cities, buildings, terrains, and structures; texture models; and depth maps. The authors of this study suggest the augmentation of this information with simple photographs to achieve various operations like dehazing, relighting, view synthesis, and expanding the field of view. Defogging/dehazing solutions relying entirely on single images are proposed in [9–12]. The single image visibility enhancement method in [9] develops an optimized cost function based on certain observations of images with and without weather effects. This cost function attempts to estimate the direct attenuation of the air light which is applied to the scene for visibility enhancement. In [10] the dark channel prior (DCP) is proposed for single image haze removal. This prior, based on some statistical observations of outdoor haze-free images, is combined with the haze imaging model to recover a haze-free image.



FIGURE 1: (a) Scene visibility without fog. (b) Scene visibility under foggy weather conditions.

The authors in [11] propose another single image dehazing method which is based on [10]. They improve the dehazing performance by modifying the DCP method by introducing a single median filter operation. They also study the effects of dehazing on image and video coding. In [12] restoration of hazy scenes from a single image is proposed by defining the image through a scene transmission model that takes into account the surface shading. A general survey of vision based vehicle detection methods for intelligent driver assistance systems is presented in [13]. Finally in [14, 15] the authors propose enhanced visibility algorithms that take fog effects into account and are particularly well suited for road images.

In this paper, we present a novel approach to improve the visibility of the scene in an image degraded by foggy weather conditions. We target enhanced visibility by generating an approximate model of the fog composition in the scene with a deep neural network [16]. This generalized model is then used for the restoration of scene quality in the image. Our proposed method performs this recovery of image scene in real time and it does not require any additional information. The proposed method is robust as it achieves good result for a large set of unseen foggy images.

2. Defogging by Deep Neural Network

2.1. Artificial Neural Networks. Artificial neural networks [17–19] are complex mathematical systems which tend to electronically simulate the working of a biological nervous system. These networks are made up of a large number of very simple computational units called the artificial neurons. The number of artificial neurons in a network can range from a few hundred to a several thousand artificial neurons. A neural network is formed by interconnecting these hundreds and thousands of artificial neurons in different topologies. One such interconnected network is shown in Figure 2. It is through this interconnection of simple computational elements that the high computational complexity of the neural networks is realized. Artificial neural networks are now being researched to provide solution to various problems like function approximation, regression analysis, time series prediction, classification, pattern recognition, optimization, decision making, data processing, filtering, and clustering

and categorizing, to name but a few. The advent of immensely high CPU processing abilities has made it possible to realize more deep architectures [16, 20] for the neural networks and hence more mathematically complex models can be accomplished. A deep neural network has many hidden layers of neurons between the input layer and the output layer.

2.2. Deep Neural Network for Visibility Enhancement. We present a deep neural network that accepts a foggy image as an input, models the corresponding fog composition in the scene, and produces a defogged version of the scene in the image as an output. The architecture of the proposed deep neural network for image defogging is shown in Figure 2. The network consists of an input layer, an output layer, and n number of hidden layers sandwiched in between the two. The multiple hidden layers of the deep neural network are advantageous in realizing more efficient representation for the corresponding fog function. The learning problem consists of finding the optimal combination of weights so that the network function δ approximates a given function f as closely as possible. The network learns this given function f through some implicit examples. In this paper, the deep neural network is tailored to solve the visibility enhancement problem by training it on several foggy images and their corresponding original images (input-output pairs). In order to learn the generalized fog function and produce the corresponding defogged images the foggy images are divided into M nonoverlapping blocks of size $N \times N$ pixels. Each block is normalized to a range that optimizes the learning and results of the DNN. The DNN is customized for a one-dimensional vector input; therefore, the two-dimensional block is transformed into a one-dimensional vector X by rasterization. This vector X is presented as an input pattern to the network. The weights of the DNN are initialized randomly and the input pattern is forward propagated through the multiple hidden layers of the DNN to generate the propagation output activation. The activation function employed in our DNN is the hyperbolic tangent transfer function which is given by

$$\mathcal{T}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (1)$$

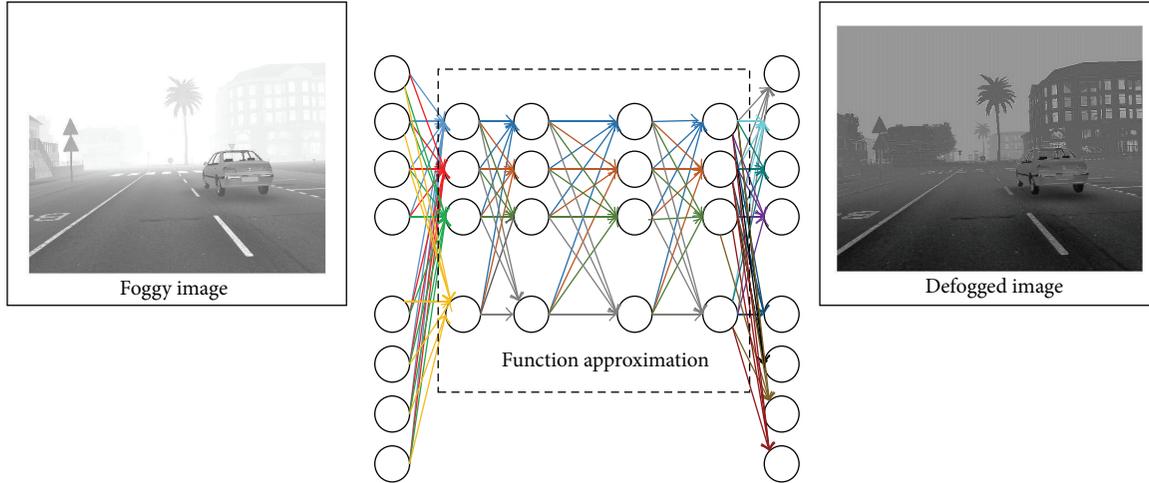


FIGURE 2: Deep neural network architecture for fog removal.

The tangent hyperbolic function produces the scaled output over the -1 to $+1$ closed range. This function shares many properties of the sigmoid function, but because the output space of this tangent function is broader, it may be more efficient for modeling complex nonlinear relations more abstractly. This input pattern X produces an output that is different from the target output (original Image). The error function of the network is the combination of errors contributed by all the hidden nodes of the network and is given as

$$E = \frac{1}{2} \sum_x \|\text{actual output} - \text{desired output}\|^2. \quad (2)$$

The combination of weights which minimizes the error function is considered to be the solution of learning. In this paper, this optimum set of weights is achieved by the backpropagation algorithm. The backpropagation is an iterative gradient descent algorithm in which the output error signals are transmitted backwards from the output layer to each node in the hidden layer that immediately contributed to the output layer. This backward propagation is continued, layer by layer, until each node in the network has received an error signal that describes its relative contribution to the overall error. Once the error signal for each node has been determined, the errors are then used by the nodes to update the values for the weights of the network. This backpropagation of errors and the updating of weights continue until the value of error function becomes sufficiently small. The training of the network is stopped and the new input patterns are presented to the network. The trained network can approximate and remove the fog in these input patterns for better visibility. The flow chart for the visibility enhancement DNN is shown in Figure 3.

3. Experimental Results

Experiments were carried out on images obtained from the FRIDA database [14]. The FRIDA database consists of sets

of synthetic images with and without fog. The evaluation of the proposed method is carried out on gray scale images. The size of each image was 640×480 pixels. The image data set was divided into training set images and the test set images. The learning of deep neural network for defogging purpose was carried out by the training set. Architectures based on multiple hidden layers (5 to 8) with various numbers of hidden nodes (16 to 128) were trained. The convergence of the network was checked on the basis of mean square error (MSE). Once the network learnt the generalized fog function, the results were applied to the test images. For evaluation purpose we show the results in Figures 4–10. In these figures the areas of interest are shown enclosed in rectangles. The proposed solution can be evaluated by analyzing these areas. Figure 4(a) shows the original scene in the image without fog, while Figure 4(b) shows the scene affected by fog. It can be seen clearly that the background in Figure 4(b) is faded and is mostly invisible. The tree enclosed in the rectangle in Figure 4(a) is barely visible in Figure 4(b). The major portion of the building (front) which is visible in Figure 4(a) is completely invisible in Figure 4(b). Also other sections of the building enclosed in rectangles in Figure 4(a) are invisible in Figure 4(b). Figure 4(c) represents the scene recovered by our deep neural network. The building that was invisible in the foggy version is recovered in the defogged version. The tree is visible and other areas enclosed in rectangles in the figure are also partially recovered by the network. Figures 5(a) and 5(b) show the original scene and the fogged scene, respectively. The front building which is clearly visible in the original scene that is Figure 5(a) is faded in the fogged scene, while the building behind it is not visible in the scene of Figure 5(b). The smaller palm tree is also invisible in the foggy version. The house and trees at the left-hand side of the foggy image are almost invisible. The traffic signs and some minor structures are missing in the foggy scene. In Figure 5(c), the fog is removed by the deep neural network. The faded front building has improved visibility in the recovered image. The building behind the front building which was not visible

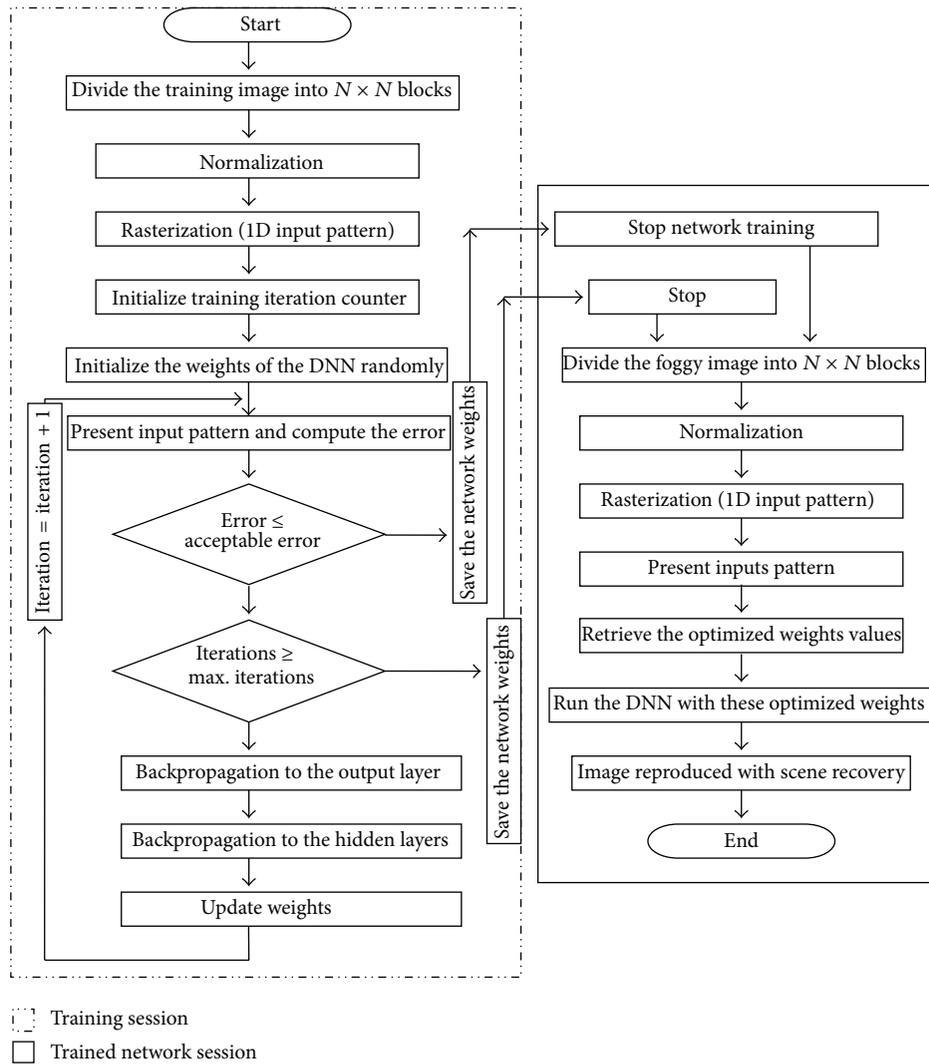


FIGURE 3: Flow chart for proposed visibility enhancement DNN.

previously can be clearly seen in our recovered scene. The smaller palm tree is visible. The house and tree at the left-hand side of the image are also recovered successfully. The traffic signs and the minor structures that were invisible due to fog are also recovered to some extent in the defogged scene. Figure 6(a) shows the original scene, while Figure 6(b) represents the image with degraded visibility due to fog. The trees enclosed in the rectangles on the left-hand side of Figure 6(a) are barely visible in Figure 6(b). The cars enclosed in the rectangles in the original image are invisible in the degraded image. Also the structures/buildings enclosed in rectangles in Figure 6(a) are invisible in the degraded version. The same scene with enhanced visibility is presented in Figure 6(c). As can be seen, the trees which are barely visible in the foggy image have been recovered. The cars which were previously invisible can be spotted in the recovered image. Also the structures/buildings missing in the foggy image are retrieved by our DNN successfully. Figure 7(a) shows the original scene in the image without any fog and Figure 7(b)

represents the scene with reduced visibility due to foggy conditions. Starting from the left-hand side of the image the tree which is visible in the original image is barely visible in the degraded image. The house and the tree adjacent to it are invisible in the degraded image. The palm tree is also missing in the degraded image. The pole, tree, and the rear view of the building on the right-hand side are severely degraded and are invisible in the foggy image. Figure 7(c) represents the scene recovered in the image by the DNN. Starting from the left-hand side the tree which is barely visible has been recovered in the enhanced version. The house and adjacent tree have also been retrieved in the reproduced image, although the details are missing but the structures are visible. The palm tree which was previously invisible is present in the recovered scene. The pole, tree, and rear portion of the building are also recovered and are clearly visible in the reproduced image. Figures 8(a), 8(b), and 8(c) show the original image, the image affected by foggy conditions and the image recovered by our DNN. Compared to Figure 8(a), in Figure 8(b), the background

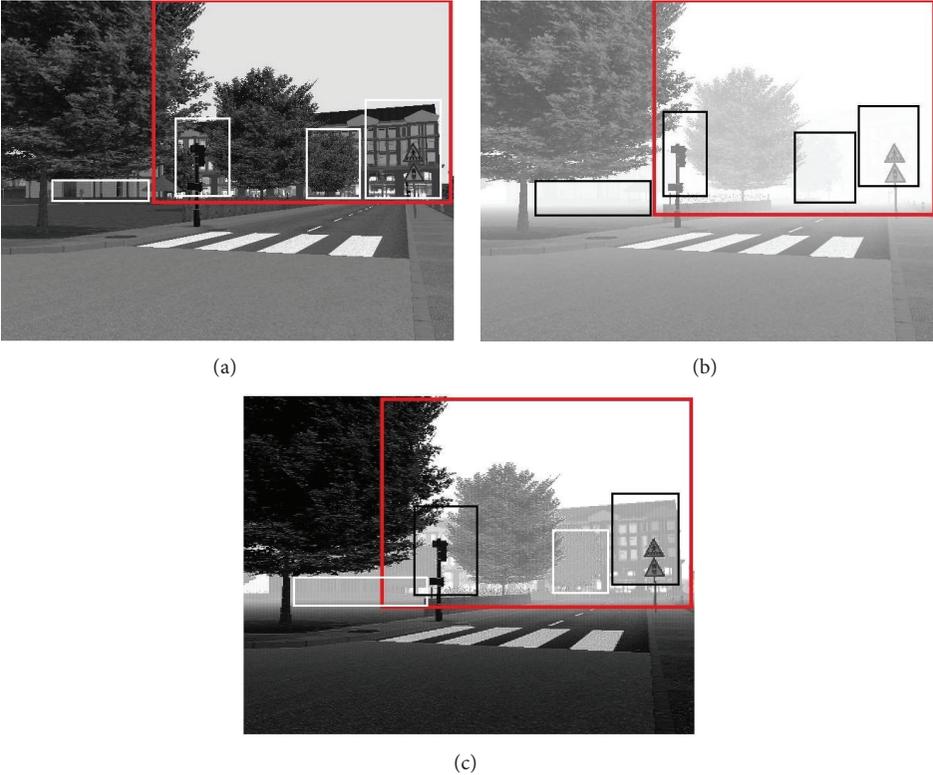


FIGURE 4: (a) Original scene without fog. (b) Scene visibility degraded under fog. (c) Visibility enhancement by deep neural network.

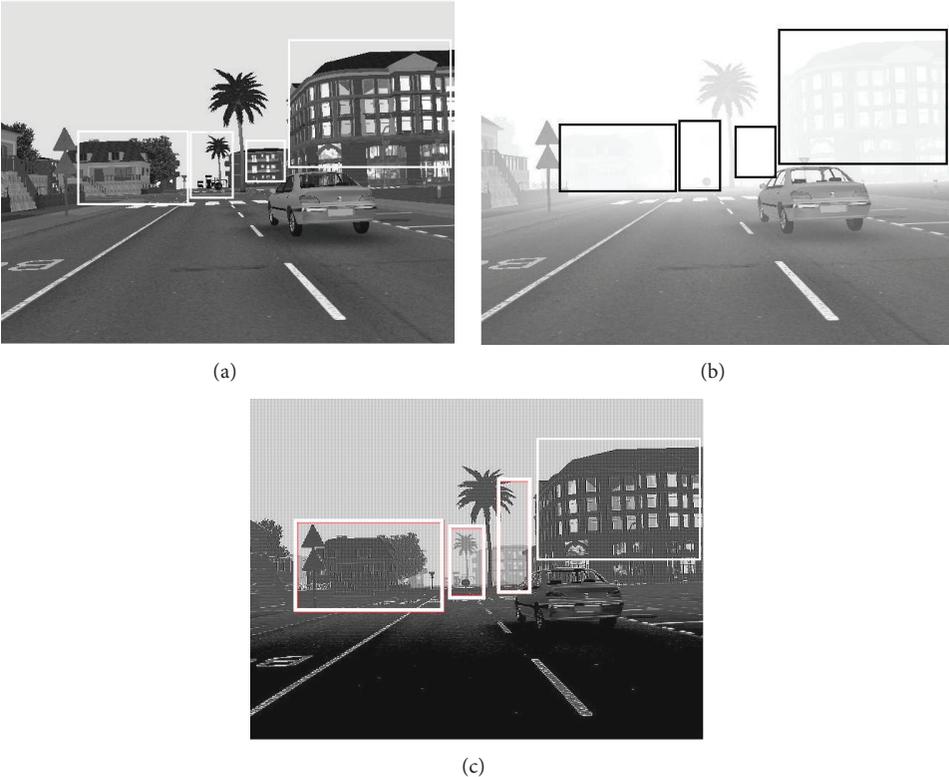


FIGURE 5: (a) Original scene without fog. (b) Scene visibility degraded under fog. (c) Visibility enhancement by deep neural network.

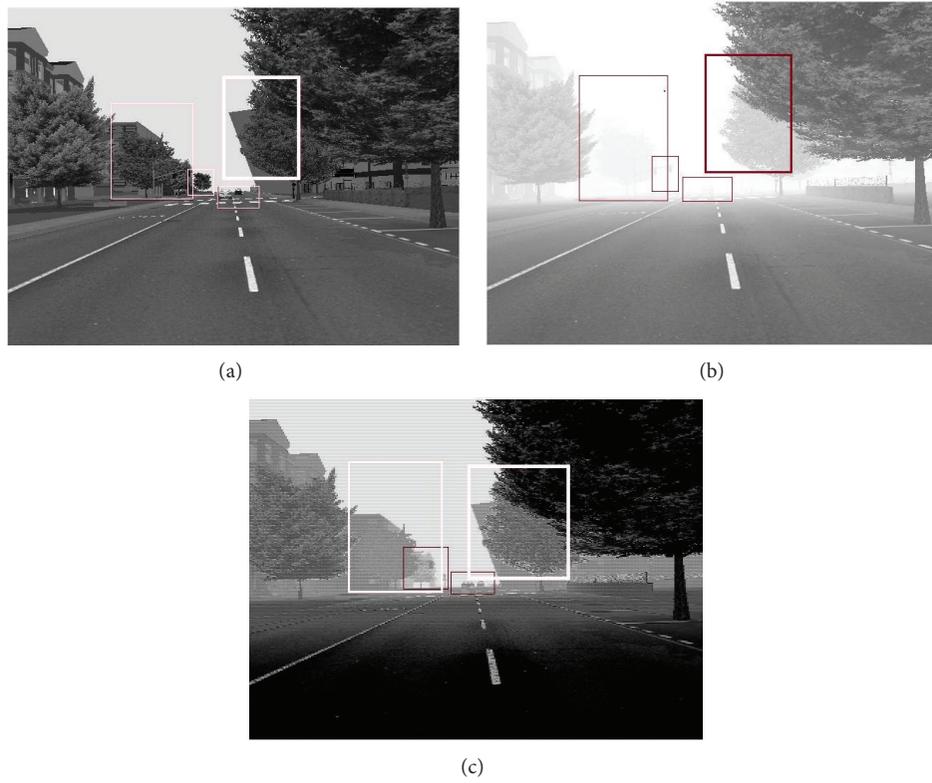


FIGURE 6: (a) Original scene without fog. (b) Scene visibility degraded under fog. (c) Visibility enhancement by deep neural network.

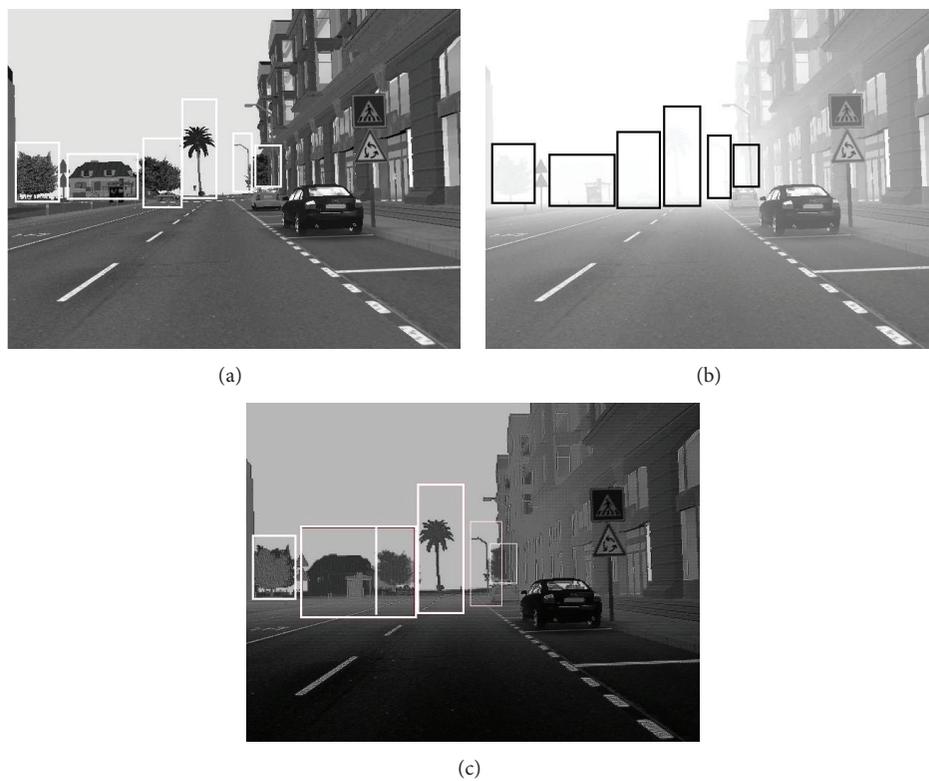


FIGURE 7: (a) Original scene without fog. (b) Scene visibility degraded under fog. (c) Visibility enhancement by deep neural network.

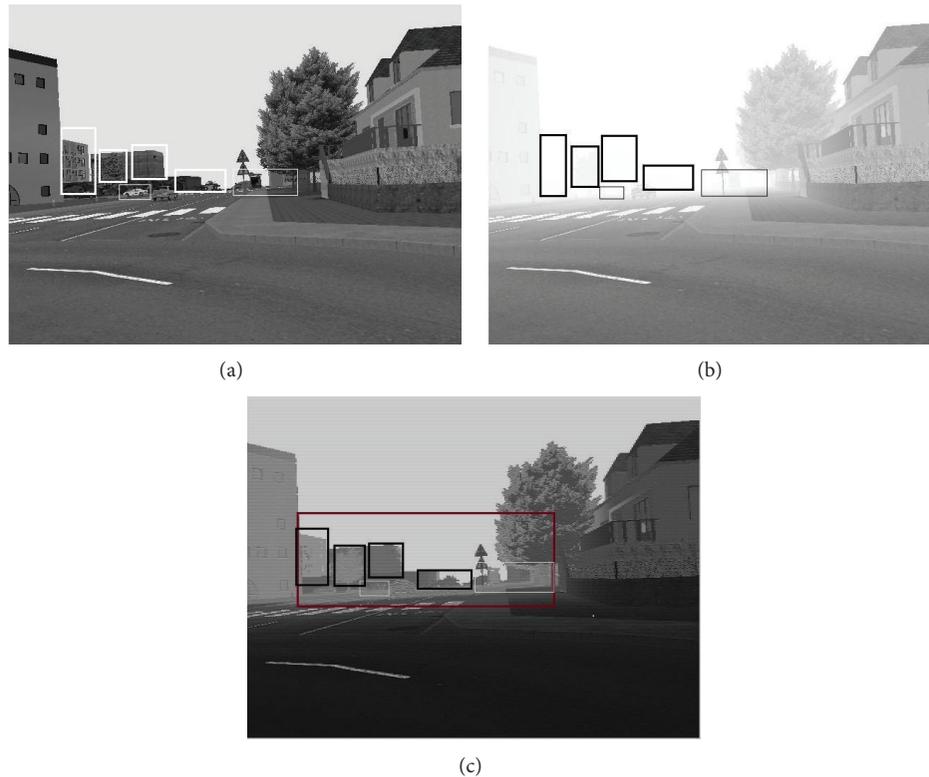


FIGURE 8: (a) Original scene without fog. (b) Scene visibility degraded under fog. (c) Visibility enhancement by deep neural network.

that is buildings, houses, trees, vegetation, and cars enclosed in the rectangular boxes is barely visible or invisible, while in Figure 8(c) our proposed DNN has recovered the above regions successfully. Figures 9(a) and 9(b) represent the original image and its foggy version, respectively. The houses at the end of the lane in the left-hand side of Figure 9(a) are almost invisible in Figure 9(b). The poles and traffic signals are also invisible. The building at the right-hand side is faded and the trees behind it are not visible due to fog. The cars in the scene are barely visible. In our recovered image that is Figure 9(c) the visibility of the scene is enhanced. The houses at the end of the lane are now quite visible. The poles and traffic signals are recovered from the foggy version. The visibility of the building at the right-hand side is improved and the trees behind it are now visible. The cars in the recovered image are more detectable than the foggy version. Figure 10(a) represents the original scene without fog. Figure 10(b) represents the image scene degraded by fog. The building at the right-hand side is faded due to fog and is barely visible. The tree behind the building at the right-hand side is invisible. The traffic sign is invisible in the foggy image, and a large section of buildings and the palm tree is also missing in the left-hand side of the foggy image. In Figure 10(c) that is the scene reproduced by the DNN the barely visible building on the right-hand side of the image is largely visible. The tree behind is also recovered. The traffic sign is also visible in the reproduced image. The rear section of buildings and the tree on left-hand side have also been recovered successfully by our DNN.

4. Conclusions

In this paper, we propose a novel approach for restoring the visibility of foggy images with the help of deep neural networks. This idea is an initial attempt to model the fog function for scene recovery with deep neural networks. The method recovers the scene structure simply with a single image input, operates in real time, and generalizes well. The restored images showed good recovery of the scene from the foggy images. In some of the images the proposed technique showed limitations in reproducing the exact details of some objects in the scene, but, nevertheless, they were fairly recognizable. This method can be tailored to be used in applications designed for driving assistant systems, intelligent vehicle systems, outdoor surveillance systems, and so forth. In future works we are going to use this idea to develop more advanced frameworks for weather removal vision systems based on deep neural networks.

Conflict of Interests

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

Acknowledgment

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Republic of Korea, under

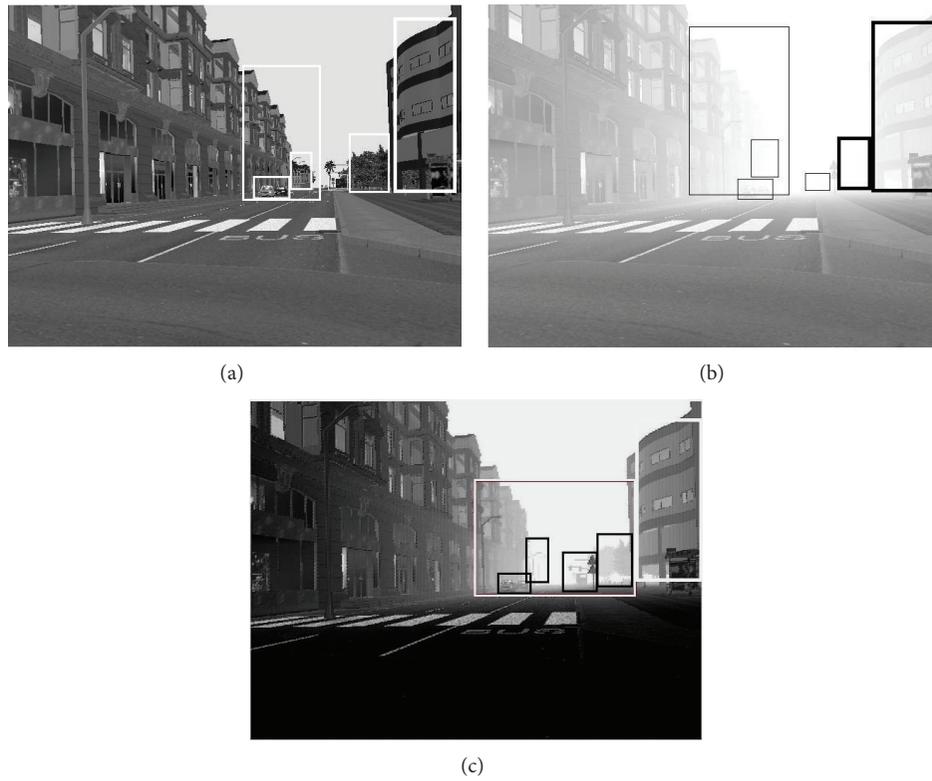


FIGURE 9: (a) Original scene without fog. (b) Scene visibility degraded under fog. (c) Visibility enhancement by deep neural network.

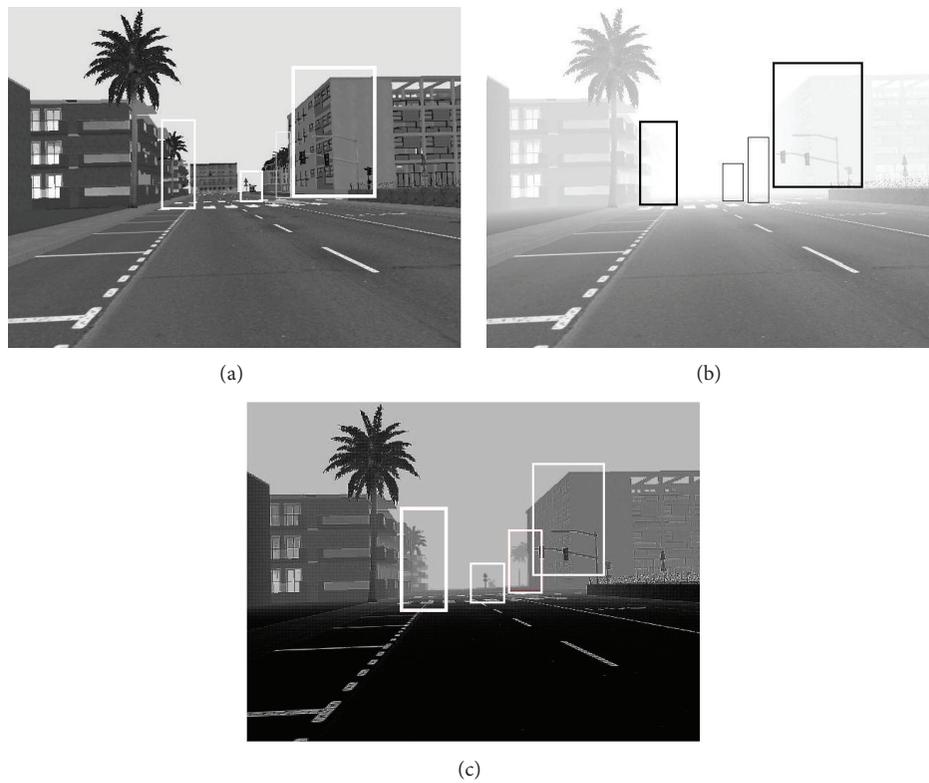


FIGURE 10: (a) Original scene without fog. (b) Scene visibility degraded under fog. (c) Visibility enhancement by deep neural network.

the project for Technical Development of Information Communication and Broadcasting, supervised by IITP (Institute of Information and Communications Technology Promotion) (IITP 2015-B0101-15-1377).

References

- [1] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.
- [2] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR' 2000)*, pp. 598–605, June 2000.
- [3] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003.
- [4] S. K. Nayar and S. G. Narasimhan, "Vision in bad weather," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 820–827, September 1999.
- [5] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 325–332, December 2001.
- [6] S. Shwartz, E. Namer, and Y. Y. Schechner, "Blind haze separation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 2, pp. 1984–1991, June 2006.
- [7] S. G. Narasimhan and S. K. Nayar, "Interactive deweathering of an image using physical models," in *Proceedings of the IEEE Workshop Color and Photometric Methods in Computer Vision, in Conjunction with IEEE International Conference on Computer Vision*, October 2003.
- [8] J. Kopf, B. Neubert, B. Chen et al., "Deep photo: model-based photograph enhancement and viewing," *ACM Transactions on Graphics*, vol. 27, no. 5, article 116, 2008.
- [9] R. T. Tan, "Visibility in bad weather from a single image," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, June 2008.
- [10] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1956–1963, IEEE, Miami, Fla, USA, June 2009.
- [11] K. B. Gibson, D. T. Vo, and T. Q. Nguyen, "An investigation of dehazing effects on image and video coding," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 662–673, 2012.
- [12] R. Fattal, "Single image dehazing," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 988–992, 2008.
- [13] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 694–711, 2006.
- [14] J.-P. Tarel, N. Hautiere, A. Cord, D. Gruyer, and H. Halmaoui, "Improved visibility of road scene images under heterogeneous fog," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '10)*, San Diego, Calif, USA, June 2010.
- [15] N. Hautière, J.-P. Tarel, and D. Aubert, "Mitigation of visibility loss for advanced camera-based driver assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 474–484, 2010.
- [16] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [17] S. O. Haykin, *Neural Networks and Learning Machines*, Prentice Hall, 3rd edition, 2008.
- [18] Y. S. Abu-Mostafa, M. M. Ismail, and H.-T. Lin, *Learning from Data*, AMLBook, 2012.
- [19] P. H. Sydenham and R. Thorn, *Handbook of Measuring System Design*, vol. 3, John Wiley & Sons, Chichester, UK, 2005.
- [20] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–27, 2009.

Research Article

RS485 Image Sensor for Digital Cinema System

Eunju Kim,¹ Seokhoon Kang,² and Sangsoo Lee³

¹High-Tech S/W Research Center, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon 406-772, Republic of Korea

²Department of Embedded Systems Engineering, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon 406-772, Republic of Korea

³Department of Computer Engineering, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Kyunggi-do, Seongnam 461-701, Republic of Korea

Correspondence should be addressed to Sangsoo Lee; sslee@gachon.ac.kr

Received 21 January 2015; Accepted 19 March 2015

Academic Editor: Wei Wu

Copyright © 2016 Eunju Kim 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.

To activate various devices using RS485, a repeater is generally used. In current digital cinema systems, each device is controlled with RS485 by mixing RS485 and DMX512. However, as today's cinema equips hundreds of 4D chairs and the environmental directors, it is nearly infeasible for the legacy system to control. To this end, this paper designs and implements a new system which makes hundreds of 4D chairs and the environmental directors be controlled simultaneously exploiting RS485 network topology and its repeaters. The proposed approach is tested in a real-time system for assessing the performance by Paessler Router Traffic Grapher (PRTG) in Windows environment. Simulation results show that the tested system supports 4D chairs and their motions are well operated simultaneously with RS485.

1. Introduction

A digital cinema indicates the usage of digital technology to deliver and present motion pictures in contrast to the historical usage of motion picture film [1]. Movies can be delivered via internet and hard drives or devoted to web or satellite links or optical disks such as blu-ray disc [2]. Digital movies are presented adopting a digital projector instead of a traditional film projector. Thus, digital cinema is different from HDTV (high-definition television) and does not rely on adopting SD or HDTV standards [3]. In general, resolutions of digital cinema are represented by the horizontal pixel count, usually 2048×1080 (or 2.2 megapixels) to 4096×2160 (or 8.8 megapixels). As digital cinema technique was enhanced in early 2010s, most of the theaters across the world have transferred to digital [4].

Four dimensional (4D) film is a term of marketing for an entertainment presentation system which is advanced form of three dimensional (3D) film with physical possessions that transpire in the theatre in operation with the film [5]. The accomplishments conducted in a 4D film may include lamp, wind, rain, and vibration. The seats in 4D place may oscillate

or move a few inches during the movie [6]. Some other effects contain water sprays, air jets, and leg and back ticklers. In addition, hall effects may include rain, smoke, lightning, and air bubbles, and scent can be considered for 4D effects. Since the physical effects are expensive to realize, 4D films are most often accomplished in custom-built theatres such as amusement or theme parks. Some examples of 4D films are "Journey to the Center of the Earth" and "Avatar," which were presented at movie theatres with 4D versions [7].

Recently, research on IT convergence techniques for film production and film screening are widely studied. Analog-type theater is changing to digital-type theater, and interests in digital-type theater increase for 4D-type theater which yields reality that 3D movies did not provide [8]. A 3D theater is giving a three-dimensional effect to viewers with special glasses such as polarized glasses [9]. In addition, a future-oriented 4D theater helps viewers to try special effects affecting five senses such as wind, vapor, chair vibration, scent, and special lighting [10, 11].

In a traditional 4D digital theater, an environmental system that generates device controller of lamping system and a device controller of sensing of reality are separated [12]. That

TABLE 1: Environmental parameters.

Equipment	Starter	Delimiter				Parameter			CRC
	1 (STX)	2	3	4	5	6	7	12	
LED lighting									
Actual value	0x02	0x4c	0x45	0x44	0~60000		0~255	0~255	
Meaning	Begin	“L”	“E”	“D”	ID top	ID down	Type booking	1~255 data	
Fan									
Actual value	0x02	0x46	0x41	0x4E	0~60000		0~255	0~255	
Meaning	Begin	“F”	“A”	“N”	ID top	ID down	Type booking	1~255 data	
Motion									
Actual value	0x02	0x4d	0x4f	0x54	0~60000		0~255	0~255	
Meaning	Begin	M	O	T	ID top	ID down	Type booking	1~255 data	

is, DMX512 protocol is adopted to control lamping and RS45 image sensor is adopted to control motions, chair devices, and environments generating devices [13]. To this end, two types of controllers must be realized to control each device and generally they cause high power consumption and complex communication lines [14]. In addition, controlling all the subdevices precisely and simultaneously is hard to achieve [15–17]. Therefore, a stable device controller system is needed which can control all the subdevices precisely and provide autoinspection and calibration of real-time operation.

In this paper, RS485 image sensor based digital cinema system is proposed. The system is designed to integrate motion bases, chair devices, and other environmental rendering devices as well as lighting under RS485 image communication. This paper is organized as follows. Section 2 introduces 4D device control system. The proposed system is explained in Section 3. Section 4 provides simulation results and conclusion remarks are described in Section 5.

2. Proposed Design

2.1. Image Coding. The digital cinema architecture utilizes the JPEG standard to realize a real-time decoder. Due to the communication traffic and storage cost, interframe such as video coding can be better choice for movie compression. However, there are some reasons that we selected JPEG:

- (1) There is no global standard for RGB compression with 30 bits or more.
- (2) Intraframe coding methods remain significant due to their supporting of video editing.
- (3) JPEG is easier standard for editing.
- (4) JPEG decoder is more reliable in the error condition.

Therefore, we used JPEG for our digital cinema system.

2.2. Real-Time Decoder. The decoder is able to realize real-time decompression with higher speed of 3G pixels per minute using parallel approach. The decoder is comprised of two circuit blocks: PC/LINUX part with GbE interface and JPEG2000 decoder boards. Four boards are established on the PCI-X-bus for processing 24 frames of 8 M pixels with 36-bit RGB color images in a second. The color standards may be changed such as RGB, YCbCr, or HSV.

2.3. The 4D Device Control System. The 4D device control system is working for the environmental rendering of digital theater. To investigate 4D device control system, we tested some characteristics including chair vibration, wind, vapor, scent, special lighting, and spray in the hardware configuration. We also tested LED lighting, fan, and vibration among the various environmental rendering devices.

As RS485 image sensor only enables multiple connections of various masters and gives half duplex communication, two wires are exploited for sender and receiver. The ID is assigned to each device. If a master sends data “1” to master “*m*,” all devices receive the data simultaneously. However, other devices except master “*m*” ignore the data once they realize the data is not for them. Table 1 shows data configuration table of RS485 image sensor protocol.

Conventional studies performed simulations for LED lighting, fan, and vibration among the diverse environmental rendering device and displayed their accuracy in real-time data transmission. Although previous researches studied data transmission for RS485 image sensor, those studies did not fully provide comprehensive understanding of environmental rendering devices, motion, and chair devices. To alleviate this issue, we verify data transmission for RS485 image sensor and design digital cinema system which is able to provide all the environmental rendering devices, motion, and chair devices.

3. Configuration of the Proposed Digital Cinema System

Let us consider structure of conventional digital cinema system.

Step 1. Main PC supports connectivity between the Internet and sites based on Ethernet. Using 4D system interface shown in Figure 1, the entire system is controlled.

Step 2. This stage is for network server, which monitors devices. In addition, network server converts data between main PCs.

Step 3. This is the last step which belongs to physical layer devices. RS485 works for communication and is designed

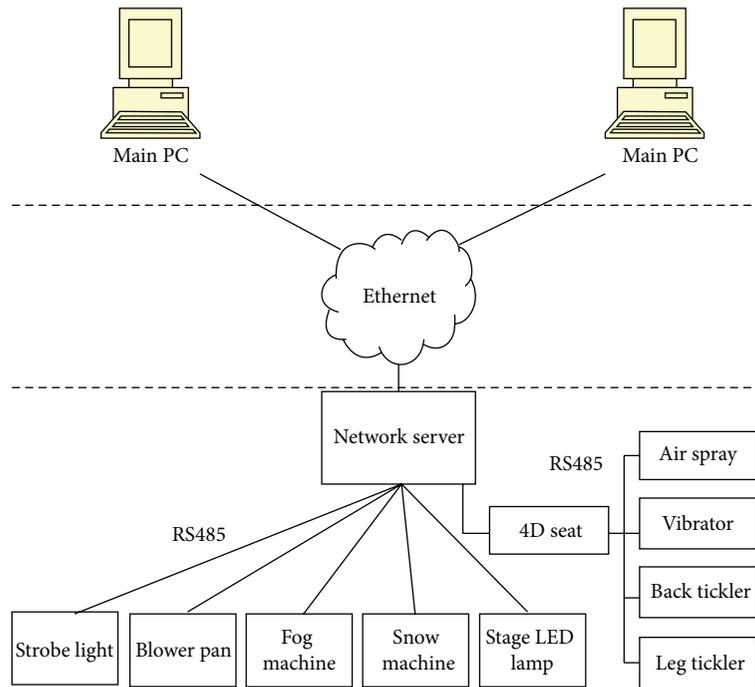


FIGURE 1: Conventional digital cinema system.

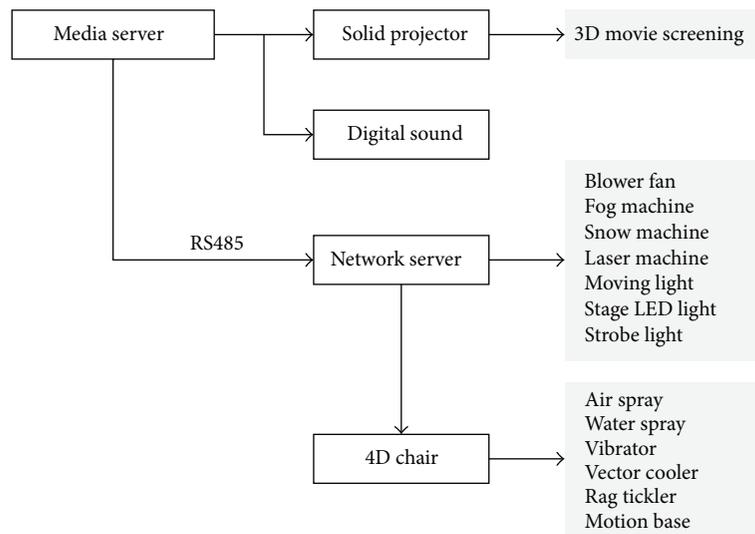


FIGURE 2: Configuration of digital cinema system.

to expand. However, cinema has hundreds of 4D chairs; therefore compatibility with existing system is a big issue. In our system, we use RS485 network bus topology and repeater to control hundreds of 4D chairs and environmental productions.

Figure 2 shows the proposed digital cinema system configuration. When a data file is entered in a media server, a driving controller delivers data to each device through RS485 image sensor. Each device provides environmental rendering effect.

3.1. Design of Digital Cinema System. The driving controller of digital cinema system activates environmental rendering devices in synchronization with screen display as well as devices which are installed on a chair such as motion base. The driving controller of digital cinema system can be installed in automatic manner through touch screen and yields tests for some operations.

3.2. Protocol Definition of RS485 Image Sensor. To define RS485 image sensor protocol, we take into account motion, chair device, and environment devices simultaneously. To

TABLE 2: Mandatory instruction set for environmental devices.

Equipment	Delimiter			Contents
LED lighting				
Actual value	0x4c	0x45	0x44	Color, brightness, and time adjustment of LED light
Meaning	“L”	“E”	“D”	
Strobe lighting				
Actual value	0x53	0x54	0x52	Brightness, blink rate, and holding time adjustment of the strobe light
Meaning	“S”	“T”	“R”	
Heater				
Actual value	0x48	0x45	0x54	Temperature of the heater
Meaning	“H”	“E”	“T”	
Fan				
Actual value	0x46	0x41	0x4E	Wind speed of the fan
Meaning	“F”	“A”	“N”	
Fog				
Actual value	0x46	0x4F	0x47	Fogging degree of fog machine
Meaning	“F”	“O”	“G”	
Curtain				
Actual value	0x43	0x54	0x4E	Curtain’s opening degree
Meaning	“C”	“T”	“N”	
Flash				
Actual value	0x46	0x4C	0x53	Brightness and action adjustment of flash
Meaning	“F”	“L”	“S”	
Ground LED light				
Actual value	0x55	0x4c	0x44	Color, brightness, and motion of ground LED lighting
Meaning	U	L	D	
Ceiling LED light				
Actual value	0x52	0x4c	0x44	Color, brightness, and motion of ceiling LED lighting
Meaning	R	L	D	
Moving light				
Actual value	0x4d	0x56	0x4c	Color of the moving light, operating time, blink rate, movement speed, and position adjustment
Meaning	“M”	“V”	“L”	
Actual value	0x4d	0x56	0x4c	
Meaning	“M”	“V”	“L”	

this end, we used separators for each two to four bytes to distinguish motion, chair device, and environment device commands. We also defined a protocol by taking into account the scalability of environment and chair devices. The basic function of each device is coded as a specific parameter in a command set. When a command is delivered to each device, it activates defined operation. The command is comprised of 12 bytes including 115,200 bps, 8 bit data, 1 stop bit, and 0 parity bit. Consider

$$\text{STX} + \text{separators (3 byte)} + \text{variable (7 byte)} + \text{CR.} \quad (1)$$

Based on (1), command sets are made as Table 2.

To connect hundreds of chairs and servers, we design RS485 repeater. This is to ensure the stability of the communication between the isolated signals and to be able to play the best communication state through the embedded MCU. The

designed system may have total of 4 channels; 1 input signal of each channel is reoutput after isolating via the SN75176.

3.3. Signal Processing. A multiple controller is executed to allot multiple signal processing among various devices, field-effect transistor (FET) board control, power check control, and multipoint control unit (MCU) monitoring, to raise stability. The controller is divided into main MCU, motor MCU, and monitor MCU to achieve mutual monitoring and help. The main MCU is responsible for signal processing and examines encoded signal, digital input/output signals. In addition, the main MCU generates pulse width modulation (PWM) signal which delivers an appropriate driving signal to the FBT board.

The FET board is controlled by the motor MCU. It stops motion base and reports possible errors to the main MCU when an error alert is delivered. The monitor MCU oversees

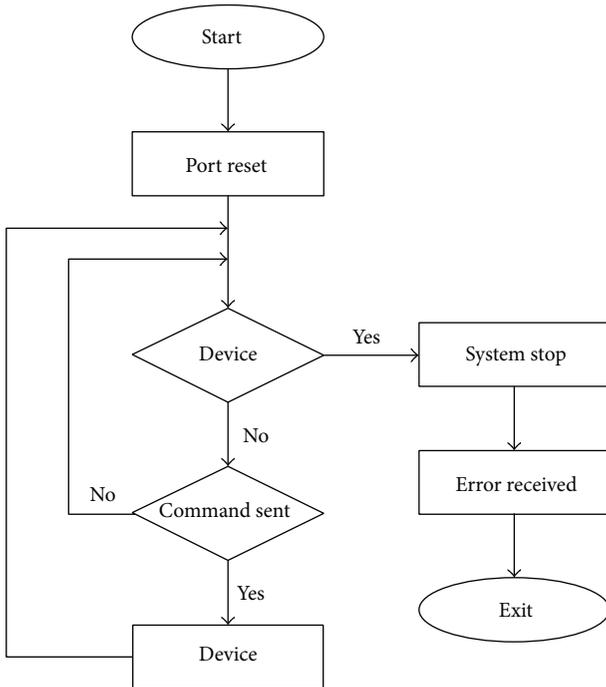


FIGURE 3: Block diagram of the motor and monitor for MCU operation.

power and monitors of other MCUs. The signals overseen by monitor MCU are significant to the safety of the entire system and can tune power supply to the system. The operation sequence of motor and monitor MCUs is shown in Figure 3.

Figure 4 shows the proposed digital cinema system. Step 1 is main PC which supports connectivity between the Internet and sites based on Ethernet. Step 2 is network server, which is able to monitor the devices. Step 3 is physical layer, which is comprised of devices, and commination between devices is RS485. In this paper, we propose a method with repeater which is free of problems on the network even when the number of 4D chairs increases. Figure 4 shows improved digital cinema system.

3.4. RS485 Repeater Design. Repeaters are used only when the network has the same structure, and they belong to physical layer. We designed RS485 repeater for multiple system communication. It can be connected to at least hundreds of 4D chair system and the server, which ensures the stability of the communication between insulated signals. The system has 4 channels; input signal of each channel is isolated via SN75176 and output again.

4. Experimental Results

Simulations of the improved digital cinema system must control hundreds of 4D chairs and a motion base to exactly and simultaneously operate through the RS485 image sensor. Therefore, our simulation was tested by installing 100 4D chairs under the corresponding motion base. The hardware environments are indicated in Table 3. The operating system

TABLE 3: Environments for simulation.

Type	Hardware spec
4D chair	(i) Seats: 2 persons
	(ii) Woofer power: DC 24 V
	(iii) Woofer quantity: 6 Ea
	(iv) Effect accessory: water spray 1 set
	(v) Air spray 1 set
	(vi) Vibrator 1 set
	(vii) Leg tickler 1 set
	(viii) Frame material: steel
	(ix) Cover material: fabric
	(x) Seats material: urethane
	(xi) Accessory: cup holder
	(xii) Side cover material: wood
	(xiii) Weight: 40 Kg
	(xiv) Include wire, hose, and connector
4D motion base	(i) Dimension ($W \times D \times H$): 1,260 × 670 × 450
	(ii) Power: 220 VAC ± 10%, 50/60 Hz, Max. 1,000 W
	(iii) Motion range: Pitch = Max. 6 degrees
	(iv) Roll = Max. 6 degrees
	(v) Heave = Max. 44 mm
	(vi) 3 DOF (degrees of freedom)
	(vii) Effect accessory: leg tickler 1 set
	(viii) Weight: 50 Kg

(OS) of the software environment was Windows 8 and the program language we used is C++. The adopted network environment was Windows 8-based Paessler Router Traffic Grapher (PRTG) simulator.

In the previous works, 4D chair, motion, air injection, water spray, and vibration were considered as personal experience. However, in this study, we excluded group experience and only considered personal experience, that is, 4D chair and motion. Table 4 shows environmental device parameter sets for 4D chair and motion. The environments guided device parameter values are set in Table 4 and stored as a metadata image file. The main screen runs both image data on left and right to produce 3D effect simultaneously. The data transmission is divided by synchronous and asynchronous connections according to the time positions of the transmitting and receiving side. The synchronous data is transmitted in a noncharacter block unit such as frame. In other words, data is made by a predetermined number of strings between the transmitting side and the receiving side, with a format of packet at a time. The asynchronous transfer is motivated by sending information to noncharacter unit blocks. The synchronous transmission has character synchronization method, bit synchronization method, and the frame synchronization method. In this paper, we used character synchronous system which is the most reliable among all synchronous processing. To align characters we add STX in front of blocks and add ETX at the end of blocks to indicate beginning and the end of the transmission data.

TABLE 4: Device parameter values for environments.

Equipment	Starter 1 (STX)	Delimiter			Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5	Parameter 6	Parameter 7	CRC
		2	3	4	5	6	7	8	9	10	11	12
Notion												
Actual value	0x02	0x4d	0x4f	0x54	0–80000		XX	XX	XX	XX	XX	0–255
Meaning	Begin	M	O	T	ID	ID	Type	Type	Type	Type	Type	Data
4D chair												
Actual value	0x02	0x53	0x4f	0x4c	0–80000		XX	XX	XX	XX	XX	0–255
Meaning	Begin	S	O	L	ID	ID	Type	Type	Type	Type	Type	Data

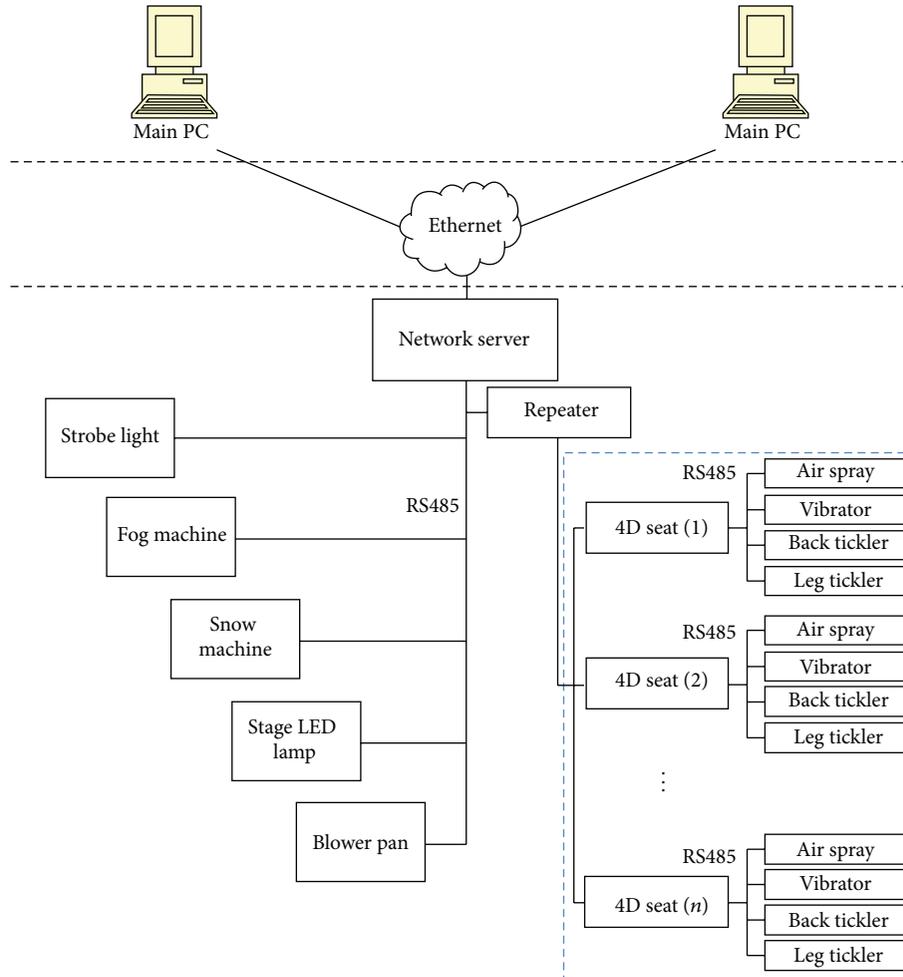


FIGURE 4: Proposed digital cinema system.

The motion chairs and environmental production device parameter values that were set in Table 4 are shown in MOT and SOT. The reason our system starts with STX is that we adopted character synchronous system for its stability in high-speed processing. To inform the beginning and the end, the character synchronous system adds STX to the front and ETX to the end of transmitted data.

From the main screen it can be found that 15 Gbyte movie file is well transmitted to receiver through the 100 Mbps interface of repeater. This confirms that 4D motion and chair were correctly operated when server and each device node

information is properly transmitted without loss. Figure 5 shows transition of motor torques and motor power for M1, M2, and M3. The M1 motor power was 280 W. Based on the simulation, we designed and produced motion base structure. Figure 6 shows a simulation for selected 3DOF motion base dimension.

5. Conclusions

The most representative tangible personal experience factors of 4D theaters are 4D chair and 4D cinema, and hands-on

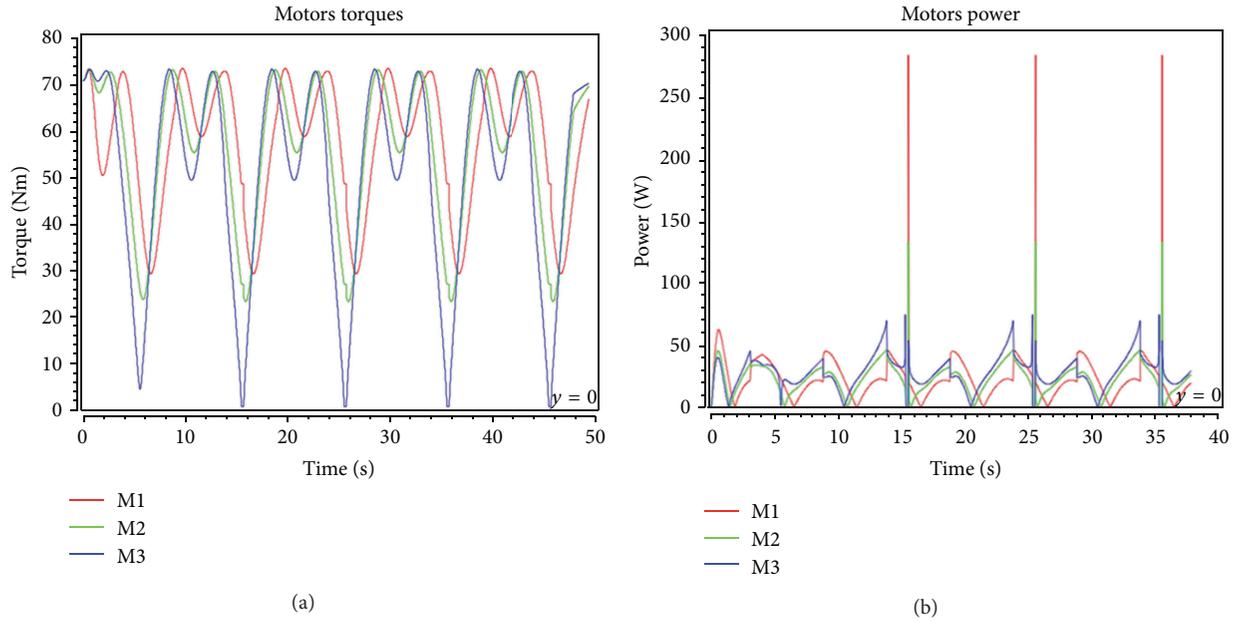


FIGURE 5: (a) Transition of motor torques for M1, M2, and M3. (b) Transition of motor power for M1, M2, and M3.

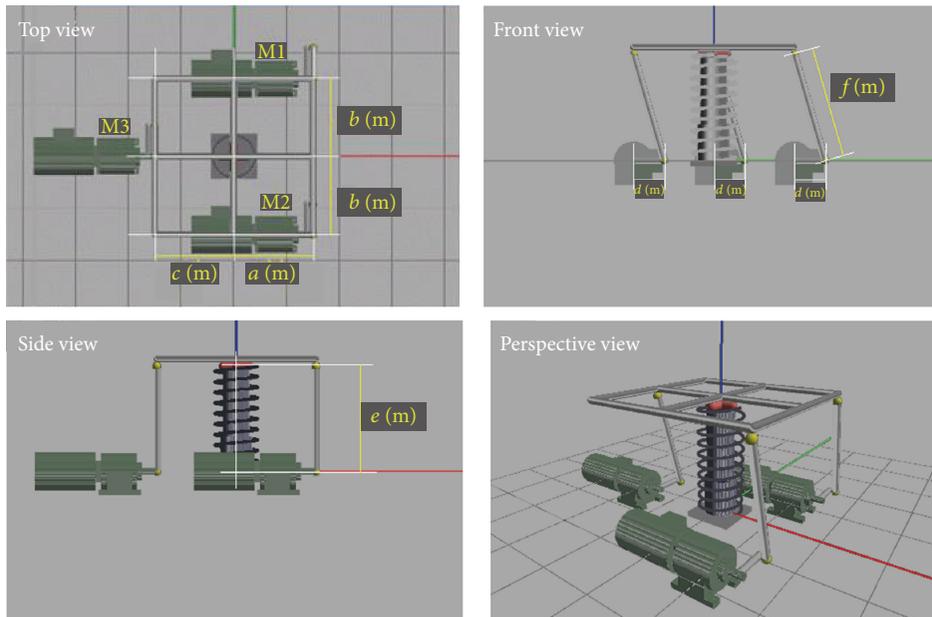


FIGURE 6: Simulation for 3DOF motion base dimension selection.

chairs, motion base, air injection, water spray, and vibration can be elements as well. In conventional studies, only personal experience was discussed for 4D element. In this paper, RS485 network topology, hundreds of 4D seats, and environment implementation were designed to be controlled using a repeater to RS485. Simulation results show that the improved digital cinema system is well designed for 4D chair and found to operate accurately under the condition of no traffic over the RS485.

Conflict of Interests

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

References

[1] T. Yamaguchi, M. Nomura, K. Shirakawa, and T. Fujii, "SHD movie distribution system using image container with 4096 ×

- 2160 pixel resolution and 36 bit color,” in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '05)*, pp. 5918–5921, Kobe, Japan, May 2005.
- [2] Digital Cinema Initiatives, *Digital Cinema System Specification V1.1*, 2007, <http://www.dcinovies.com/>.
 - [3] Specific Targeted Research Project, “IST call 2, Layered compression technologies for digital cinema to graphy and cross media conversion,” WorldScreen, 2004, <http://www.worldscreen.org>.
 - [4] EDCine Project, “IST 6th framework program of the European Commission,” <https://ec.europa.eu/research/fp6/pdf/fp6-in-brief.en.pdf>.
 - [5] WiMAX Forum, “Mobile WiMAX—Part I: A Technical Overview and Performance Evaluation,” 2006, <http://www.wimaxforum.org/>.
 - [6] J. Leigh, L. Renambot, A. Johnson et al., “The global lambda visualization facility: an international ultra-high-definition wide-area visualization collaboratory,” *Future Generation Computer Systems*, vol. 22, no. 8, pp. 964–971, 2006.
 - [7] T. Shimizu and D. Shirai, “International real-time streaming of 4K digital cinema,” *Future Generation Computer Systems*, vol. 22, no. 8, pp. 929–939, 2006.
 - [8] A. J. Woods, “How are crosstalk and ghosting defined in the stereoscopic literature?” in *Stereoscopic Displays and Applications XXII*, vol. 7863 of *Proceedings of SPIE*, San Francisco, Calif, USA, 2011.
 - [9] W.-P. K. Yiu, X. Jin, and S.-H. G. Chan, “Challenges and approaches in large-scale P2P media streaming,” *IEEE Multimedia*, vol. 14, no. 2, pp. 50–59, 2007.
 - [10] H.-H. Yen, S. S. W. Lee, and B. Mukherjee, “Traffic grooming and delay constrained multicast routing in IP over WDM networks,” in *Proceedings of the IEEE International Conference on Communications (ICC '08)*, pp. 5246–5251, Beijing, China, May 2008.
 - [11] L. H. Sahasrabuddhe and B. Mukherjee, “Light-trees: optical multicasting for improved performance in wavelength-router networks,” *IEEE Communications Magazine*, vol. 37, no. 2, pp. 67–73, 1999.
 - [12] T. Inoue and H. Ohzu, “Accommodative responses to stereoscopic three-dimensional display,” *Applied Optics*, vol. 36, no. 19, pp. 4509–4515, 1997.
 - [13] C. Wheatstone, “Contributions to the physiology of vision—part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision,” *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394, 1838.
 - [14] N. A. Dodgson, A. J. Woods, J. O. Merritt, S. A. Benton, and M. T. Bolas, “Variation and extrema of human interpupillary distance,” in *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291 of *Proceedings of SPIE*, pp. 36–46, 2004.
 - [15] R. S. Allison, B. J. Gillam, and E. Vecellio, “Binocular depth discrimination and estimation beyond interaction space,” *Journal of Vision*, vol. 9, no. 1, article 10, 2009.
 - [16] N. S. Holliman, J. P. Dakin, and R. G. W. Brown, *Handbook of Optoelectronics*, vol. 2, Taylor & Francis, 2006.
 - [17] H.-J. Jia and Z.-H. Guo, “Research on the technology of RS485 over ethernet,” in *Proceedings of the International Conference on E-Product E-Service and E-Entertainment (ICEEE '10)*, pp. 1–3, November 2010.

Research Article

Feature Coverage Indexes for Dual Homography Estimation in Constructing Panorama Image

Kyungkoo Jun¹ and Sijung Kim²

¹Department of Embedded Systems Engineering, University of Incheon, Incheon 406-772, Republic of Korea

²Department of Entertainment Engineering and Design, University of Nevada, Las Vegas, NV 89154, USA

Correspondence should be addressed to Kyungkoo Jun; kjun@incheon.ac.kr

Received 24 January 2015; Accepted 21 March 2015

Academic Editor: Wei Wu

Copyright © 2016 K. Jun and S. Kim. 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.

Enlarged images can be obtained by various methods. Stitching is one of the efficient methods. It can produce panoramic images by stitching adjacent images which contain overlapping regions even though they are obtained through separate image sensors. Images that contain multiple different planes are hard to be stitched together because each plane has a different homography matrix for perspective warping. For this, a dual homography was proposed. However its performance varies depending on feature detectors which are used to find matching feature points between images. In this paper, we propose three feature coverage indexes which evaluate the stitching performance of feature detectors and predict the outcomes of the stitching. We evaluate four well-known feature detectors by the proposed indexes by applying them to the image stitching process and show that the prediction by the index values coincides with the stitching results.

1. Introduction

Enlarged images can be obtained by various methods. Stitching is one of the efficient methods and has long drawn attention of researchers from graphics and computer vision fields. Its primary goal is to integrate multiple images into a single panorama [1].

Stitching depends on a perspective transformation which warps pixels from one coordinate frame to another. Its algorithms have traditionally sought to parameterize the warping using a transformation matrix, such as the 3×3 affine or homographic matrixes.

This matrix-based parameterization of the warping provides robustness at the cost of flexibility and is only accurate as long as a set of restrictive conditions are met [2]. For example, the homographic transformation is only applicable for planar scenes or parallax-free camera motion between adjacent images. Thus it requires that the one who takes pictures is not allowed to change one's location but only move in a rotational way.

As the transformation must keep visually accurate alignment of large image regions, it must be tolerant to significant

view point shift. Also, as outdoor environments are beyond control, the transformation must also be robust to illumination changes and motion of objects.

The distribution of detected features across an image is known to affect the accuracy of homography calculated from them [3]. It is desirable that the features are evenly distributed across the image because many vision algorithms are robust only when such conditions are met.

When images contain more than one plane, it is hard to stitch them together into a single panoramic image. For example, an image containing both a distant plane and a ground plane that stretches out from the camera's view point is one of such difficult images. Since both planes have different homography transform matrix, it is hard to build a single universal matrix to apply for the whole image stitching.

For the dual-plane image stitching, existing approaches estimate a single planar perspective transform to align two adjacent images. However a single homography cannot warp the images correctly, requiring postprocessing to remove misalignments.

In [4], it proposes a method to address dual-plane panoramic scene by estimating two-perspective transform per

image-pair, resulting in improved alignment before the post-processing. It estimates dual homographies from matched points and applies different weights to each homography depending on distances to corresponding pixels.

The dual homography approach divides the matched feature points into two groups. Then a perspective matrix is estimated per group. Each pixel uses a weighted sum of the homography matrices to warp into new position as follows:

$$H_{ij} = \omega_{ij}H_g + (1 - \omega_{ij})H_d, \quad (1)$$

where H_g and H_d mean the ground plane and distant plane homographies, respectively, and ω_{ij} is a weight to the pixel at (i, j) , representing which plane is closer to the pixel.

In [5], it improves the problem of the dual homography: curve effects; that is, straight lines in original images are bended after stitching. It lessens such side effect by adding another homography to the weighted sum, resulting in triple homography:

$$H_{ij} = [\omega_{ij}H_g + (1 - \omega_{ij})H_d] \times \frac{W - x_i}{W} + H \times \frac{x_i}{W}, \quad (2)$$

where W is the width of image and x_i represents the horizontal coordinate of the pixel.

Both the dual homography and its enhanced version need to find two sets of feature points to estimate homography. Therefore they are dependent on the feature detection algorithms.

In this paper, we propose three feature coverage indexes which evaluate the stitching performance of feature detectors and predict the outcomes of the stitching. It has been attempted to provide a set of indexes to evaluate the quality of image processing results. In [6], the convex hull was employed to indicate the spatial coverage of feature points. In [7], spatial relationship between feature points was measured by dense sampling scheme. We compare four feature detection algorithms for the dual homography. We use SIFT [8], SURF [9], ORB [10], and BRISK [11] to detect features for estimating homographies.

The rest of the paper is organized as follows. Section 2 describes the details of the dual homography procedure and introduces the feature coverage indexes. In Section 3, we experiment with three sets of images for stitching and evaluate the detectors by the proposed indexes. Section 4 concludes the paper.

2. Dual Homography

Figure 1 shows the flowchart of stitching two images by using the dual homography. At first, feature points are extracted from images. The features are then matched one another, resulting in a set of matched pairs. Then the pairs are clustered based on which plane they belong to: the ground and distant planes. By using the two groups of feature pairs, two homographies are estimated, respectively. The stitching is actually the backward projection: the weighted sum of the inverse homographies is used to calculate the pixel location to fill up the position on the resulting stitched image.

Feature detectors have an influence on the success of the image stitching because the following depends on feature detectors: the number detected feature points, the number of matched pairs, the number of feature points per cluster, the location of cluster centroids, and estimated homography of clusters.

We propose three indexes to measure and evaluate the efficiency of feature detectors for the estimation of homography. Those indexes can be used in a way that if a feature detector obtains high score over all three indexes, we can expect the dual homography estimated by the detector to produce seamless stitched images with high probability.

The first index k_1 measures how many detected features still remain after matching step:

$$k_1 = \frac{m}{n} \leq 1, \quad (3)$$

where n is the number of detected feature points and m is the number of matched feature points. It indirectly represents the efficiency of feature detectors. During the feature detection, the detector requires computing resources such as CPU processing and memory to describe and store feature points. Thus, the more feature points are matched, the less resources are wasted.

The second index k_2 is the ratio between the number of feature points belonging to each cluster:

$$k_2 = \frac{c_n^d}{c_n^g}, \quad (4)$$

where c_n^d is the number of matched feature points belonging to the distant plane and c_n^g is the number of matched feature points belonging to the ground plane. If each cluster contains the same or similar number of feature points, the probability to successfully estimate homography matrices becomes higher. Otherwise, the cluster with less number of feature points is more likely to fail to estimate the homography, resulting in the failure of the dual homography warping.

The third index k_3 is the variance of the distances of feature points to its cluster centroid:

$$k_3 = \frac{1}{n} \sum_{i=1}^n (d_i - \mu_d)^2, \quad (5)$$

where μ_d is the average distance from each of the cluster points to a centroid, d_i is the distance from i th cluster point to the centroid, and n is the number of feature points of the cluster. If feature points are distributed evenly over its plane, the estimated homography becomes more robust.

In summary, these three indexes specify indirectly three conditions required to estimate homography with success; feature points are extensively distributed and evenly over different planes with a sufficiently large number of detected points.

3. Performance Evaluation

We evaluate feature detectors by comparing three proposed indexes and the quality of generated panorama images. We consider four detectors: SIFT, SURF, ORB, and BRISK.

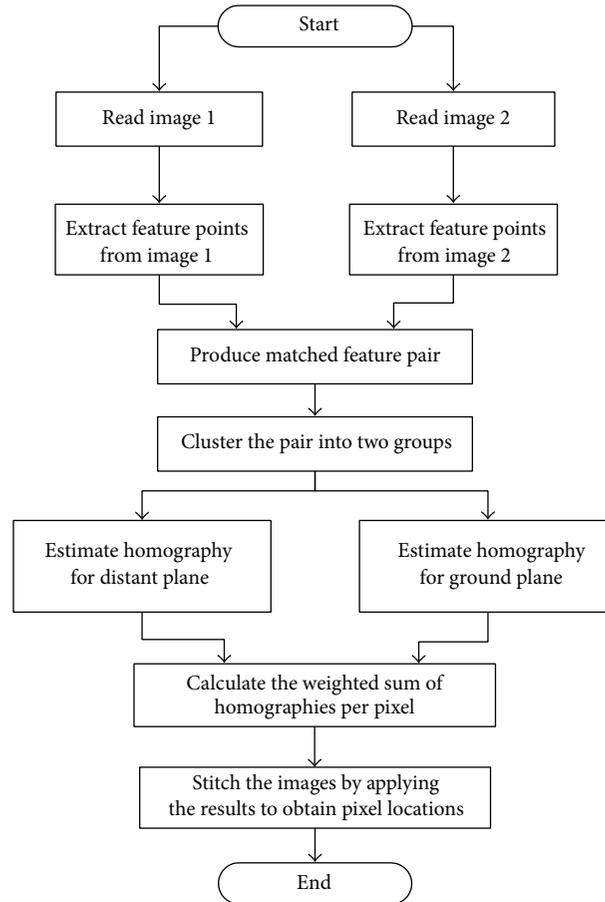


FIGURE 1: The flowchart of the image stitching using the dual homography.

For evaluations, we use three sets of images. For the first set, we divide a single image into three pieces, each of which has overlapped regions with adjacent subimages. Each subimage contains two planes: ground and distant shown in Figures 2(a)–2(c).

By this set, the stitched results can be easily evaluated by comparing with the undivided original image.

For the second set, each image of the first set is transformed by using random perspective transformation matrices as shown in Figures 2(d)–2(e). Note that the center and the right images are warped, leaving the left one unchanged. It evaluates whether the dual homography performs perspective transformation on images correctly.

The third set contains three separately taken images, each of which has overlapped regions as shown in Figures 5(a)–5(c). Each image contains two planes: ground and distant. Stitching the images of this set results in a panoramic image.

Figure 3 shows the stitching results of the first image set when using four feature detectors. All of them stitched subimages successfully into the original image. Since each subimage is not modified with any perspective transformation, the stitching is the process of simple translation.

Figure 4 shows the stitching results of the second image set. Three of the feature detectors were able to stitch the images while BRISK failed to do so. The results of SURF and

ORB are not satisfactory. Only SIFT managed to produce the result similar to the original image.

Figure 6 shows the stitching results of the third set images. All of the four feature detectors were able to stitch the images into larger panorama images. However, there exist differences in seamlessness among the results. Both SURF and SIFT produce the results with higher quality than those of ORB and BRISK. In particular, the contour lines of the buildings in the result of BRISK are indistinguishable because of warping.

We now evaluate the four feature detectors by comparing the proposed index values. Figure 7 shows the values of the three indexes of k_1 , k_2 , and k_3 when the first set of the images are processed by the four detectors. For k_1 , the higher the value is, the more efficient the detectors are. Since the first image set contains the vertical dividing of a larger image, finding matching feature points among them is not difficult. Thus, the four detectors show similar performance. For k_2 , the closer to 1 the value is, the better the chances are to obtain the dual homography. All of the k_2 's of the four detectors have the values around 1, implying that the stitching by the dual homography proceeds smoothly. For k_3 , the higher the value is, the more accurate the estimated homography is because it means that the feature points are evenly distributed over planes. All of the k_3 's have the value of over 200, implying that all of the four detectors are able to stitch the images



FIGURE 2: The three images of the first set (a)–(c) and the three warped images of the second set (d)–(f).



FIGURE 3: The stitching results of the first set from four feature detectors.



FIGURE 4: The stitching results of the second set from SURF, SIFT, and ORB. BRISK failed to produce results.



FIGURE 5: Three images of the third set to be stitched together.

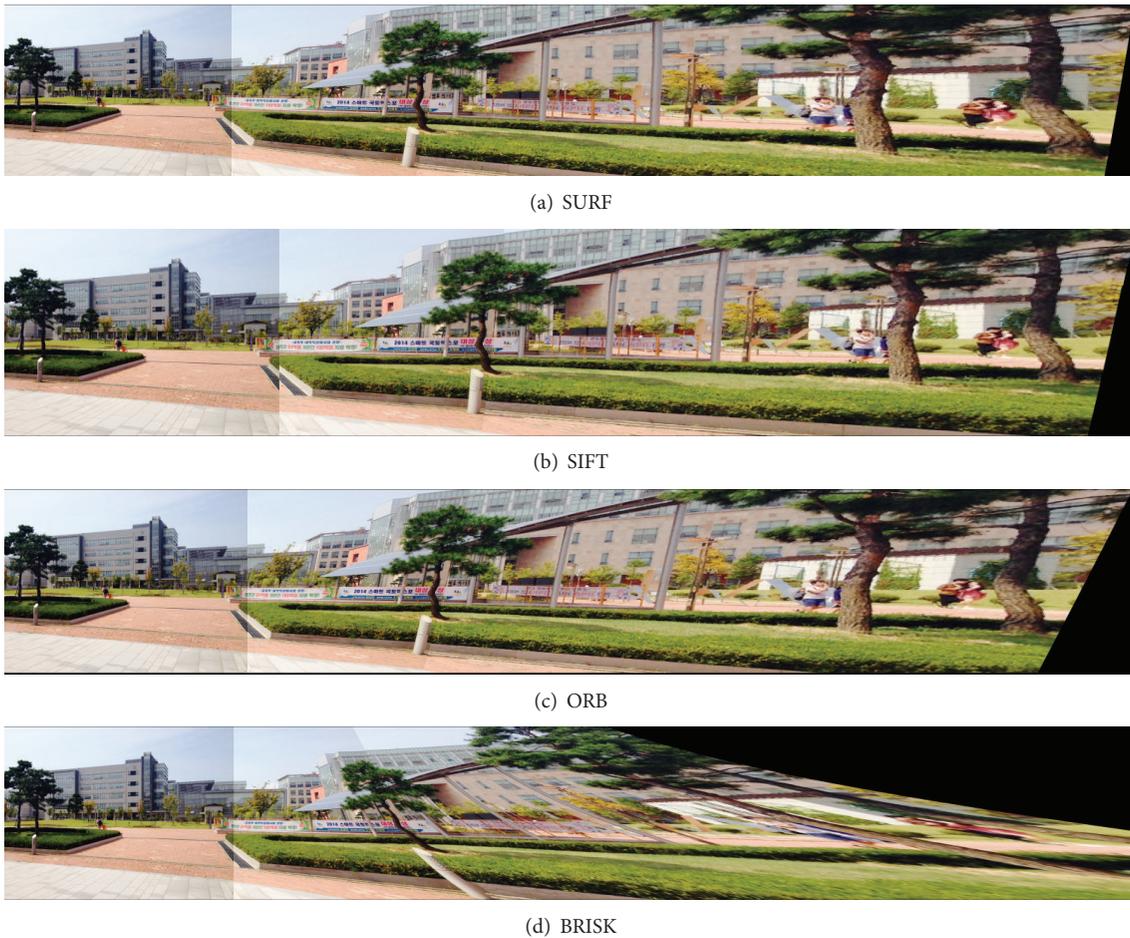


FIGURE 6: The stitching results of the third image set by using four feature detectors.

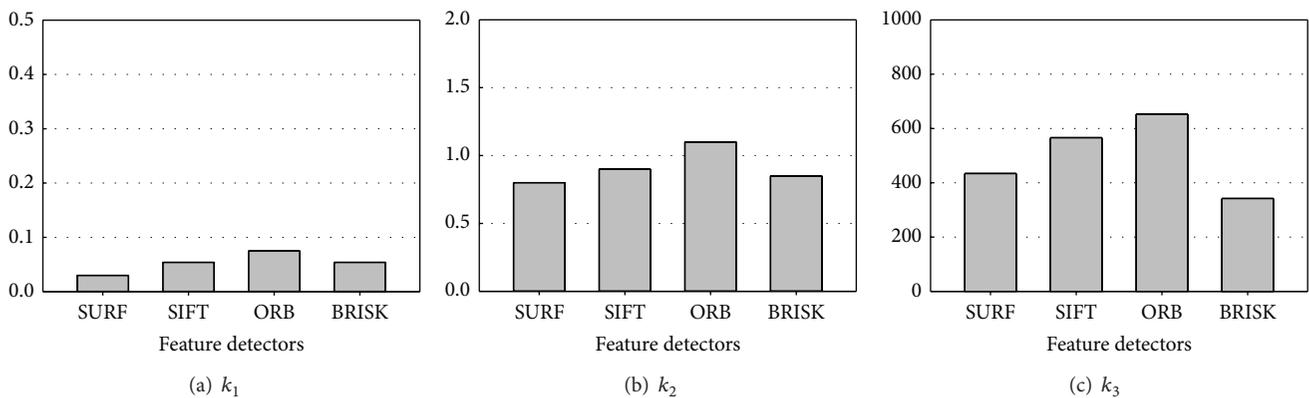


FIGURE 7: The measurement of the three indexes for the first image set.

successfully. Since all of the four detectors have similar index values, we can predict that they produce similar stitching results, which is true when observing the panorama images of Figure 3.

Figure 8 shows that the index values for the second image set vary depending on the detectors. It is because the images of the second set are distorted after dividing a larger image.

Thus finding matching feature points is challenging for some of the detectors. Note that only SIFT has the values which fall into proper ranges; in particular k_2 is very close 1 and k_3 is as high as 600. BRISK has no results because it failed to stitch images. SURF and ORB have the k_2 values which are away from 1 and the k_3 values around 200 implying that the feature points are not proper to estimate correct homography.

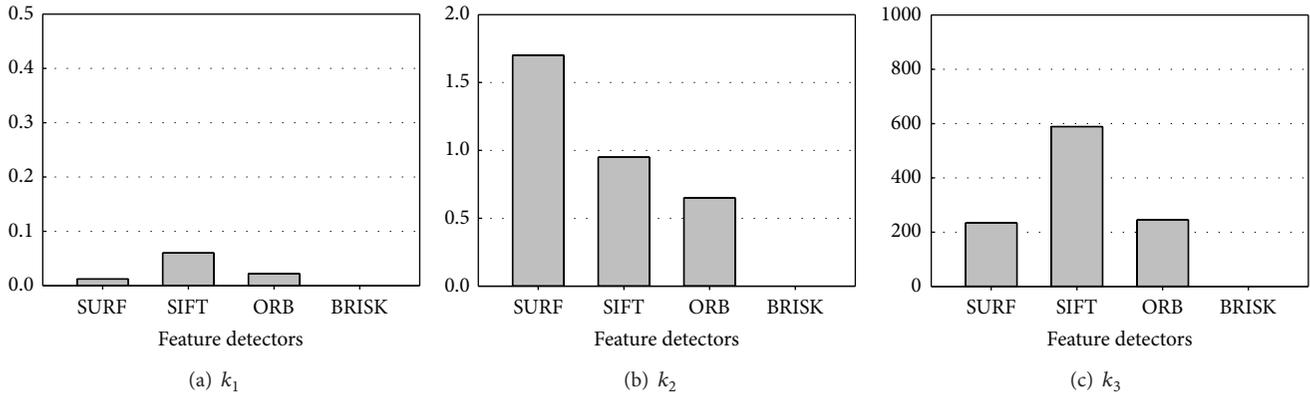


FIGURE 8: The measurement of the three indexes for the second image set.

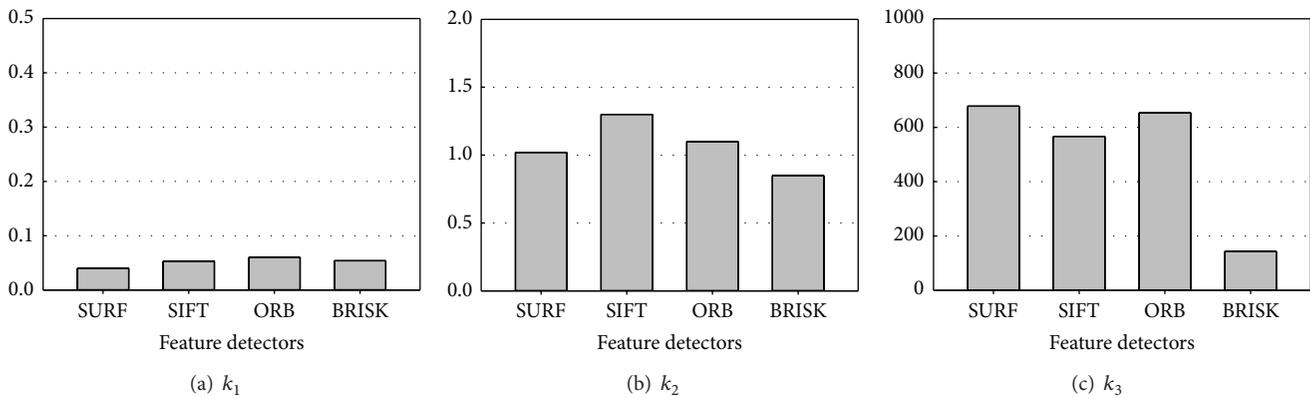


FIGURE 9: The measurement of the three indexes for the third image set.

From these, it is predicted that only SIFT can stitch the images correctly while others cannot, which is true when observing the results of Figure 4.

Figure 9 shows the index values for the third image set. All of the feature detectors excluding BRISK have similar values for all the indexes. Note that the k_3 value of BRISK has less than 200, implying that the estimated homography has errors. From these, it is predicted that the stitching results of SURF, SIFT, and ORB are similar while BRISK produces an incorrect result, which is proved true by the results of Figure 6.

In summary, it is possible to predict the stitching results only by observing the three indexes which capture the capability of the feature detectors to contribute to the estimation of the dual homography. Particularly, we can observe that k_2 and k_3 are closely related with the correct estimation of the homography.

4. Conclusions

We proposed three feature coverage indexes which evaluate the stitching performance of feature detectors and predict the outcomes of the stitching. Particularly, these indexes are developed to evaluate the stitching process involving the dual homography which are for the images containing multiple different planes. We evaluated the four well-known feature

detectors by the proposed indexes by applying them to the image stitching process and showed that the prediction by the index values coincides with the stitching results.

We note that the proposed indexes need improvements in the following area in future works. Firstly, the indexes are applicable only to the cases when stitched images contain more than two planes. It is particularly because of k_2 which involves the number of planes as parameter. Secondly, the indexes are not sufficient enough to evaluate the completeness of panoramic results because the indexes are mostly related with registration while the blending part is not covered. Future works will extend the indexes to be able to evaluate the quality of the stitched boundary.

Conflict of Interests

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

Acknowledgment

This work was supported by the University of Incheon International Cooperative Research Grant in 2012.

References

- [1] B. He, G. Zhao, and Q. Liu, "Panoramic video stitching in multi-camera surveillance system," in *Proceedings of the 25th International Conference of Image and Vision Computing New Zealand (IVCNZ '10)*, pp. 1–6, Queenstown, New Zealand, November 2010.
- [2] E. Bostanci, N. Kanwal, and A. F. Clark, "Feature coverage for better homography estimation: an application to image stitching," in *Proceedings of the 19th International Conference on Systems, Signals and Image Processing (IWSSIP '12)*, pp. 448–451, Vienna, Austria, April 2012.
- [3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2003.
- [4] J. Gao, S. J. Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 49–56, June 2011.
- [5] C. Yuan, J. Pan, M. Sheu, and T. Chen, "Fast image blending and deghosting for panoramic video," in *Proceedings of the 9th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 104–107, Beijing, China, October 2013.
- [6] T. Dickscheid and W. Förstner, "Evaluating the suitability of feature detectors for automatic image orientation systems," in *Computer Vision Systems*, vol. 5815 of *Lecture Notes in Computer Science*, pp. 305–314, Springer, Berlin, Germany, 2009.
- [7] T. Tuytelaars, "Dense interest points," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2281–2288, San Francisco, Calif, USA, June 2010.
- [8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [9] H. Bay, T. Tuytelaars, and L. Gool, "SURF: speeded up robust features," in *Computer Vision—ECCV 2006*, vol. 3951 of *Lecture Notes in Computer Science*, pp. 404–417, Springer, Berlin, Germany, 2006.
- [10] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2564–2571, November 2011.
- [11] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: binary robust invariant scalable keypoints," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2548–2555, Barcelona, Spain, November 2011.

Research Article

Patch Classifier of Face Shape Outline Using Gray-Value Variance with Bilinear Interpolation

Seokhoon Kang¹ and Seonwoon Kim²

¹*Department of Embedded Systems Engineering, University of Incheon, Incheon 406-772, Republic of Korea*

²*TecAce Solutions Inc., Seoul 153-782, Republic of Korea*

Correspondence should be addressed to Seokhoon Kang; hana@incheon.ac.kr

Received 12 January 2015; Accepted 26 February 2015

Academic Editor: Wei Wu

Copyright © 2016 S. Kang and S. Kim. 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 proposes a method to classify whether a landmark, which consists of the outline in a face shape model in the shape model based approaches, is properly fitted to feature points. Through this method, the reliability of information can be determined in the process of managing and using the shape. The enlarged face image by image sensor is processed by bilinear interpolation. We use the gray-value variance that considers the texture feature of skin for classification of landmarks. The gray-value variance is calculated in skin area of the patch constructed around the landmark. In order to make a system strong to poses, we project the image of face to the frontal face shape model. And, to fill out each area, the area with insufficient pixel information is filled out with bilinear interpolation. When the fitting is properly done, it has the variance with a low value to be calculated for smooth skin texture. On the other hand, the variance for misaligned landmark shows a high variance by the background and facial contour gradient. We have proposed a classifier using this characteristic and, as a result, classified the true and false in the landmark with an accuracy of 83.32% through the patch classifier.

1. Introduction

The approaches based on the shape model to faces are the technique that matches the shape model to objects in an image through shapes or textures and have shown good results by applying the various methods [1–3]. In addition, researches are in progress, which consider the illumination in the real world, the impact of a background, and the case that some part of a face is covered [4, 5]. However, there are still some known imitations in the techniques that use the shape model. Representatively, the feature detection and tracking techniques, which estimate an approximation [1] with the least square solution that calculates the tracking parameter used in the shape alignment, have limitations in each [6].

The purpose of the techniques related to the face does not aim to simply fit shapes. Mainly using the location information of the landmark, they are used in systems that perform the face recognition [7], gaze estimation [8], and expression recognition [9]. If a landmark is not properly placed on a feature point, each system causes serious problems in these approaches. Furthermore, when tracking is progressed for

a continuous image from a video stream, when a landmark of shape stays in the background, when the face is obscured, and when the object becomes a variation in the type that cannot be generated by a variant of the shape model, the situation occurs in which the accurate fitting cannot be made according to the environmental changes.

Therefore, we have decided that another verification method is required in addition to the optimization strategies and the shape alignment method. If it is gone through the verification, prior to using the result of shape fitting, the reliability of the current shape can be determined. For example, it can be determined whether to use the result or not before using the data. Also, if the verification result is not reliable as the result of fitting in the video stream, the current tracking is abandoned and it can be operated to start again from the search process. Or searching for a better spot for the landmark can be induced.

We propose a method of verifying the landmark constituting the outline of the face to achieve this. But it may not be enough to go through the verification process only for the outlines of the landmark of the face shape model. However,

every landmark is associated with each other in the form of the shape model. Thus, verification on the outline of the landmark can help to separate the fitting in the overall shape model.

Landmarks of a face contour have a feature suitable to be distinguished as compared to other landmarks because it is located at the boundary between the background and other objects in the two-dimensional image. Our approach is to classify landmarks through the analysis of the combination of a texture and a shape for a landmark. If a patch is configured that contains the pixel information around each landmark, this patch will all include the background, the skin of the face, and the boundary gradient. Therefore, as analyzing texture features targeting patches around the landmarks of the face outline, the result placed on the facial contour can be verified.

We named the classifier “patch classifier” that classifies landmarks using features described above. The patch classifier configures patches for the landmarks of the shape model and classifies whether the fitting is correct or not by analyzing the features from the gray-value variance calculation.

Most of the shape models go through the fitting process in accordance with variations in poses. At this point, the case occurs that the pixel information is not shown, so it goes through the process of filling content of a missing pixel using bilinear interpolation. Since the variance represents the relationship with the average, the bilinear interpolation is suitable for filling the missing information without alteration of variance results.

This paper is developed in the following flow. First, in Section 2, we review the approaches associated with our system. In Section 3, the patch classifier that we propose is introduced. In Section 4, the verification of the patch classifier is shown through an experiment. Finally, Section 5 refers to the conclusion, and Section 6 concludes with the contributions and suggestions for future research of the paper.

2. Relative Work

2.1. Face Shape Model and Fitting Method. Showing good performances by the method of matching the shape model with an image and the technique that matches the shape alignment with the image object, various approaches have been studied. The representative methods of the shape are Active Shape Models [1, 10–12] as start, the Active Appearance Models [2, 13–15] which uses the texture, and the Constrained Local Model [3, 6, 16] which uses the patch texture around a landmark. These studies are common to estimate the position of a landmark, each feature point. And it updates the optimized shape parameter and goes through the process of transforming the reference shape, and generates a shape that fits an object.

Starting with these techniques, a more extensive research direction is created. Typically, by aligning the nonrigid shape model on the face image and distinguishing the expression, it estimates the emotion of a person [9, 17]. A pose is estimated using the geometric information of the shape [18]. And by using the pose and eye position information of the shape, the research on gaze tracking is achieved [8]. Also, by generating

the frontal face image from the nonfrontal face, it is used to identify faces [7, 19]. Above these, the various studies are underway to make good use of the shape in a number of areas. In this paper, before using the shape model, by judging whether the result can be trusted or not, it can be usefully utilized.

2.2. Gray-Value Variance. Gray-value variance is an attribute that can calculate texture as the feature. In Tracking-Learning-Detection, as the first state from the 3 states in object detection method, the gray-value variance of the patch was used [20]. If it is scanned by a scanning window and the variance is less than 50% compared to the target object, it is rejected from the candidate group. This process is commonly a nonobject patch, and it is determined from the background like sky, street, and so forth.

Also, in [21], similar to the previous paper, a cascade filter based tracking algorithm of multifeature is proposed in tracking the object. In this way, in the stage of cascade, if the gray-value variance is lower than 80% and greater than 120%, it is used in the method of rejection.

The Fingerprint image segmentation research has proposed the algorithm using gray-level variance as threshold [22]. In this way, the gray-value variance is used to filter objects or the background in the object detection research.

2.3. Most Related Approaches. In [7], as suggesting the view based active shape model, it deals with the content of the boundary extraction. In this research, the whole active shape mode is initialized with the Procrustes analysis; however, due to the individual differences, in order to solve the case that the shape grows apart from the actual face area, the process of extracting the boundary of the face is added. It is limited through the boundary in the square that includes the top, bottom, and sides from the eyebrow to the chin and the end of nose. From the bottom to the top of the rectangular area in which the edge strength and smooth combination are optimized, it was formulated by the curve running. Optimal curve is found by dynamic programming in seam carving. The seam means path of least importance. First, the importance function is measured by measuring neighbor pixels of image. The seam carving consists of finding the path of minimum energy cost. Then the optimal curve is excluded and this part is maximized instead of minimizing the edge strength. Later, it updates the boundary of AAM points using curve points.

The study is different from our purpose, but it is consistent with the point that it draws the result through the face outline. However, it extracts the contour of a face by using the edge components of the face outline proposed in the above mentioned paper. The method of using the edge represents the form feature of an object, but this would not be able to avoid the influence on the background of real life.

3. Patch Classifier

The patch classifier proposed in this paper has a role to classify whether the landmark that consists of a face outline is

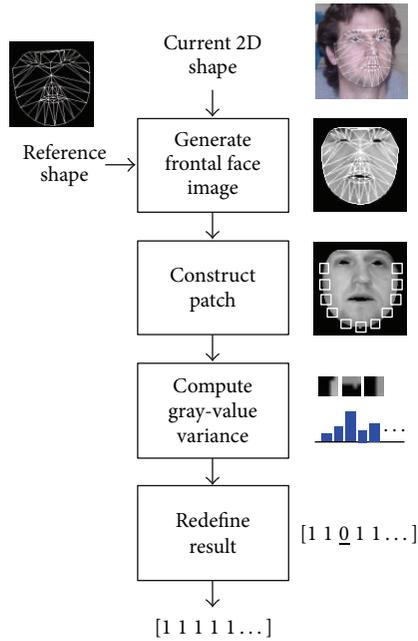


FIGURE 1: Flow of patch classifier.

properly placed on the outline of the face image. The key idea of the patch classifier is that the inside of a shape composed of the outskirts of a landmark consists of skin. Since a region other than the outer face is only composed of smooth skin texture, when calculating the gray-value variance, it shows a low value. To calculate this, we construct the area around the landmark with patches. In addition, the outside of the shape composed of landmarks that constitute the face contour is treated with a mask. The patch that excludes each mask should not include the background and gradient boundary of the face. Using that, the patch classifier judges whether the landmark is properly placed on the facial contour or not. Prior to these works, it is necessary to create a frontal image of the face so that you can work in the free pose of the shape model. Figure 1 shows the flow with the simple picture. We use current shape in tracking method and trained reference shape. First, the frontal face image is generated by projecting image warped current shape to reference shape. And we construct patches for set of pixel in region around each landmark of reference shapes. Then the gray-value variance is calculated to analyze the characteristics of the patch. Finally, we decide classification results using the calculated gray-value variance and geometric feature of face shape model.

This process will be described in the following flow. First, Section 3.1 deals with the process that configures the patch after creating a frontal face image composed of gray-value in the preprocessing section. Section 3.2 explains the process of calculating the variance of each patch. Finally, Section 3.3 describes the process of determining the variance calculation result again considering the relationship between landmarks.

3.1. Preprocessing. A frontal face image is produced with the shape and image which are in tracking to compute

the variance of the patch classifier. Next, it goes through the preprocess for the calculation of variance by constructing the patch made of gray-value of landmarks.

3.1.1. Frontal Face Shape Model. The shape model is basically performed through the calculation of the mean shape and the parameter [1].

In this paper, we use the current shape, image, and the reference shape of the Point Distribution Model (PDM) used in the shape alignment. Reference shape is the mean shape of the shape of the training set that is used when the PDM is training, and it is generally from the front face. The shape model is utilized only for the face when the background was removed. Since our approach is pose-invariant system, the frontal image is prepared for the fact that the patch calculates under the same condition regardless of the pose.

3.1.2. Generate 2D Frontal Face Image. Using the reference shape, the current face image is transformed into the frontal image. In a number of studies, Piecewise Affine Warp [23] is used to produce a face image. This technique is the texture mapping, which maps the different area from the image. First, it configures the Delaunay triangulation in the shape and reference shape currently being tracked. Then, using a bilinear interpolation correction with the center coordinates, it maps the specific triangle to the target triangle. When a pose variation occurs in the shape model, it is to fill with bilinear interpolation for the case that 2D face image is not shown. The bilinear interpolation is an extension of the simple linear interpolation. It fills the empty area by the method that does not affect the gray-value variance which will be calculated later.

In general, when generating the front face image through the Piecewise Affine Warp, the area other than the face region is thereby excluded by the mask. That is, the other area except the set of the triangle region is treated with a mask. The mask is possible to be obtained by a convex hull to the reference shape. Figure 2 shows an example of the result of the process.

This mask is not used solely for the Piecewise Affine Warp transformation. In the variance calculation, the mask area is excluded to calculate only the skin area. In this way, it is to naturally review the outer boundary of the face. The background is included in the outside area of connecting lines between neighboring landmarks. On the other hand, only the skin is included in the inside of the boundary.

3.1.3. Construct Patch. Each patch specifies a rectangular area around the landmarks of the frontal face shape. The size of the rectangular area determines the size of the patch in proportion to the whole size of the shape model. All the created patches have to include a boundary of the face contour.

The reason to specify the patch is to reduce the effect on the skin or the light amount when the variance is calculated. The gray-value variance is the number that indicates the extent of the fact that the gray-value image is spread from the average. That is, the value indicates the distribution of the pixel in the image. Therefore, if we specify a range of image as the full face, the large numerical value must be



FIGURE 2: Examples of frontal face image.

acquired due to the gradient by the illumination, eyes, nose, mouth, and so forth. If the variance is locally calculated by configuring the patch, on the other hand, it is possible to reduce the effect on the gradient of the overall light in the face. Because we only calculate the variance to determine whether the texture inside of the shape model is soft, calculating in a limited scale is more stable and efficient.

3.2. Gray-Value Variance. This section describes the details for calculating a variance for each patch via the previous preprocessing. Variance is a value that measures how much the number sets of the values are spread out. Low variance means that the value is very closely gathered to the average. On the other hand, a high variance value indicates that the numbers are located far from the average. That is, if there is no change in the gray-value in the generated patch, variance will be low. As a result, if the variance of the gray-value is low, it means it has a smooth texture. Conversely, if the calculated value of variance is high, it means that the background or any other object is included in the patch. Thus we calculate the variance to the patch consisting of the landmark as the center and determine the skin texture detection by calculating the variance. For this, the gray-value variance σ_i to the i th landmark l_i is calculated as the following equation. First, the mean of gray-level values except mask M for image I is calculated by

$$E(I, M) = \frac{1}{N} \sum_{x=0}^n \sum_{y=0}^n i(x, y) m(x, y), \quad (1)$$

where $i(x, y)$ is the gray-level value of coordinate (x, y) of the image I , $m(x, y)$ is the pixel of coordinate (x, y) of the mask M , N is number of pixel when the value of the coordinate (x, y) is 1, and n is size of image. The image type should be a grayscale. And the mask consists of 0 or 1. Equation (1) is used for calculating variance except points where value of mask is 0.

The mask has been used in the preprocessing. The ideal result should be that the mask is cut out along the face contour. Through this fact, the gray-value variance of frontal face image excepting the mask is

$$\sigma_k = E(P_k^2, M_k) - E^2(P_k, M_k), \quad (2)$$

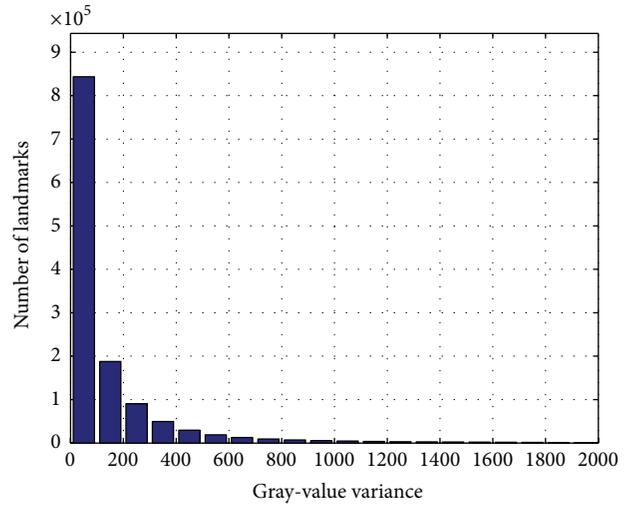


FIGURE 3: Gray-variance of the patch around the ground truth landmark in the Multi-PIE.

where P_k is the patch that contains $n \times n$ gray-level value in the center of the k th landmark. We can calculate the variance of the range which should contain only the gray-value of the skin through this formula as noted above. We have collected the appropriate range of variance through the landmark passively placed. In Table 1, the patch, the actual gray-value figures, and a gray-value variance are shown. The mask M area is shown as black in the patch of which gray-value is 0.

We measured the gray-value variance from the face image of the Multi-PIE database [24] and the ground truth landmarks in order to determine the numerical value for the calculation results in the general image and the correct position of the landmarks. The face images of the Multi-PIE database contain a variety of light direction, the number of people, and even the beard. The results are shown in Figure 3. This graph shows the result of collecting the number of landmark that has a gray-value variance included in the gray-value variance range shown on the x -axis. In the result of the measurement, in the landmarks of the entire face outline, approximately 87.52% of the gray-value variance showed a value of less than 200. Through the result, we come to believe that we are likely to determine the possibility

TABLE 1: An example of gray-value variance.

k	Patch	Gray-values					Variance
0		0	0	0	0	125	1491.06
		0	0	0	32	126	
		0	0	0	32	123	
		0	0	0	55	127	
		0	0	0	73	129	
2		0	0	0	110	119	93.8496
		0	0	0	117	119	
		0	0	0	110	121	
		0	0	0	120	122	
		0	0	0	88	109	
6		0	119	142	148	139	137.96
		0	113	139	132	141	
		0	0	120	132	127	
		0	0	0	122	134	
		0	0	0	0	106	
9		120	113	112	110	108	105.604
		103	98	111	102	101	
		0	0	0	79	0	
		0	0	0	0	0	
		0	0	0	0	0	
11		91	81	71	67	75	98.7288
		80	63	64	61	70	
		61	59	58	0	0	
		68	52	0	0	0	
		0	0	0	0	0	
13		102	90	85	78	0	32.2301
		90	88	87	0	0	
		88	87	80	0	0	
		86	84	0	0	0	
		85	79	0	0	0	

of whether the fitting is correctly done in the proposed method. Some landmark represents high gray-value variance value by a gradient according to the direction of the light source and the beard. Since this phenomenon can occur more frequently in the real world, the patch classifier determines the classification through one more step.

3.3. Considering Relation between Landmarks. As described above, the result of whether it is placed in the correct position of the landmark can be estimated through the calculation of the variance of the gray-value patch. However, due to the environment like the amount of light, shadow, skin characteristics, and the beard may cause different results. In addition, there is a weak point in what will be calculated from the variance. For example, let us assume that the front face is being tracked in front of a single color background. If the patch only contains the background due to the failure of the face alignment, the result indicates a very low variance and the incorrect fitting result, more likely, will be classified as normal. To solve this problem, we redefine the result of the variance through the relationship between the landmarks of the shape.

As a result of the trace for the first situation in Figure 4(a), gray-value variance of one landmark is calculated over 300; it is classified as an incorrect result. In this case, due to the gradient of the facial contour, it came to have a high variance. However, the RMS error of the manually placed landmark and location is actually a very small number, and it is the point needed to be determined as the correct fitting.

Since these patches are created with the landmarks that consist of a face, features of a face can be used. The landmarks of which variances are calculated have the shape of curve that forms the face boundary. Also, because it keeps the reference shape by transforming using the parameters in the shape alignment technique, it is difficult that only one landmark takes the protrusive shape. Thus, the result of the protrusive location is treated as the same as the result of the peripheral landmark.

As the same phenomenon of this, the different situation in the protrusive result is shown in Figure 4(b). This result obtained the very low value of variance since the around patches of protrusive landmarks remain on the background of a solid color. Further, a single landmark to start escaping showed a high variance. Until this situation, it is the same as

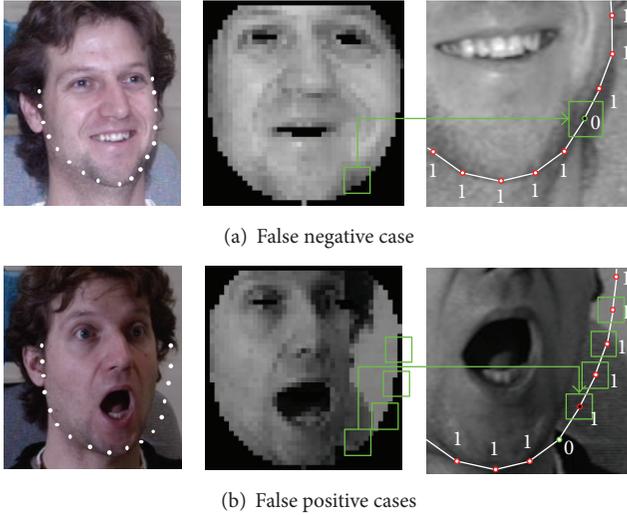


FIGURE 4: Examples of the situation that needs to consider the relationship between landmarks. The first column represents result of fitting, second column represents frontal face image, and third represents parts of classified results using gray-value variance (1 is true, 0 is false).

the previous situation. In order to prevent this, the mean to the patch that does not exclude the mask used in the Piecewise Affine Wrap after the landmark is

$$E'(I) = \frac{1}{n^2} \sum_{x=0}^n \sum_{y=0}^n I(x, y). \quad (3)$$

Therefore, the gray-value variance to the whole patch is

$$\sigma'_i = E'(p_i^2) - E'^2(p_i). \quad (4)$$

This calculation will calculate the variance to the patch including the background face and so on. If the landmark stays in a solid colored background, since the face contour does not exist, the similar variance will be retained.

4. Experimental Results

In this paper, the patch classifier, as described above, classifies the correct fitting result by calculating the variance of each landmark nearby. We did an experiment based on the CMU Multi-PIE database [24] to evaluate the accuracy of the result. Also, after performing by applying a patch classifier to the actual shape fitting technique in the video frame of FGNet [25], according to the RMS error, the measurement of the gray-value variance and the result of the classification are checked. As the same experiment, the patch classifier in real-life images with a complex background is reviewed.

Verification of the basic patch classifier proceeds in the Multi-PIE database. Multi-PIE has various pose, facial expression changes, and face images to the illumination. 68 ground truth landmark points for this have attached to comments. Each image is the 640×480 size. A total of face images for 346 people exist in various poses and, at the same

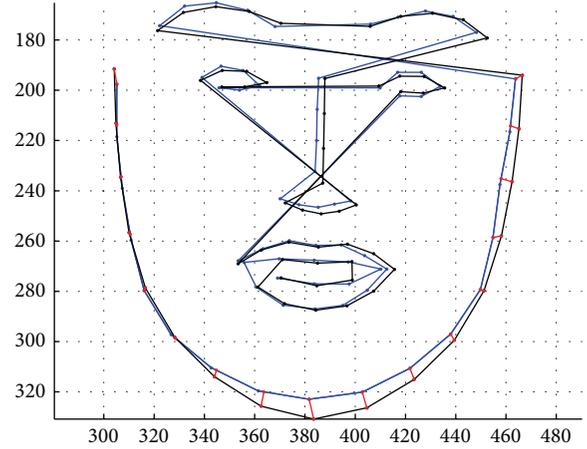


FIGURE 5: Shape RMS error to the face contour.

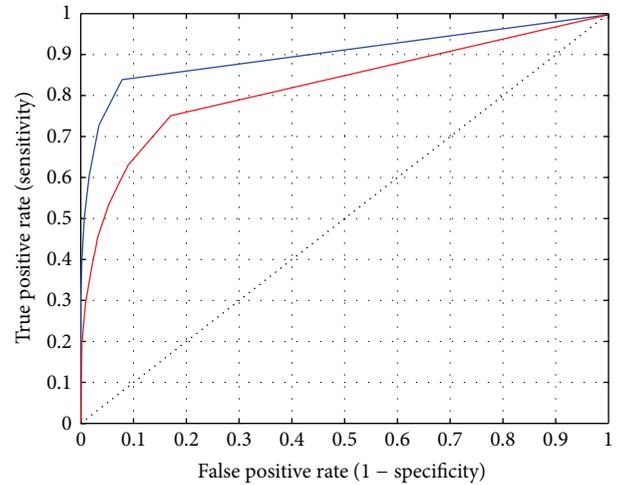


FIGURE 6: ROC curve according to the wide range of variance in the Multi-PIE. Red line is the ROC curve that is measured only by the threshold of the gray-value variance to the Multi-PIE. Blue line is the adjusted ROC curve considering the threshold of the gray-value variance to the Multi-PIE and features of the curve of the facial shape.

time, 20 sets of direction of light are composed. Also, it is made up of images of faces taken from the camera in a total of 15 directions. We use the total 75,360 images by utilizing a set of three directions. This database is a database that is primarily used in the face shape fitting performance comparison, so we carry out the experiment about the gray-value variance compared with the ground truth landmark advances.

In this paper, because it determines whether landmarks are correctly positioned in the facial contours or not, errors are estimated by the distance between the existing ground true landmark and the face contour, not by the RMS error. This example is shown in Figure 5.

In order to find an appropriate variance threshold σ to distinguish between true and false in the patch classifier, the Receiver Operating Characteristic (ROC) curves for a wide range of gray-value variance are presented in Figure 6.

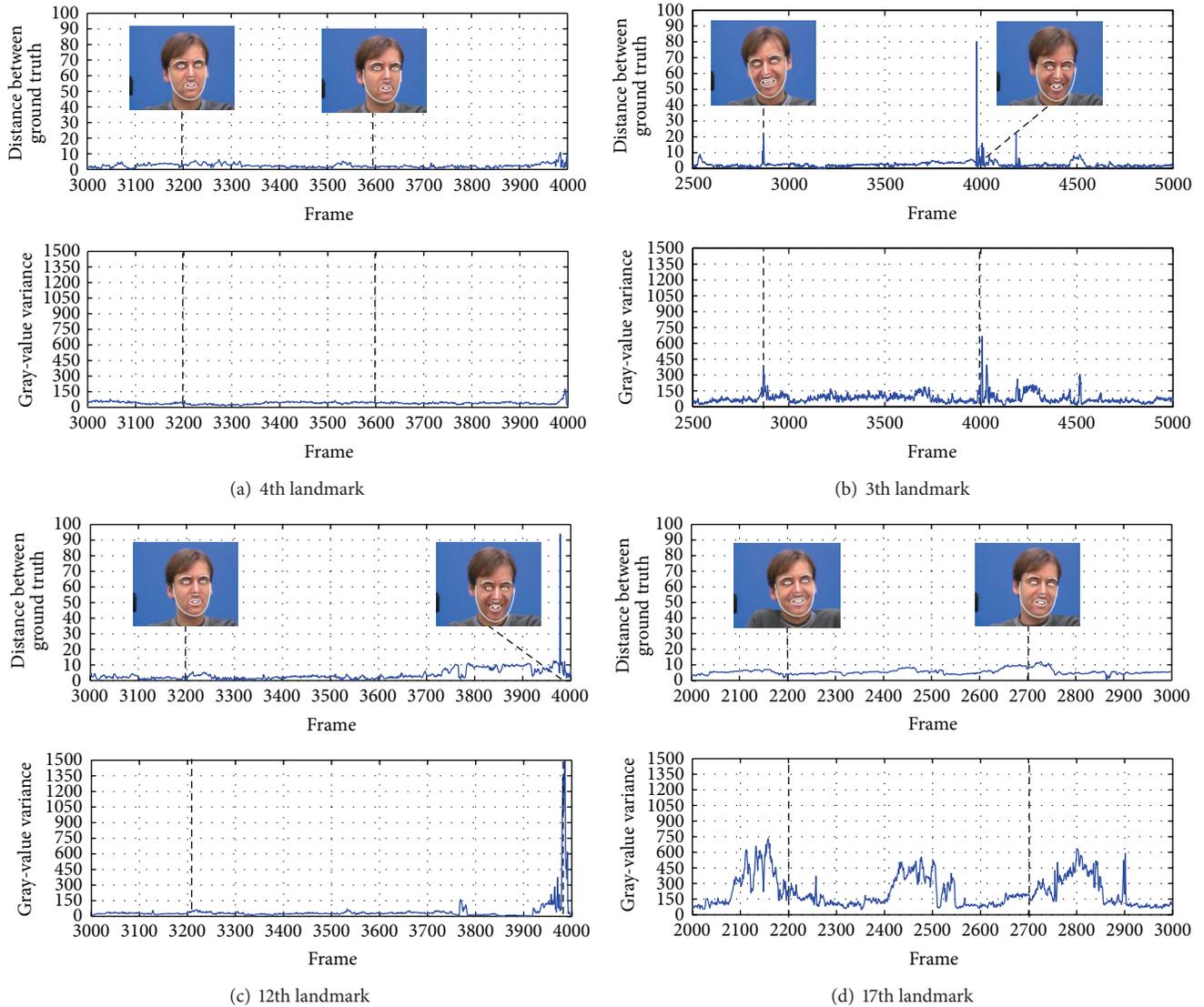


FIGURE 7: Each landmark error and gray-value variance in the FGNet database.

The curve is calculated only for the outline of the landmark of the shape model. It is used as the ground truth landmark of the Multi-PIE for each sample image and goes through the fitting process for the same image using the technique of [6]. We thus apply the patch classifier to the around of the landmark of the facial contour for this fitting. The fitting result is defined as true when the distance of the ground truth landmark is less than 3 pixels and defined as false when it is greater than that. In ROC curve, the red mark is the case that the result is determined only by the gray-value variance. The blue mark is the result of considering the relationship between the landmarks. By considering the relationship between the landmarks, it was identified that the result was improved in the true positive rate before the correction, and also the false positive rate became much lower. The critical gray-value variance in which the classifier shows the best performance is showed as 228. If it is classified

by using this critical value, the true positive rate is 87.71%, and the true negative rate is 71.84%.

The experiment is processed for continuous images using the calculated critical value of the gray-value variance. In Figure 7, the result for the distance between fitted landmarks and ground truth landmarks in the FGNet talking face video database [25] and the gray-value variance is showed. As the same as in the experiment of Multi-PIE, the fitting technique [23] was used. The graph at the top represents the distance between the landmark, the result of the fitting, and the ground truth landmark. As shown in the graph, this fitting technique maintains the distance of approximately 10 pixels on the whole in tracking. According to this, it is confirmed that the gray-value variance for each landmark shows less than the value of 300. As shown in Figure 7(a), if the fitting result is stable, the gray-value variance can be seen to be stable. And Figures 7(b) and 7(c) show the relationship between

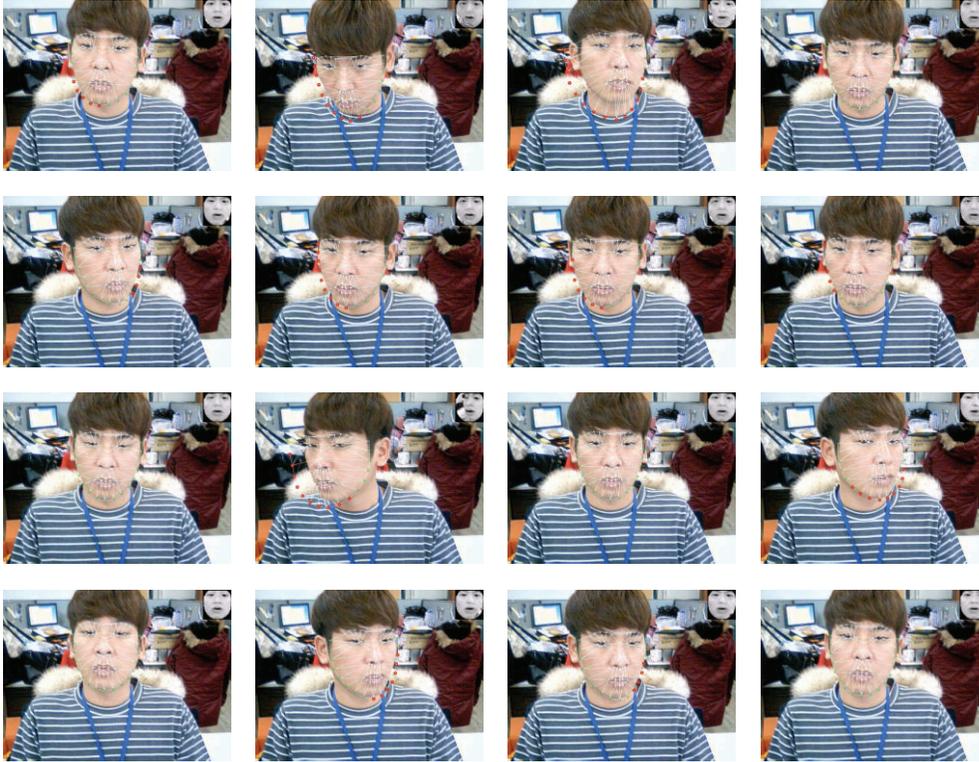


FIGURE 8: Example of the result of patch classifier in the complicated background. The red point represents the failure of fitting and the green point represents the success.

gray-value variance and error. As shown in Figure 7(d), the 17th landmark, which are the gray-value variance at both ends of the landmarks of the outline of the face, do not seem to be stable. Since this position does not only contain the skin due to hairs, the value must be high. However, according to considering the relationship between the landmarks of the shape as mentioned, the result of the patch classifier can be returned as true according to neighbor landmarks.

Finally, the result of applying the patch classifier on real images is shown in Figure 8. The green dots represent the landmarks of which normal fitting is made and the red dots represent the landmarks that are classified by the fitting failure by the patch classifier. In the upper right of each image, the frontal image is placed that its gray-value variance is measured, after configuring the patch. The image represents the result that the patch classifier classifies the case that the shape cannot keep up with or the landmark stays on the background by the rotation. It can be identified that, in the frontal face image in the upper right, the nonface contour area is included. This method showed better performance in complex background. The gray-value variances in complex background are higher than the gray-value variances in simple background. When the face rotated yaw axis, our approach can classify landmarks properly.

5. Conclusion

In this study, the patch classifier has been studied which determines the fitting result of landmarks that consist of

the face contour lines of the face shape model. The patch classifier approaches to estimate a texture of skin through a gray-value variance measurement. To show the performance invariant to pose, it transforms to the frontal shape model. In this transformation, we were able to fill a hidden pixel by a pose using the bilinear interpolation without modification in measuring the gray-value variance. Further, in order to reduce the interference with the illumination or the like, the method of configuring a patch around a landmark is used. A gray-value variance is calculated only in this patch. If you apply this approach to the Multi-PIE database, approximately 87.52% of the gray-value variance of the landmark is confirmed to appear as a value of less than or equal to 200. The texture and shape features are dealt with in the approach of the existing fitting methods and showed the potential for the use of the features in this study. In addition, by applying this technique to the outskirts of the landmark face shape model, we proposed a method to classify whether a fitting is successful or not. As a result, it was able to classify the correct fitting result in the probability of 83.32%. We could verify whether a gray-value variance is placed on the landmark of the face outline in the simple and effective way.

6. Limitations and Future Work

We used the outline landmarks by which the success or failure of fitting can be determined in the relatively simple way. As mentioned above, since landmarks of the shape model are composed of geometric relations, we can estimate the fitting

state of the shape being tracked. Estimation of the fitting state might be helpful in verifying the reliability of the shape model used in various directions and determining the usage.

First, there is a case that uses it as a measure for improving the tracking of the shape model itself. The systems that are related to the shape model consisting of the landmarks are flowed in two steps. After the initial searching of a face area and the initial fitting process of shape, fitting process only occurs. If a landmark fails to fit the facial feature in tracking, it can be occur that a shape repeats the incorrect fitting. In this case, with analyzing the current status and returning to the initial phase of the system by using this research, the recovery is possible.

Also, it can be used as a correction of the shape fitting result. The fitting technique keeps track of each feature and maintains the natural form through the shape alignment. In this process, the outer landmark directly evaluates and finds the better fitting points. After verification of the shape, furthermore, it can be used in various systems to take advantage of the shape model. The systems of gaze tracking, face recognition, motion recognition, and facial expression recognition are applicable. The role of the landmark in the system is critical. If landmarks are incorrectly fitted, a more serious problem might occur in these systems.

Like this, the patch classifier can achieve the better results by applying after the shape fitting. In addition, by utilizing the failed data, it contributes to the way of escaping from the bigger problem.

Conflict of Interests

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

References

- [1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [2] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [3] D. Cristinacce and T. Cootes, "Automatic feature localisation with constrained local models," *Pattern Recognition*, vol. 41, no. 10, pp. 3054–3067, 2008.
- [4] X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1513–1520, IEEE, December 2013.
- [5] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D Face recognition under expressions, occlusions, and pose variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013.
- [6] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200–215, 2011.
- [7] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. V. Rohith, "Fully automatic pose-invariant face recognition via 3D pose normalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 937–944, November 2011.
- [8] C. Xiong, L. Huang, and C. Liu, "Gaze estimation based on 3D face structure and pupil centers," in *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR '14)*, pp. 1156–1161, IEEE, Stockholm, Sweden, August 2014.
- [9] C. Cao, Y. Weng, S. Lin, and K. Zhou, "3D shape regression for real-time facial animation," *ACM Transactions on Graphics*, vol. 32, no. 4, article 41, 2013.
- [10] B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. Ter Haar Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp. 924–933, 2002.
- [11] T. F. Cootes and C. J. Taylor, "Active shape models—'smart snakes'," in *BMVC92: Proceedings of the British Machine Vision Conference, organised by the British Machine Vision Association 22–24 September 1992 Leeds*, pp. 266–275, Springer, London, UK, 1992.
- [12] C. A. R. Behaine and J. Scharcanski, "Enhancing the performance of active shape models in face recognition applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2330–2333, 2012.
- [13] F. Dornaika and J. Ahlberg, "Fast and reliable active appearance model search for 3-D face tracking," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 4, pp. 1838–1853, 2004.
- [14] J. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.
- [15] A. U. Batur and M. H. Hayes, "Adaptive active appearance models," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1707–1721, 2005.
- [16] D. Cristinacce and T. Cootes, "Feature detection and tracking with constrained local models," in *Proceedings of the 17th British Machine Vision Conference (BMVC '06)*, pp. 929–938, September 2006.
- [17] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," *International Journal of Computer Vision*, vol. 107, no. 2, pp. 177–190, 2014.
- [18] M. D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool, and H. Pfister, "Real-time face pose estimation from single range images," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.
- [19] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 1701–1708, IEEE, Columbus, Ohio, USA, June 2014.
- [20] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012.
- [21] S. Hu, S. Sun, X. Ma, Y. Qin, and B. Lei, "A fast on-line boosting tracking algorithm based on cascade filter of multi-features," in *Proceedings of the 3rd International Conference on Multimedia Technology (ICM '13)*, Atlantis Press, Guangzhou, China, November 2013.
- [22] I. S. Msiza, F. V. Nelwamondo, and T. Marwala, "On the segmentation of fingerprint images: how to select the parameters of a block-wise variance method," in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICIPV '12)*, pp. 1107–1113, July 2012.

- [23] C. A. Glasbey and K. V. Mardia, "A review of image-warping methods," *Journal of Applied Statistics*, vol. 25, no. 2, pp. 155–171, 1998.
- [24] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition (FG '08)*, September 2008.
- [25] T. F. Cootes, C. J. Twining, V. Petrović, R. Schestowitz, and C. J. Taylor, "Groupwise construction of appearance models using piece-wise affine deformations," in *Proceedings of 16th British Machine Vision Conference (BMVC '05)*, vol. 5, pp. 879–888, 2005.

Research Article

Efficient Deep Neural Network for Digital Image Compression Employing Rectified Linear Neurons

Farhan Hussain and Jechang Jeong

Department of Electronics & Computer Engineering, Hanyang University, Seoul 133-791, Republic of Korea

Correspondence should be addressed to Jechang Jeong; jjeong@hanyang.ac.kr

Received 16 December 2014; Accepted 11 February 2015

Academic Editor: Wei Wu

Copyright © 2016 F. Hussain and J. Jeong. 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 compression technique for still digital images is proposed with deep neural networks (DNNs) employing rectified linear units (ReLU). We tend to exploit the DNNs capabilities to find a reasonable estimate of the underlying compression/decompression relationships. We aim for a DNN for image compression purpose that has better generalization property and reduced training time and support real time operation. The use of ReLUs which map more plausibly to biological neurons, makes the training of our DNN significantly faster, shortens the encoding/decoding time, and improves its generalization ability. The introduction of the ReLUs establishes an efficient gradient propagation, induces sparsity in the proposed network, and is efficient in terms of computations making these networks suitable for real time compression systems. Experiments performed on standard real world images show that using ReLUs instead of logistic sigmoid units speeds up the training of the DNN by converging markedly faster. The evaluation of objective and subjective quality of reconstructed images also proves that our DNN achieves better generalization as most of the images are never seen by the network before.

1. Introduction

Digital image compression plays an extremely important role in the transmission and storage of digital image data. Usually the amount of data associated with visual information is so large that its storage requires enormous memory and its transmission requires high bandwidths. Image compression is the process of effectively coding digital images to reduce the number of bits required to represent an image. This compression allows transmission of image at very low bandwidths and minimizes the space requirement for storage of this data. As consecutive frames of still images constitute the video data, algorithms designed for 2D still image compression can be effectively extended to compress video data. The image compression algorithms can be broadly classified into lossless and lossy compression algorithms. A lossless compression algorithm reproduces the original image exactly without any loss of information. These methods are used in applications where data loss is unacceptable, for example, text data and medical images. In this paper we are going to target the lossy

compression where the reproduced image is not an exact replica of the original image. Some information is lost in the coding process. These algorithms provide a way of tradeoff between the image quality and the degree of compression. The goal is to achieve higher degree of compression without much degradation in image quality.

Artificial neural networks (ANNs) [1–3] get their inspiration from the manner in which human brain performs calculations and makes decisions. In fact ANNs tend to imitate the functionality of human brain, that is, the biological neuron system. An ANN achieves this abstraction and modeling of the information processing capabilities of human brain by interconnecting a large number of simple processing elements called the artificial neurons. An artificial neuron is an electronically modeled biological neuron. The complexity of a real biological neuron is highly abstracted while modeling an artificial neuron. The theme is to mimic the working of human nervous system with the help of these artificial neurons. In order to achieve this each artificial neuron is

equipped with some computational strength similar to that possessed by a biological neuron. An artificial neuron like a biological neuron can take many input signals and then based on an internal weighting system produces a single output signal that is typically sent as input to another neuron. Figure 1 shows a representation of an artificial neuron.

Here X_0, X_1, \dots, X_n are the n input signals to the artificial neuron. Each of these input signals is multiplied by a connection weight. These weights are represented by W_0, W_1, \dots, W_n , respectively. All of these products are summed and fed to an activation function to generate an output for the artificial neuron. An ANN is set up by creating connections between these artificial neurons analogous to the connection between biological neurons in a human nervous system as shown in Figure 2. Recent developments in the processing capabilities of the CPU and advancement in the architecture of neural networks have attracted many researchers from various scientific disciplines to investigate these networks as a possible solution to solve various problems encountered in the fields of pattern recognition, prediction, optimization, function approximation, clustering, categorizing, and many more. ANNs are also being researched and developed to address the problem of compression of still images and video data [4–11] besides the conventional algorithms [12]. The target is to achieve high compression ratios, maximize the quality of reproduced images, and design systems that utilize minimum computational resources and support real time applications. In this work we attempt for these goals by a deep neural network [13], that is, an artificial neural network having multiple hidden layers of neurons between the input and output layers. These DNNs can model complex nonlinear relationships more efficiently. The extra layers of the DNNs enable the network to capture the highly variant features of the input and learn more higher abstract representations of the data. The common issues for the DNNs are that its training is time consuming and is computationally very expensive. In this work we propose a DNN for still image compression that has much reduced training time and is computationally efficient. Our DNN achieves this speed-up in training time and reduction in complexity by employing rectified linear units [14]. This arrangement also leads to better generalization of the network and reduces the real compression-decompression time. The network meets the targets guaranteed for a good compression system; it converges fast, does not have a vanishing gradient problem, and has a sparse representation (only part of the network is active for a given input). These characteristics make the proposed network suitable for real time compression systems.

The rest of the paper is organized as follows. Section 2 gives a brief overview of the literature concerning image compression via ANNs and states the preliminaries. Section 3 introduces the proposed DNN. Section 4 reports the experimental results. Section 5 provides the conclusion to the paper.

2. Background and Preliminaries

Several studies have been proposed to address the problem of digital image compression via artificial neural networks.

The most elemental and simple network, that is, the single structured neural network, is described in [4]. This network uses a 3-layer network with logistic transfer function to achieve image compression. Parallel architectures for neural networks for image compression are proposed in [5–7]. The idea is to compress different part of an image (i.e., according to some complexity) by different neural networks in parallel in order to increase the compression ratio and quality of reconstructed images. In [8] the authors propose the use of novel normalization function along with the single structured neural network to improve the compression quality. An ANN for calculating the discrete cosine transform for image compression is described in [9]. The authors in [10, 11] provide a summary of different neural network models and techniques such as vector quantization that neural networks can be complemented with to improve the compression results.

The process of image compression can be formulated as designing a compressor and decompressor module as shown in Figure 3, where I is the original image, I_c is the compressed data, and I' is the reconstructed image. The number of bits in the compressed data is much less than the number of bits required to represent the original image. Usually these modules are realized by heavy complex algorithms with a lot of complicated calculations involved. As described before many researches are carried out to approximate the image compression/decompression modules with the help of ANN by generating an internal data representation. These networks are trained on several training images (input-output pairs), the task being to approximate the corresponding compression-decompression algorithm compactly and model it so it can be generalized to a large set of test data. This is achieved by rigorously training the network with the help of learning algorithms. This training process is characterized by the use of a given output that is compared to the predicted output and by adaption of all parameters according to this comparison. The parameters of a neural network are its weights. This paper uses the back propagation for training the ANN. The back propagation [15] is a supervised learning algorithm and is especially suitable for feed-forward networks. The feed-forward neural networks refer to multilayer perceptron network in which the outputs from all the neurons go to following but not the preceding layers, so there are no feedback loops and the information flows in only one direction. The term back propagation is abbreviated for “backward prorogation of errors” and it implies that the errors (and therefore the learning) propagate backwards from the output nodes to the inner nodes. Hence back propagation is used to calculate the gradient of the error with respect to the network’s modifiable weights. This gradient is then used in a simple gradient descent algorithm to find weights that minimize the error. The term back propagation refers to the entire procedure encompassing both the calculation of gradient and its use in gradient descent. Back propagation requires that the transfer function used by the artificial neurons (or “nodes”) to be differentiable. Figure 2 shows a generalized architecture for an ANN. The neurons are tightly interconnected and organized into different layers. The input layer receives the input; the output layer produces the final

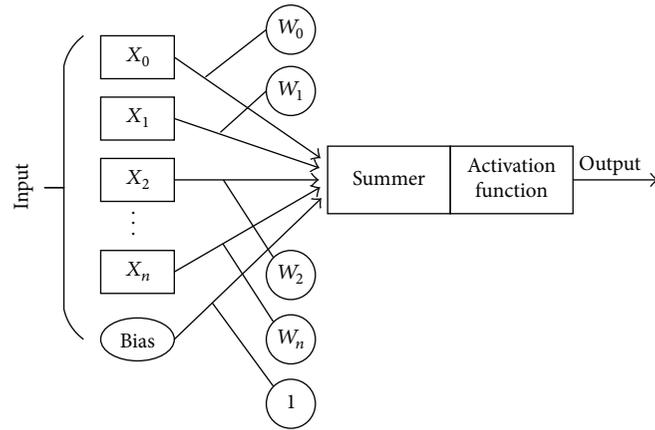


FIGURE 1: An artificial neuron.

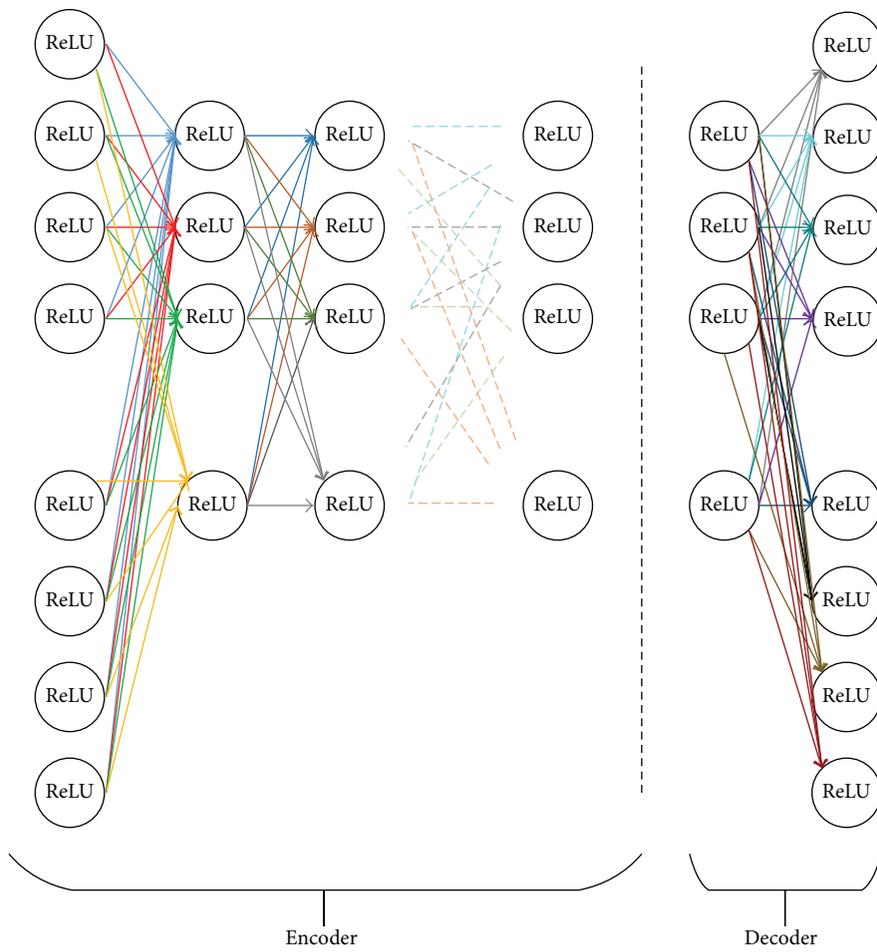


FIGURE 2: Deep neural network architecture for image compression.

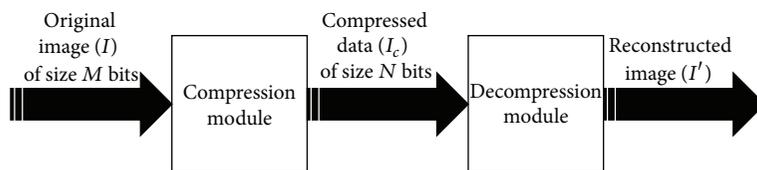


FIGURE 3: Block diagram of an image compression system.

output. Usually one or more hidden layers are sandwiched in between the two.

3. Proposed Deep Neural Network for Image Compression

In this paper digital image compression is achieved by a deep neural network that employs rectified linear units. The multiple hidden layers of the DNN are advantageous in realizing more efficient internal data representations of the underlying compression-decompression function. Researchers conclude that the biological neurons can be better approximated and modeled by the ReLUs. As these ReLUs are more biologically plausible they can be engaged as better activation functions than the widely used logistic sigmoid and hyperbolic tangent functions. The ReLU is given by

$$\text{rectifier}(x) = \max(0, x), \quad (1)$$

where x is the input to the neuron. The function behaves linearly for a favorably excitatory input pattern and “0” otherwise. As can be seen the function is one sided; it does not possess the property of sign symmetry or antisymmetry. The function is not differentiable at “0” and its derivative, whenever it exists, can take only two values, “0” or “1.” Although the ReLUs are not entirely differentiable, nor symmetric and possess hard linearity they can outperform the sigmoid and hyperbolic tangent neurons. Experiments show that the ReLUs are worth the tradeoff with these much sophisticated counter parts. They allow fast convergence and better generalization of the DNN. These ReLUs are computationally much cheaper. The efficient computations required for both its value and its partial derivatives enable much larger network implementations. Engaging ReLUs induces sparsity in the network; that is, only a subset of neurons are active in the hidden layers. This leads to faster computation and better learning. Other advantages of sparsity are discussed in [14]. The DNN with ReLUs used for still image compression is shown in Figure 2. The network consists of an input layer, n number of hidden layers, and an output layer. As this network is targeting image compression/decompression it must have equal number of input and output neurons, N (an N dimensional input is mapped to an N dimensional output). The number of neurons in the input layer or the output layer corresponds to the size of image block to be compressed. Compression can be achieved by allowing the number of neurons at the last hidden layer, K , to be less than that of the neurons at both the input and the output layers ($K < N$). The number of hidden layers and hidden neurons is determined by the number of input and output neurons as well as desired compression ratio. The compression ratio of this DNN is the ratio of input neurons to the number of neurons in the last hidden layers.

The training of the DNN is carried out with a set of images selected from the training set. The training images are divided into nonoverlapping blocks of size W by W pixels. The pixels in each of these blocks are normalized by a normalizing function f , usually from a grey scale value between 0 and 255 to a range of values between 0 and 1. These

normalized blocks are fed into the input layer of the DNN at random; each neuron in the input layer corresponds to one pixel; that is, $N = W \times W$. As in case of supervised learning the desired output for the network is known in advance, which in case of image compression purpose is the same as the input to the network. We tend to resolve the network to produce at the output what it sees at the input. As we use back propagation for training the DNN, the difference between the actual output and desired output is calculated and the error is propelled backwards to adjust the parameters of the network accordingly. With the new weights the output is again calculated and compared with the desired one, errors are repropagated, the parameters are readjusted, and the process continues in an iterative fashion. In our implementation of DNN the training is stopped when the iterations reach their maximum limit or when the average mean square error drops below a certain threshold. Once the training of the network is completed, the parameters of the network are saved. With these finalized weights we utilize this DNN to compress and decompress the test images. The test image to be compressed is divided into nonoverlapping blocks. Each block is fed into the input of the network after normalization. The input layer and the n hidden layers act as the compressor module and perform a nonlinear and nonorthogonal transformation S . The compressed data is found at the output of the last hidden layer. The output layer acts as the decompressor module and reconstructs the normalized input data block by performing a second transformation T . The decompressed image block can be found at the output neurons of the output layer. The dynamic range of the reconstructed data block is restored by applying the inverse normalization function f^{-1} . The transformations S and T are optimized by training the network on several training images.

4. Experimental Results

Experiments were performed on test images taken from the standard set of images: Lena, baboon, cameraman, pepper, and boats. The size of the test images was 512 by 512. The number of neurons in the input layer, output layer, and the last hidden layer was adjusted to achieve different compression ratios, that is, 4:1, 8:1, and 16:1. The epoch versus mean square error (mse) curves for the DNN employing the rectified linear neurons and the logistic sigmoid neurons were plotted for different compression ratios (CRs) and are shown in Figures 4–6.

A comparison of these plots shows that the DNN with ReLUs converges several orders of magnitude faster than the one with logistic sigmoid units. This fact is illustrated in Figure 7 which shows that to reduce the mse to 0.0019 the DNN with sigmoid neurons takes 500 epochs while the same mse is achieved by the DNN with ReLUs after 50 epochs for a CR of 4:1. Similarly it takes 500 epochs for the sigmoid neurons to reduce the mse to 0.0039 at a CR of 8:1 while the same is achieved after 35 epochs by employing ReLUs. This high convergence rate validates the fact that the network with ReLUs trains much faster than their logistic counterparts.

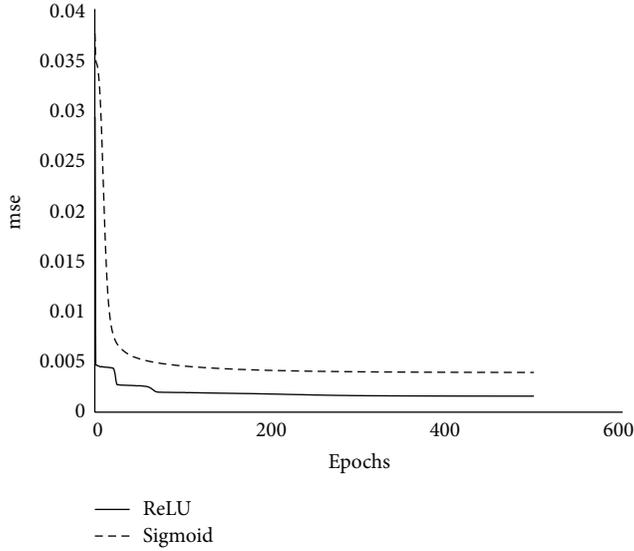


FIGURE 4: Comparison of epoch versus mse for rectified linear neurons and logistic sigmoid neurons at a CR of 8 : 1.

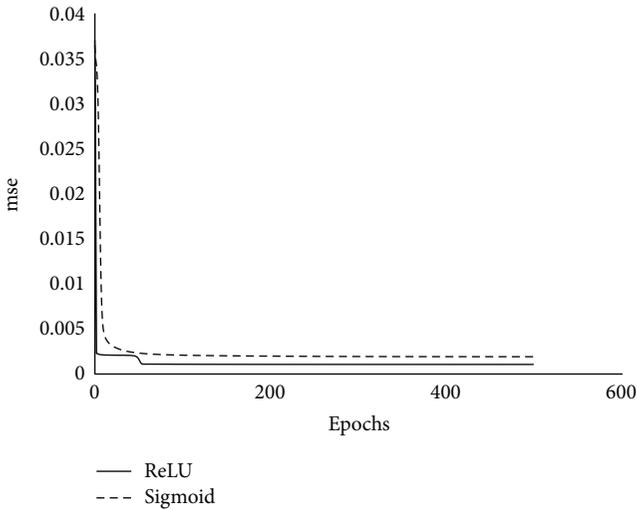


FIGURE 5: Comparison of epoch versus mse for rectified linear neurons and logistic sigmoid neurons at a CR of 4 : 1.

The performance of the network is measured by the peak signal to noise ratio (PSNR) of image reconstructed at the output layer and is defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right), \quad (2)$$

where MAX_I is the maximum possible pixel value of the image and MSE is the mean square error given by

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_O(i, j) - I_R(i, j)]^2, \quad (3)$$

where m and n represent the size of the image in the horizontal and vertical dimensions, respectively, I_O is the

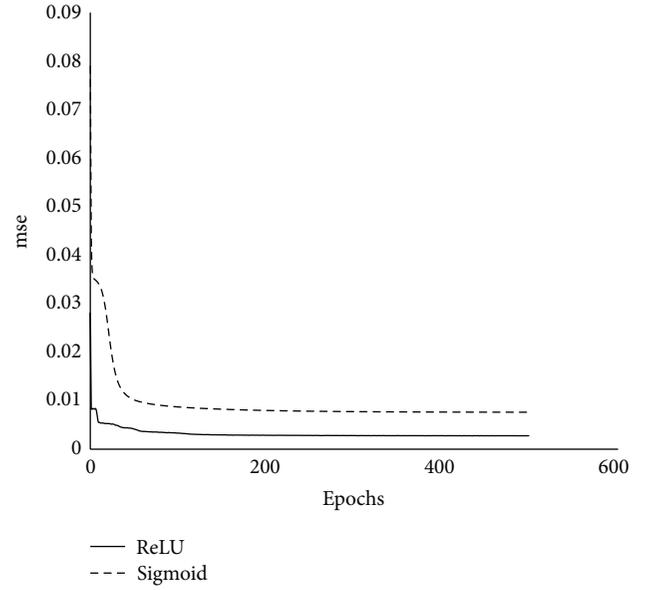


FIGURE 6: Comparison of epoch versus mse for rectified linear neurons and logistic sigmoid neurons at a CR of 16 : 1.

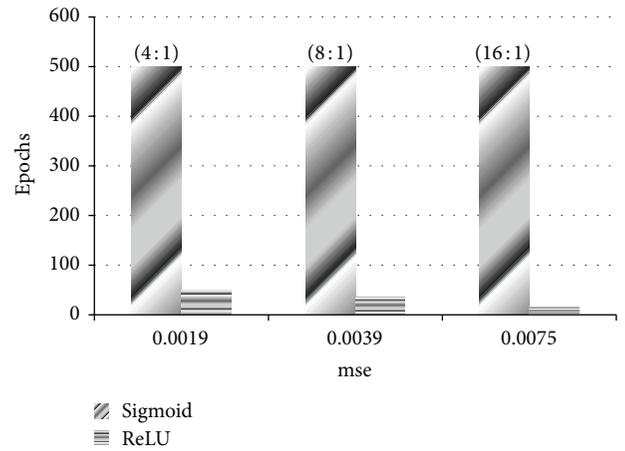


FIGURE 7: Epochs required to reduce the mean square error of the DNN using sigmoid neurons and rectified linear neurons at different CRs.

original image, and I_R is the reconstructed image. Tables 1–3 show the PSNR achieved by the set of five standard real world images compressed at different compression ratios by (i) the rectified linear DNN and (ii) the logistic sigmoid DNN. The DNNs were trained using the same data set and by the back propagation algorithm. There is a significant improvement in the PSNR of the reconstructed images for rectified linear units compared to the logistic units. This result proves better generalization ability of our network with ReLUs as none of the sequences in Tables 1–3 were used for the training of the DNN.

To evaluate the subjective quality of the decompressed images Figures 8–10 show some of the original images and their reconstructed counterparts.



FIGURE 8: (a) Original image, (b) image reconstructed by rectified deep neural network at a CR of 8 : 1, and (c) image reconstructed by deep neural network employing sigmoid transfer function at a CR of 8 : 1.



FIGURE 9: (a) Original Image, (b) image reconstructed by rectified deep neural network at a CR of 4 : 1, and (c) image reconstructed by deep neural network employing sigmoid transfer function at a CR of 4 : 1.

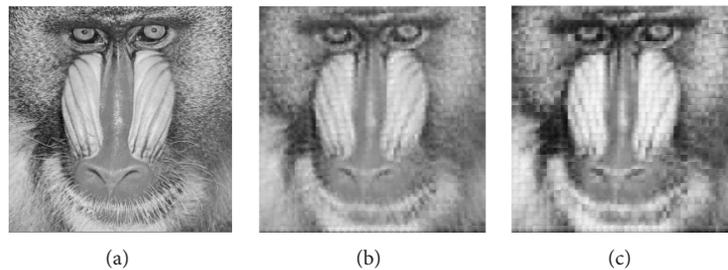


FIGURE 10: (a) Original image, (b) image reconstructed by rectified deep neural network at a CR of 16 : 1, and (c) image reconstructed by deep neural network employing sigmoid transfer function at a CR of 16 : 1.

TABLE 1: PSNR of the reconstructed images at a CR of 8 : 1.

Sequence	Compression scheme	
	ReLU	Sigmoid
Lena	26.20	24.37
Baboon	22.56	20.59
Cameraman	24.47	23.24
Peppers	24.02	25.54
Boats	26.31	24.54

TABLE 2: PSNR of the reconstructed images at a CR of 4 : 1.

Sequence	Compression scheme	
	ReLU	Sigmoid
Lena	30.48	27.30
Baboon	23.70	23.95
Cameraman	28.02	27.07
Peppers	28.40	28.46
Boats	26.68	26.13

5. Conclusions

In this paper we use a deep neural architecture for the purpose of still image compression. The proposed DNN learns the compression/decompression function very well. We advocate the use of ReLUs in this DNN as these units can be realized

by very simple functions. They greatly accelerate the convergence of the network, are computationally inexpensive, and have better learning characteristic. It has been shown experimentally that these networks train faster and generalize better; hence we argue that this network can be realized into real time compression systems. Future work on this study will

TABLE 3: PSNR of the reconstructed images at a CR of 16 : 1.

Sequence	Compression scheme	
	ReLU	Sigmoid
Lena	25.19	22.78
Baboon	21.25	20.06
Cameraman	24.54	21.53
Peppers	23.25	22.70
Boats	24.11	22.44

leverage implementing these DNNs on specialized hardware like GPUs and also extending this idea to the compression of video data.

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 MSIP (Ministry of Science, ICT and Future Planning), Republic of Korea, under the project for technical development of information communication and broadcasting (2014-044-057-001).

References

- [1] S. O. Haykin, *Neural Networks and Learning Machines*, Prentice Hall, 3rd edition, 2008.
- [2] Y. S. Abu-Mostafa, M. M. Ismail, and H.-T. Lin, *Learning from Data*, AMLBook, 2012.
- [3] P. H. Sydenham and R. Thorn, *Handbook of Measuring System Design*, vol. 3, John Wiley & Sons, Chichester, UK, 2005.
- [4] G. L. Sicuranza, G. Ramponi, and S. Marsi, "Artificial neural network for image compression," *Electronics Letters*, vol. 26, no. 3, pp. 477–479, 1990.
- [5] S. Carrato and S. Marsi, "Parallel structure based on neural networks for image compression," *Electronics Letters*, vol. 28, no. 12, pp. 1152–1153, 1992.
- [6] G. Qiu, M. R. Varley, and T. J. Terrell, "Image compression by edge pattern learning using multilayer perceptions," *Electronics Letters*, vol. 29, no. 7, pp. 601–603, 1993.
- [7] A. Namphol, S. H. Chin, and M. Arozullah, "Image compression with a hierarchical neural network," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 326–338, 1996.
- [8] Y. Benbenisti, D. Kornreich, H. B. Mitchell, and P. A. Schaefer, "A high performance single-structure image compression neural network," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 3, pp. 1060–1063, 1997.
- [9] K. S. Ng and L. M. Cheng, "Artificial neural network for discrete cosine transform and image compression," in *Proceedings of the 4th International Conference on Documents Analysis and Recognition*, vol. 2, pp. 675–678, August 1997.
- [10] C. Cramer, "Neural networks for image and video compression: a review," *European Journal of Operational Research*, vol. 108, no. 2, pp. 266–282, 1998.
- [11] J. Jiang, "Image compression with neural networks—a survey," *Signal Processing: Image Communication*, vol. 14, no. 9, pp. 737–760, 1999.
- [12] M. Rabbani and P. W. Jones, *Digital Image Compression Techniques*, SPIE Publications, 1991.
- [13] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [14] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural network," in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS '11)*, vol. 15 of *JMLR: W&CP*, pp. 315–323, Fort Lauderdale, Fla, USA, 2011.
- [15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," in *Neurocomputing: Foundations of Research*, MIT Press, Cambridge, Mass, USA, 1988.

Research Article

Modified Hybrid Freeman/Eigenvalue Decomposition for Polarimetric SAR Data

Shuang Zhang, Shuang Wang, and Bo Chen

*Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education,
International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi 710071, China*

Correspondence should be addressed to Shuang Zhang; shzhang_work@163.com

Received 25 November 2014; Accepted 30 November 2014

Academic Editor: Gwanggil Jeon

Copyright © 2015 Shuang 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.

Because of the rapid advancement of the airborne sensors and spaceborne sensors, large volumes of fully polarimetric synthetic aperture radar (PolSAR) data are available, but they are too complex to interpret difficultly. In this paper, a modified hybrid Freeman/eigenvalue decomposition method for the coherency matrix derived from the fully PolSAR sensors is proposed. The proposed modified hybrid Freeman/eigenvalue decomposition uses a real unitary transformation on the coherency matrix to release correlations between the copolarized term and cross polarized term, and the scattering models are derived from eigenvectors of the coherency matrix with reflection symmetry condition. The anisotropy and entropy are used to determine whether the volume scattering component is derived from the man-made structures or not. Moreover, the scattering powers from the proposed hybrid Freeman/eigenvalue decomposition are all nonnegative values. Fully PolSAR data on San Francisco acquired by AIRSAR sensor are used in the experiments to prove the efficacy of the proposed decomposition.

1. Introduction

Since 1985, the first fully polarimetric AIRSAR at L-band was launched by the Jet Propulsion Laboratory (JPL) [1] which started a rapid advancement stage of PolSAR sensors, such as the well-known airborne sensors: EMISAR [2] by the Technical University of Denmark and E-SAR [3] by DLR of Germany, and the famous spaceborne sensors: TerraSAR-X [4] by DLR of Germany and Radarsat-2 [5] by CSA of Canada and so forth. Large volumes of the fully PolSAR data need to be interpreted. Polarimetric target decomposition is an important and useful tool for understanding the PolSAR data [6] by separating received measurements into basic scattering mechanisms. The scattering mechanisms of the PolSAR media are analyzed for the purpose of parameter inversion, terrain classification, and so forth. Currently, eigenvector-based decompositions and model-based decomposition methods are commonly used on the second-order statistics matrix of the PolSAR data. Cloude and Pottier developed the most notable eigenvector-based decomposition, that is, Entropy/Alpha method [7]. The classical model-based decomposition was Freeman-Durden decomposition (FDD)

developed by Freeman and Durden [8], which decomposed the coherency matrix of PolSAR data into three components: surface scattering, double-bounce scattering, and volume scattering with the reflection symmetry condition. The reflection symmetry condition implies that the correlation between copolarized term and cross polarized term is zero. Freeman-Durden decomposition is used in various applications since it is easy to understand and accomplish.

When applying Freeman-Durden decomposition on the real PolSAR data, some scattering powers are negative those most frequently occur in the double-bounce scattering powers. To solve this deficiency, various modified methods have been developed [9–15]. Cloude improved Freeman-Durden decomposition via setting surface scattering model and double-bounce scattering model to be orthogonal [14], that is, well-known hybrid Freeman/eigenvalue decomposition. After the rotation of the coherency matrix, the scattering powers derived from hybrid Freeman/eigenvalue decomposition are effective at avoiding negative values. Singh improved the hybrid Freeman/eigenvalue decomposition by using different volume scattering models [15] for vegetation areas and oriented structures.

In this paper, a modified version of hybrid Freeman/eigenvalue decomposition for PolSAR data is proposed by using different volume scattering models for the vegetation areas and man-made structures. Moreover, the surface scattering model and double-bounce scattering model are defined as the eigenvectors of the coherency matrix of PolSAR data. The eigenvector with scattering angle α greater than $\pi/4$ denotes the double-bounce scattering model, while the other scattering angle α less than $\pi/4$ represents the surface scattering model. The volume scattering model for the man-made structures is also derived from the eigenvectors of the coherency matrix. We will show how the eigenspace of the coherency matrix enables the proposed Freeman/eigenvalue decomposition and solve the scattering powers as the linear combinations of eigenvalues. In addition, the scattering powers are all nonnegative values.

The rest of this paper is organized as follows. The hybrid Freeman/eigenvalue decomposition is summarized in Section 2. The proposed decomposition is presented in Section 3. Results and discussion of experiments performed on the real PolSAR data are provided in Section 4. Section 5 presents our conclusions.

2. Freeman/Eigenvalue Decomposition

In this section, the Freeman/eigenvalue decomposition is simply introduced for the integrity of this paper. The detail content has been shown in [14].

For monostatic PolSAR sensor with $\{H, V\}$ basis, a Pauli vector is used to represent the single look PolSAR data as (1) and the coherency matrix is used to represent the multilook PolSAR data as (2) as follows:

$$\vec{k}_p = \frac{1}{\sqrt{2}} [S_{HH} + S_{VV} \quad S_{HH} - S_{VV} \quad 2S_{HV}]^t \quad (1)$$

$$\langle [T] \rangle = \langle \vec{k}_p \cdot \vec{k}_p^{*t} \rangle = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{12}^* & T_{22} & T_{23} \\ T_{13}^* & T_{23}^* & T_{33} \end{bmatrix}, \quad (2)$$

where t denotes a transposition operator, $*$ implies a complex conjugation processing, and $\langle \rangle$ denotes the multilook processing.

In the model-based decomposition, such as Freeman-Durden decomposition [8] and hybrid Freeman/eigenvalue decomposition [14], the measured coherency matrix of the PolSAR data is expanded into three components, surface scattering, double-bounce scattering, and volume scattering as follows:

$$\langle [T] \rangle = m_s T_s + m_d T_d + m_v T_v, \quad (3)$$

where T_s , T_d , and T_v are surface scattering model, double-bounce scattering model, and volume scattering model, respectively.

In hybrid Freeman/eigenvalue decomposition, those three scattering models are defined in (4), (5), and (6).

m_s , m_d , and m_v are the corresponding scattering powers in the proper order:

$$T_s = \begin{bmatrix} \cos^2 \alpha_s & \cos \alpha_s \sin \alpha_s e^{-j\phi_s} & 0 \\ \cos \alpha_s \sin \alpha_s e^{j\phi_s} & \sin^2 \alpha_s & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \alpha_s \leq \frac{\pi}{4}, \quad (4)$$

$$T_d = \begin{bmatrix} \cos^2 \alpha_d & \cos \alpha_d \sin \alpha_d e^{-j\phi_d} & 0 \\ \cos \alpha_d \sin \alpha_d e^{j\phi_d} & \sin^2 \alpha_d & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \alpha_d > \frac{\pi}{4}, \quad (5)$$

$$T_v = \begin{bmatrix} F_s & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

α_s and α_d imply the type of the scattering model; namely, $\alpha_s \leq \pi/4$ denotes the surface scattering, while $\alpha_d > \pi/4$ represents the double-bounce scattering. In [14], the condition $\alpha_s + \alpha_d = \pi/2$ is set to reduce the number of unknowns. F_s is the volume parameter. If $F_s = 2$, hybrid Freeman/eigenvalue decomposition becomes equivalent to Freeman-Durden decomposition [8]. The cross polarized term T_{33} only exists in the volume scattering model, so the volume scattering power m_d is solved as follows:

$$m_v = T_{33}. \quad (7)$$

Since the rank of either surface scattering model T_s or double-bounce scattering model T_d is equal to 1, the corresponding scattering powers m_s and m_d are the eigenvalues of T_{SD} as (10), and m_s and m_d can be solved as (11). α_s and α_d are obtained as follows:

$$\alpha_{d,s} = \cos^{-1} \left[\left(1 + \left| \frac{T_{12}}{T_{22} - T_{33} - m_{d,s}} \right|^2 \right)^{-1/2} \right]. \quad (8)$$

Whichever of m_s and m_d is smaller is always set to zero, and then volume parameter F_s can be estimated as

$$F_s = \frac{T_{11} \cdot (T_{22} - T_{33}) - |T_{12}|^2}{T_{33} \cdot (T_{22} - T_{33})}. \quad (9)$$

It enables solving for the scattering powers m_s and m_d by substituting (9) into (11):

$$\begin{aligned} T_{SD} &= \langle [T] \rangle - m_v T_v \\ &= \begin{bmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_s & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (10)$$

$$\begin{aligned}
& m_{d,s} \\
& = \left((T_{11} + T_{22} - (F_s + 1) T_{33}) \right. \\
& \quad \left. \pm \sqrt{(T_{11} - T_{22} - (F_s - 1) T_{33})^2 + 4 |T_{12}|^2} \right) \\
& \quad \cdot (2)^{-1} \\
& \langle [T(\theta)] \rangle \\
& = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta \\ 0 & -\sin 2\theta & \cos 2\theta \end{bmatrix} \langle [T] \rangle \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & -\sin 2\theta \\ 0 & \sin 2\theta & \cos 2\theta \end{bmatrix}. \tag{11}
\end{aligned}$$

3. Proposed Decomposition

3.1. Rotation of the Coherency Matrix. To reduce the power of cross polarized term T_{33} , a real unitary transformation has been proposed to be implemented on the coherency matrix [12, 16] before decomposition as (12).

θ is the angle of rotation about the radar line of sight. To minimize T_{33} , two angles are solved as

$$\theta = \frac{1}{4} \tan^{-1} \left(\frac{2 (\operatorname{Re}(T_{23}))}{T_{22} - T_{33}} \right). \tag{13}$$

After the rotation of the coherency matrix, the imagery element of T_{23} is equal to zero.

3.2. Eigenvalue Decomposition of the Rotated Coherency Matrix. We assume that the reflection symmetry condition (i.e., $\langle S_{HH} S_{HV}^* \rangle \approx \langle S_{VV} S_{HV}^* \rangle \approx 0$) holds; then each pixel for multilook PolSAR data is represented as

$$\langle [T(\theta)] \rangle = \begin{bmatrix} T_{11}(\theta) & T_{12}(\theta) & 0 \\ T_{12}^*(\theta) & T_{22}(\theta) & 0 \\ 0 & 0 & T_{33}(\theta) \end{bmatrix}. \tag{14}$$

The average coherency matrix after rotation is decomposed [7] as

$$\langle [T(\theta)] \rangle = U \cdot \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \cdot U^{*t}, \tag{15}$$

where λ_1 , λ_2 , and λ_3 are the eigenvalues of the rotated coherency matrix $\langle [T(\theta)] \rangle$, and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$. Because of the reflection symmetry condition and the rotation of the coherency matrix, U consists of the corresponding eigenvectors k_1 , k_2 , and k_3 as

$$U = [k_1 \ k_2 \ k_3] = \begin{bmatrix} \cos \alpha_1 & \cos \alpha_2 & 0 \\ \sin \alpha_1 e^{j\gamma_1} & \sin \alpha_2 e^{j\gamma_2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{16}$$

where eigenvectors k_1 , k_2 , and k_3 are orthogonal unit vectors, $\alpha_1 + \alpha_2 = \pi/2$.

3.3. Scattering Models. Because in hybrid Freeman/eigenvalue decomposition, the condition $\alpha_s + \alpha_d = \pi/2$ is set, and, in eigenvectors of the coherency matrix, it can be seen that $\alpha_1 + \alpha_2 = \pi/2$, we can draw the conclusions: the eigenvectors k_1 and k_2 contain a surface scattering component and a double-bounce scattering component, and the scattering angles α_1 and α_2 are equivalent to the angles α_s and α_d of the hybrid Freeman/eigenvalue decomposition. The eigenvector k_1 or k_2 in U with $\alpha \leq \pi/4$ is shown in (17) which denotes the surface scattering targets, and the other eigenvector with $\pi/4 \leq \alpha \leq \pi/2$ is a double-bounce scattering target (18) under the reflection symmetry condition:

$$k_s = \begin{bmatrix} \cos \alpha_s \\ \sin \alpha_s e^{j\gamma_s} \\ 0 \end{bmatrix}, \quad \alpha_s \leq \frac{\pi}{4}, \tag{17}$$

$$k_d = \begin{bmatrix} \cos \alpha_d \\ \sin \alpha_d e^{j\gamma_d} \\ 0 \end{bmatrix}, \quad \frac{\pi}{4} \leq \alpha_d \leq \frac{\pi}{2}. \tag{18}$$

If $\alpha_1 \leq \pi/4$, we obtain $\alpha_s = \alpha_1$, the corresponding eigenvector $k_s = k_1$ as (17), and surface scattering model $T_s = k_1 k_1^{*t}$ whose form is the same as the one in (4); $\alpha_d = \pi/2 - \alpha_1 = \alpha_2$ and double-bounce scattering model $T_d = k_2 k_2^{*t}$ is identical to (5). In other cases, if $\alpha_2 \leq \pi/4$, $\alpha_s = \alpha_2$, surface scattering model $T_s = k_2 k_2^{*t}$ as (4), $\alpha_d = \pi/2 - \alpha_2 = \alpha_1$, and double-bounce scattering $T_d = k_1 k_1^{*t}$ as (5).

3.3.1. Volume Scattering Model from the Man-Made Structures. The entropy and anisotropy are defined in the eigenspace of the coherency matrix to measure the randomness of the targets [7, 17]:

$$H = - \sum_{i=1,2,3} p_i \log_3(p_i),$$

$$\text{with } p_i = \frac{\lambda_i}{(\lambda_1 + \lambda_2 + \lambda_3)}, \tag{19}$$

$$A = \frac{(\lambda_2 - \lambda_3)}{(\lambda_2 + \lambda_3)}.$$

For the man-made structures, usually the value of entropy (H) is large (usually $H > 0.7$), and anisotropy is often larger than 0.5 [7, 17]. Based on these, we defined the volume scattering model as

$$T = \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{20}$$

The rank of the volume scattering model is equal to two. The volume scattering model is designed as a distributed target that is diffused from the surface scatterers whose scattering angle $\alpha = 0$ and orientation angle $\beta = 0$ and the oriented objects whose scattering angle $\alpha = 90$ and orientation angle $\beta = 90$. For the real PolSAR data, the oriented objects can be modeled as angle $\alpha = 90$ and orientation angle $\beta = 90$ because of the reflection symmetry

condition. But the surface scatterers usually are not the same as “ $\alpha = 0$ and $\beta = 0$,” so the surface scattering case is relaxed as (17), and the volume scattering model for the man-made structures is defined as

$$T_{v,2} = \frac{1}{2} (k_s k_s^{*t} + k_3 k_3^{*t}). \quad (21)$$

3.3.2. Volume Scattering Model from the Vegetation Areas. The volume scattering model is used as a unit diagonal matrix whose rank is equal to 3 which has been proposed by An et al. [12]. Because the unit diagonal matrix has the largest entropy ($H = 1$) and least anisotropy ($A = 0$), the unit diagonal matrix is also used for the vegetation areas in this paper:

$$T_{v,1} = \frac{1}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (22)$$

3.4. Freeman/Eigenvalue Decomposition of the Rotated Coherency Matrix. For vegetation areas, the cross polarized term only presents in the volume scattering model, and volume scattering power can be solved as $m_v = 3T_{33}(\theta)$. In the eigenspace of the rotated coherency matrix under the reflection symmetry condition, the cross polarized term $T_{33}(\theta)$ is one of the eigenvalues. After the real unitary transformation, $T_{33}(\theta)$ reaches the minimum value, usually, $T_{33}(\theta) = \lambda_3$. So the volume scattering power is solved as

$$m_v = 3\lambda_3. \quad (23)$$

The surface scattering power m_s and double-bounce scattering power m_d are the eigenvalues of T_{SD} . According to the eigenvalue decomposition of the rotated coherency matrix with reflection symmetry condition, T_{SD} can be expanded as

$$\begin{aligned} \langle [T(\theta)] \rangle_{SD} &= m_s T_s + m_d T_d \\ &= \langle [T(\theta)] \rangle - m_v T_{v,1} \\ &= \langle [T(\theta)] \rangle - 3\lambda_3 T_{v,1} \\ &= (\lambda_1 - \lambda_3) k_1 k_1^{*t} + (\lambda_2 - \lambda_3) k_2 k_2^{*t}. \end{aligned} \quad (24)$$

From (23), $\alpha_1 \leq \pi/4$, the surface scattering power m_s , and double-bounce scattering power m_d are solved as (25) while $\alpha_2 \leq \pi/4$, m_s , and m_d are solved as (26):

$$m_s = \lambda_1 - \lambda_3, \quad (25)$$

$$\begin{aligned} m_d &= \lambda_2 - \lambda_3, \\ m_s &= \lambda_2 - \lambda_3, \\ m_d &= \lambda_1 - \lambda_3. \end{aligned} \quad (26)$$

For man-made structures, that is, $H < 0.7$ and $A > 0.5$ [7, 17], similar to the case of vegetation areas, the volume scattering power is solved as

$$m_v = 2\lambda_3. \quad (27)$$

Expand the coherency matrix after the real unitary transformation into eigenspace:

$$\begin{aligned} \langle [T(\theta)] \rangle &= m_s T_s + m_d T_d + m_v T_v \\ &= \langle [T(\theta)] \rangle_{SD} + m_v T_{v,1} \\ &= \langle [T(\theta)] \rangle_{SD} + 2 \cdot \lambda_3 \cdot \frac{1}{2} (k_s k_s^{*t} + k_3 k_3^{*t}) \\ &= \lambda_1 k_1 k_1^{*t} + \lambda_2 k_2 k_2^{*t} + \lambda_3 k_3 k_3^{*t} \\ &= (\lambda_1 - \lambda_2) k_1 k_1^{*t} + \lambda_2 k_2 k_2^{*t} \\ &\quad + 2 \cdot \lambda_3 \cdot \frac{1}{2} (k_1 k_1^{*t} + k_3 k_3^{*t}) \\ &= \lambda_1 k_1 k_1^{*t} + (\lambda_2 - \lambda_3) k_2 k_2^{*t} \\ &\quad + 2 \cdot \lambda_3 \cdot \frac{1}{2} (k_2 k_2^{*t} + k_3 k_3^{*t}). \end{aligned} \quad (28)$$

If $\alpha_1 \leq \pi/4$, it can be seen that, $\alpha_s = \alpha_1$, $k_s = k_1$, then the surface scattering model $T_s = k_1 k_1^{*t}$, and the volume scattering model $T_{v,2} = (1/2)(k_1 k_1^{*t} + k_3 k_3^{*t})$. In addition, $\alpha_d = \alpha_2$, $k_d = k_2$, and double-bounce scattering model $T_d = k_2 k_2^{*t}$. From the fifth line of (28), we can solve surface scattering powers m_s and double-bounce powers m_d as

$$\begin{aligned} m_s &= \lambda_1 - \lambda_3, \\ m_d &= \lambda_2. \end{aligned} \quad (29)$$

If $\alpha_2 \leq \pi/4$, then $\alpha_s = \alpha_2$, $k_s = k_2$, $T_s = k_2 k_2^{*t}$, and $\alpha_d = \alpha_1$, $T_d = k_1 k_1^{*t}$. The volume scattering model is $T_{v,2} = (1/2)(k_2 k_2^{*t} + k_3 k_3^{*t})$. From the last line of (28), we can solve surface scattering powers m_s and double-bounce powers m_d as

$$\begin{aligned} m_s &= \lambda_2 - \lambda_3, \\ m_d &= \lambda_1. \end{aligned} \quad (30)$$

Because of $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ and because the scattering powers are solved as (23), (25), and (26) or (27), (29), and (30), the scattering powers are all nonnegative values.

4. Experimental Study

To prove the efficacy of proposed Freeman/eigenvalue decomposition, the experiments are conducted on the L-band fully PolSAR data of San Francisco which were acquired by NASA/JPL ARISAR. The spatial resolution and range resolution are both about 10 m and the radar incidence angle is from 5° to 60° . The used PolSAR data is open access from the internet [18]. The original image is shown in Figure 1, with the selected regions, which are used in the later tests. The size of the used PolSAR image in these experiments is 900×1024 . Before expanding the coherency matrix, the Sigma filter is used to process the speckle of the PolSAR data.

In order to demonstrate the efficacy of the proposed hybrid Freeman/eigenvalue decomposition, the compared



FIGURE 1: Original image of San Francisco.

TABLE 1: Means of dominated scattering powers in Region 1.

	$mean_m_s$	$mean_m_d$	$mean_m_v$
FDD 1	0.9192	0.0013	0.0794
FDD 2	0.9475	0.0033	0.0492
HFED 1	0.9348	0.0164	0.0488
HFED 2	0.9348	0.0164	0.0488
Proposed method	0.9634	0.0053	0.0313

methods used Freeman-Durden decomposition [8] (FDD 1), Freeman-Durden decomposition with the rotation of the coherency matrix [12] (FDD 2), hybrid Freeman/eigenvalue decomposition [14] (HFED 1), and hybrid Freeman/eigenvalue decomposition with extended volume scattering model [15] (HFED 2). The PolSAR data are decomposed into three components: surface scattering power m_s (blue), double-bounce scattering power m_d (red), and volume scattering power m_v (green) in Figure 2.

From Figures 2(a)–2(e), three main terrain types in the used data, that is, ocean areas, city blocks, and forests are reconstructed well by the scattering powers derived from these five decompositions. For further analysis, we compare the scattering powers in three selected regions in Figure 1. Those regions are Region 1, Region 2, and Region 3, respectively. The sizes of these regions are 60×100 , 70×70 , and 60×100 in proper order. The types of ground truth are ocean areas, city blocks, and forests. The mean values of surface scattering power $mean_m_s$, double-bounce scattering power $mean_m_d$, and volume scattering power $mean_m_v$ in the three regions are listed in Tables 1, 2, and 3. The mean scattering powers are all normalized by the total scattering powers (i.e., $m_s + m_d + m_v$). In Region 1, it can be seen that $mean_m_s$ given by the proposed decomposition is 0.9634, which is about 4.8%, 1.7%, 3.1%, and 3.1% higher than the other four decompositions, respectively. The efficacy of all the five decompositions on Region 1 (sea) provides the excellent performance. In Region 2, the average double-bounce scattering power $mean_m_d$ of the proposed decomposition also outperforms the other methods. It is about 3.9%, 2.8%, 3.0%, and 0.6% larger than these four decompositions in proper sequence. But, in Region 3, $mean_m_v$ given by HFED 1 is the best.

TABLE 2: Means of dominated scattering powers in Region 2.

	$mean_m_s$	$mean_m_d$	$mean_m_v$
FDD 1	0.1082	0.4307	0.4611
FDD 2	0.1676	0.5017	0.3307
HFED 1	0.1686	0.5008	0.3306
HFED 2	0.3245	0.5137	0.1618
Proposed method	0.2404	0.5158	0.2438

TABLE 3: Means of dominated scattering powers in Region 3.

	$mean_m_s$	$mean_m_d$	$mean_m_v$
FDD 1	0.0314	0.1535	0.8151
FDD 2	0.0247	0.1555	0.8198
HFED 1	0.0673	0.1098	0.9228
HFED 2	0.0996	0.1097	0.7904
Proposed method	0.2380	0.0754	0.6865

TABLE 4: Classification results (%).

	Region 1	Region 2	Region 3
FDD 1	100	36.49	95.97
FDD 2	100	84.65	99.32
HFED 1	100	84.82	99.47
HFED 2	100	86.63	95.50
Proposed method	100	95.92	98.05

We classified the selected regions into three classes. The rule is simple; that is, the max scattering power determines the label of the pixel. If the surface scattering power is the biggest in the three scattering powers, the pixel is labeled “surface scattering class.” It can be seen that the true labels in Region 1, Region 2, and Region 3 are surface scattering class, double-bounce scattering class, and volume scattering class, respectively. The accuracy of the selected zones is listed in Table 4. It can be seen, in Region 1, that surface scattering powers of these decompositions are all the biggest ones, and, in Region 3, these classification accuracies are all higher than 95%. In contrast to the other decompositions, the average scattering powers of the proposed method in Region 3 are not very good, but the classification accuracy is higher than 98%. Moreover, in Region 2, the proposed method obtains the best results, as large as 95.92%, which are about 59.43%, 11.72%, 11.10%, and 9.29% higher than FDD 1, FDD 2, HFED 1, and HFED 2, respectively.

5. Conclusions

In this paper, a novel version of hybrid Freeman/eigenvalue decomposition for polarimetric SAR data is proposed. Three conclusions can be drawn about the proposed method. Firstly, the eigenvectors of the rotated coherency matrix with reflection symmetry condition are used as surface scattering or double-bounce scattering models. Secondly, in contrast to conventional Freeman-Durden decompositions, the derived scattering powers are all nonnegative values. Thirdly,

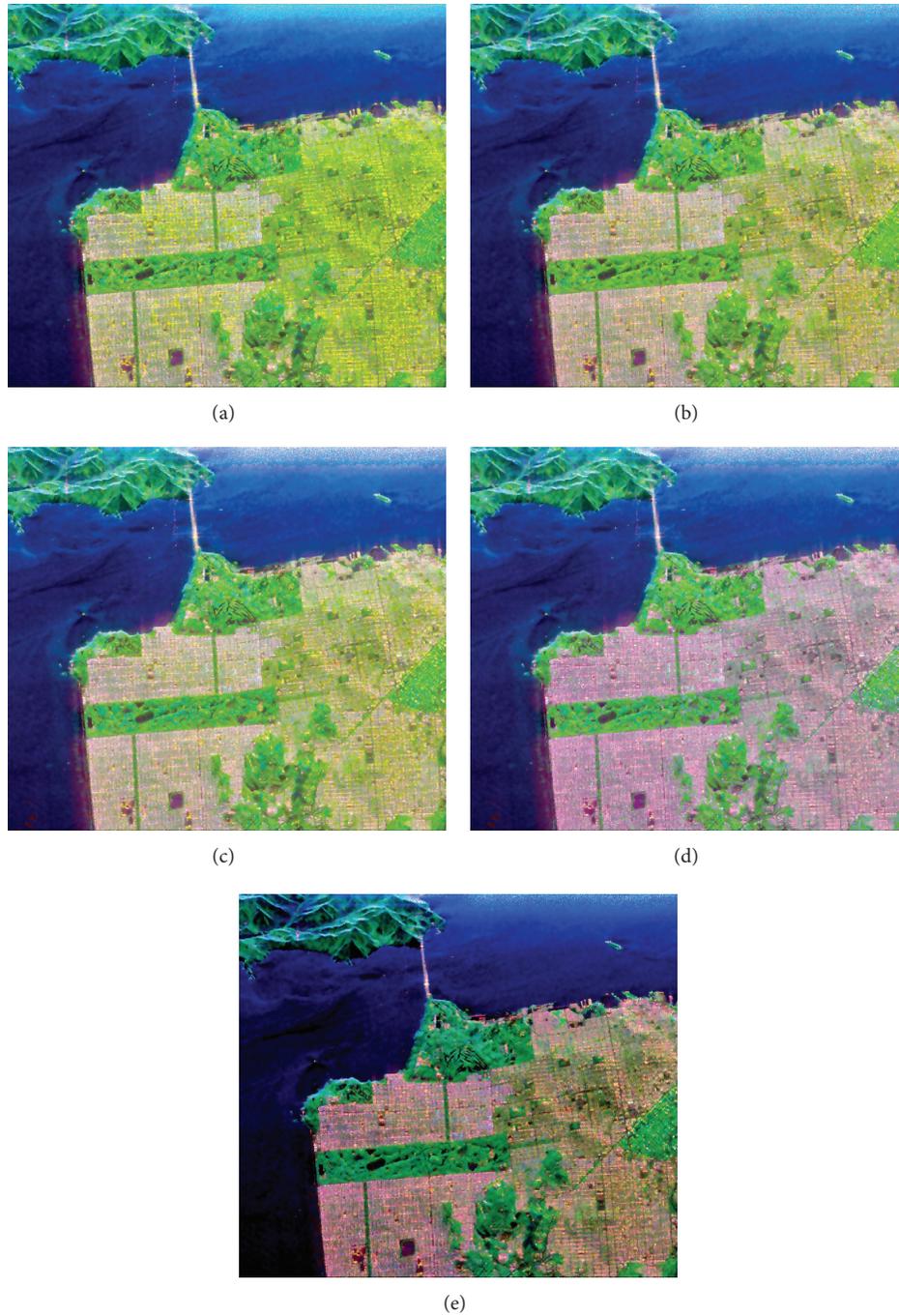


FIGURE 2: Decompositions of AIRSAR data on San Francisco with m_s for blue, m_d for red, m_v for green. (a) FDD 1, (b) FDD 2, (c) HFED 1, (d) HFED 2, and (e) proposed decomposition.

the volume scattering model is determined by the entropy and anisotropy of the coherency matrix. Experimental results have proved the efficacy of the proposed decomposition.

Conflict of Interests

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

Acknowledgments

This work was supported in part by the National High Technology Research and Development Program (863 Program) of China (no. 2013CB329402), the National Natural Science Foundation of China (nos. 61173092, 61377011, and 61272282), and the Fund for Foreign Scholars in University Research and Teaching Programs (no. B07048).

References

- [1] J. J. van Zyl, R. Carande, Y. Lou, T. Miller, and K. Wheeler, "The NASA/JPL three-frequency polarimetric AIRSAR system," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS '92)*, vol. 1, pp. 649–651, Houston, Tex, USA, May 1992.
- [2] E. L. Christensen, N. Skou, J. Dall et al., "EMISAR: an absolutely calibrated polarimetric L- and C-band SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 6, pp. 1852–1865, 1998.
- [3] R. Horn, "The DLR airborne SAR project E-SAR," in *Proceedings of the International Geoscience and Remote Sensing Symposium "Remote Sensing for a Sustainable Future" (IGARSS '96)*, vol. 3, pp. 1624–1628, Lincoln, Neb, USA, May 1996.
- [4] S. Buckreuss, R. Werninghaus, and W. Pitz, "The German satellite mission TerraSAR-X," in *Proceedings of the IEEE Radar Conference (RADAR '08)*, pp. 1–5, Rome, Italy, May 2008.
- [5] <http://www.asc-csa.gc.ca/eng/satellites/radarsat2>.
- [6] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 498–518, 1996.
- [7] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 68–78, 1997.
- [8] A. Freeman and S. L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 963–973, 1998.
- [9] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 8, pp. 1699–1706, 2005.
- [10] Y. Yamaguchi, A. Sato, W.-M. Boerner, R. Sato, and H. Yamada, "Four-component scattering power decomposition with rotation of coherency matrix," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2251–2258, 2011.
- [11] O. Antropov, Y. Rauste, and T. Hame, "Volume scattering modeling in PolSAR decompositions: study of ALOS PALSAR data over boreal forest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3838–3848, 2011.
- [12] W. An, Y. Cui, and J. Yang, "Three-component model-based decomposition for polarimetric sar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2732–2739, 2010.
- [13] G. Singh, Y. Yamaguchi, and S.-E. Park, "General four-component scattering power decomposition with unitary transformation of coherency matrix," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 3014–3022, 2013.
- [14] S. R. Cloude, *Polarisation: Applications in Remote Sensing*, Oxford University Press, London, UK, 2009.
- [15] G. Singh, Y. Yamaguchi, S.-E. Park, Y. Cui, and H. Kobayashi, "Hybrid freeman/eigenvalue decomposition method with extended volume scattering model," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 1, pp. 81–85, 2013.
- [16] F. Xu and Y.-Q. Jin, "Deorientation theory of polarimetric scattering targets and application to terrain surface classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2351–2364, 2005.
- [17] J. S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, CRC Press, Boca Raton, Fla, USA, 2009.
- [18] <https://earth.esa.int/web/polsarpro/datasets.html>.

Research Article

Superresolution of Hyperspectral Image Using Advanced Nonlocal Means Filter and Iterative Back Projection

Jin Wang,¹ Zhensen Wu,¹ and Young-Sup Lee²

¹*School of Physics and Optoelectronic Engineering, Xidian University, Xi'an, Shaanxi, China*

²*Department of Embedded Systems Engineering, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon 406-772, Republic of Korea*

Correspondence should be addressed to Young-Sup Lee; ysl@incheon.ac.kr

Received 27 September 2014; Accepted 4 November 2014

Academic Editor: Marco Anisetti

Copyright © 2015 Jin 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.

We introduce an efficient superresolution algorithm based on advanced nonlocal means (NLM) filter and iterative back projection for hyperspectral image. The nonlocal means method achieves the to-be-interpolated pixel by the weighted average of all pixels within an image, and the unrelated neighborhoods are automatically eliminated by the trivial weights. However, spatial location distance is also an important issue to reconstruct the missing pixel. Therefore, we proposed an advanced NLM (ANLM) filter considering both neighborhood similarity and patch distance. In the conventional NLM method, the search region was the whole image, while the proposed ANLM utilizes the limited search to reduce the complexity. The iterative back projection (IBP) is a very famous method to deal with the image restoration. In the superresolution issue, IBP is able to recover the high-resolution image iteratively from the given low-resolution image which is blurred due to the noise by minimizing the reconstruction error, while, because the reconstruction error of IBP is back projection and isotropic, the conventional IBP suffers from jaggy and ringing artifacts. Introducing the ANLM method to improve the visual quality is necessary.

1. Introduction

Recently, the space technology developed very fast and it could take the visual observation of earth from the artificial satellites. Hyperspectral image obtains terrestrial information in various contiguous and narrow spectral bands. Hyperspectral image is attracting more and more researchers due to its wide applications in aerial and space imagery fields, agriculture, ecology, geology, medicine, and meteorology. Hyperspectral image has a high resolution in spectral domain, while the resolution in spatial domain is limited. Moreover, when we acquire the hyperspectral image, many issues degrade the quality of hyperspectral image such as atmospheric scattering, secondary illumination, and sensor noise. Therefore, improving spatial resolution is a key issue in the hyperspectral image applications and high resolution can make various applications easier. However, it is very expensive to change the resolution in the hardware way. Thus, it is necessary to develop the technology to improve the spatial resolution in software method and superresolution is becoming a popular technique to solve this problem.

Image superresolution is a technique that reconstructs a high-resolution (HR) image from one or few given low-resolution (LR) images. These LR images are downsampled and blurred due to degradation of imagery system [1]. Huang and Tsai [2] firstly introduced the superresolution problem in the frequency domain but disregarded the blur in the image processing. Some researchers present maximum a posteriori (MAP) estimator-based algorithms to solve the interpolation and deblurring problems [3–5]. Projection onto convex sets (POCS) is also a very popular technique in image restoration issues and therefore is implemented in superresolution algorithms [6–8].

The superresolution methods have been developed for a long time for natural images. However, with the development of space technology, hyperspectral images become more popular and widely used. Many researchers start to research the superresolution method on hyperspectral images to improve the spatial resolution. For example, Akgun et al. [1] present a hyperspectral image acquisition model to simulate the degradation process and provide a POCS-based

superresolution method to improve the spatial resolution of hyperspectral images. Mianji et al. [9] reviewed some superresolution methods for hyperspectral imagery and provided the challenges in this area. Ma et al. [10] proposed an operational superresolution approach for multitemporal and multiangle remotely sensed imagery.

In this paper, we propose an advanced nonlocal means (ANLM) and iterative back projection- (IBP-) based superresolution method for hyperspectral imagery. IBP is a famous superresolution method proposed by Irani and Peleg [11]. The main concept of IBP is to minimize the error during the iterative process. It is reported that IBP can give good performance and achieve the superresolution and deblurring simultaneously. However, this iterative process often causes jaggy and ringing artifacts in the detail region. The proposed ANLM filter is an improved version of the traditional nonlocal means (NLM) filter established by Buades et al. in [12], which was inspired from Yaroslavsky neighborhood filter [13]. The NLM has been widely used in image processing such as denoising and deblurring [14, 15]. NLM estimates the missing pixel as the weighted average of the pixels whose neighborhoods seem like the neighborhood of the missing pixel. Known from Yaroslavsky neighborhood filter, the spatial patch distance also affects the performance of the superresolution method because long distance has smaller chance to have similar neighborhood than short distance. Moreover, NLM needs to compute all the neighborhood similarities in the whole image which is computationally expensive. Thus, we propose ANLM where the missing pixel is smoothed as the weighted average of all pixels that have similar Gaussian neighborhoods and close patch distance.

The rest of the paper is organized as follows. The proposed ANLM method is described in Section 2. Simulation results and their corresponding discussion are explained in Section 3. Finally, conclusions are drawn in Section 4.

2. Proposed Method

2.1. Advanced Nonlocal Means (ANLM) Filter. A basic digital signal superresolution system can be realized as weighted linear average. Given a discrete LR image \mathbf{f} , we want to estimate the HR image \mathbf{g} ; the conventional methods adopt the uniform weighted sum of average values and can be expressed as

$$g(\mathbf{x}) = \frac{\sum_{\mathbf{y} \in S_{\mathbf{x}}} w_{\mathbf{y}} f(\mathbf{y})}{\sum_{\mathbf{y} \in S_{\mathbf{x}}} w_{\mathbf{y}}}, \quad (1)$$

where $S_{\mathbf{x}}$ denotes the search region of a fixed size and is centered at the missing pixel located at \mathbf{x} , $w_{\mathbf{y}}$ is the weight, $f(\mathbf{y})$ is the pixel value of the given LR image, and $g(\mathbf{x})$ is the estimation of the missing pixel of the reconstructed HR image.

The NLM filter restores the missing pixel by taking an average value of the neighboring pixels with a similar neighborhood. Using a Gaussian kernel for evaluating the similarity, the formula can be expressed as

$$g_{\text{NLM}}(\mathbf{x}) = \frac{\sum_{\mathbf{y} \in S_{\mathbf{x}}} w_{\text{NLM}}(\mathbf{x}, \mathbf{y}) f(\mathbf{y})}{\sum_{\mathbf{y} \in S_{\mathbf{x}}} w_{\text{NLM}}(\mathbf{x}, \mathbf{y})}, \quad (2)$$

where $S_{\mathbf{x}}$ is the search region and w_{NLM} is the weight of NLM filter defined as

$$w_{\text{NLM}}(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{v}(N_{\mathbf{x}}) - \mathbf{v}(N_{\mathbf{y}})\|_2^2}{h^2}\right), \quad (3)$$

where $N_{\mathbf{x}}$ and $N_{\mathbf{y}}$ are square neighborhood centered at \mathbf{x} and \mathbf{y} , $\mathbf{v}(N_{\mathbf{x}}) = \{f(\mathbf{x}) \mid \mathbf{x} \in N_{\mathbf{x}}\}$ and $\mathbf{v}(N_{\mathbf{y}}) = \{g(\mathbf{y}) \mid \mathbf{y} \in N_{\mathbf{y}}\}$ are the intensity gray level vectors which are composed of pixels in the neighborhoods $N_{\mathbf{x}}$ and $N_{\mathbf{y}}$, and h is a filtering parameter. It is apparent that the similarity between two pixels at \mathbf{x} and \mathbf{y} depends on the similarity of the intensity gray level vectors $\mathbf{v}(N_{\mathbf{x}})$ and $\mathbf{v}(N_{\mathbf{y}})$.

Most NLM-based methods only consider neighborhood similarity and ignore the spatial locality distance which results in poor performance. Therefore, we propose an ANLM method in order to obtain an improved performance while considering the spatial distance [15]. In ANLM, the patch distance $d(N_{\mathbf{x}}, N_{\mathbf{y}}) = |\mathbf{x} - \mathbf{y}|$ is introduced as a new weight; then we redesign the weight function as

$$w_{\text{ANLM}}(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{d^2(N_{\mathbf{x}}, N_{\mathbf{y}})}{\rho^2} - \frac{\|\mathbf{v}(N_{\mathbf{x}}) - \mathbf{v}(N_{\mathbf{y}})\|_2^2}{h^2}\right), \quad (4)$$

Where ρ and h are filtering parameters. Then, we apply $w_{\text{ANLM}}(\mathbf{x}, \mathbf{y})$ to the ANLM filter as follows:

$$g_{\text{ANLM}}(\mathbf{x}) = \frac{\sum_{\mathbf{y} \in S_{\mathbf{x}}} w_{\text{ANLM}}(\mathbf{x}, \mathbf{y}) f(\mathbf{y})}{\sum_{\mathbf{y} \in S_{\mathbf{x}}} w_{\text{ANLM}}(\mathbf{x}, \mathbf{y})}. \quad (5)$$

The search region of conventional NLM was the entire image. In the proposed ANLM, the introduced patch distance $d(N_{\mathbf{x}}, N_{\mathbf{y}})$ is increasing when the given pixel \mathbf{y} is farther from the missing pixel \mathbf{x} . When $d(N_{\mathbf{x}}, N_{\mathbf{y}})$ is larger than some constant value, $\exp(-d^2(N_{\mathbf{x}}, N_{\mathbf{y}})/\rho^2)$ is close to zero. Particularly, the weight $w_{\text{ANLM}}(\mathbf{x}, \mathbf{y})$ is trivial for the larger patch distance. Thus, we can neglect these pixels that are located far from the missing pixel \mathbf{x} . We search the similar neighborhood in a limited search region instead of the whole image as in NLM, which reduce the computational complexity considerably.

2.2. Iterative Back Projection (IBP). The goal of superresolution algorithms is to restore the desired HR image from downsampled and blurred LR image. Irani and Peleg proposed an IBP superresolution reconstruction method [11]. In IBP, the difference between the simulated and given LR images is repeatedly iteratively back projected to obtain the desired HR image. The process continues until some conditions are satisfied or the maximum iteration number is reached. IBP can be formulated as

$$\mathbf{g}_{n+1} = \mathbf{g}_n + \mathbf{g}_e, \quad (6)$$

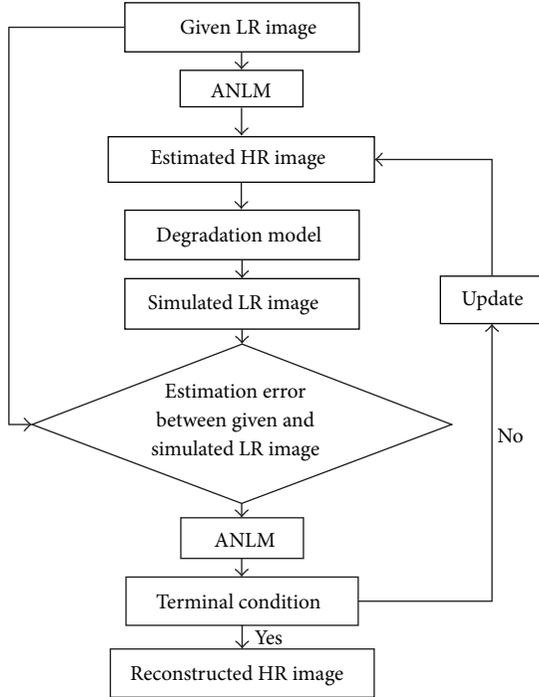


FIGURE 1: Flowchart of the proposed ANLM-IBP method.

where \mathbf{g}_{n+1} is estimated HR image in the $(n + 1)$ th iteration, \mathbf{g}_n is estimated HR image in the n th iteration, and \mathbf{g}_e is the estimation error which is computed by

$$\mathbf{g}_e = (\mathbf{f}_n - \mathbf{f}) \uparrow s * p, \quad (7)$$

where \mathbf{f}_n is the simulated LR image which is the downsampling and blurring of \mathbf{g}_n , $\uparrow s$ is upsampling, and p is back projection kernel.

2.3. The Proposed ANLM-IBP Method. To improve the performance of IBP, we propose an ANLM-IBP-based superresolution algorithm for hyperspectral imagery. First, we use ANLM to reconstruct the initial HR image. Then, during the iterative process of IBP, ANLM guides the error propagation of IBP. By combining ANLM and IBP, one can achieve better objective and subjective performance. The flowchart of the proposed ANLM-IBP is drawn in Figure 1.

In summary, the proposed ANLM-IBP method is described in the following steps:

- (1) apply ANLM to obtain initially estimated HR image;
- (2) use degradation model for downsampling and blurring the estimated HR image to obtain the simulated LR image;
- (3) calculate estimation error between the given original LR image and the simulated LR image;
- (4) utilize ANLM to interpolate the estimation error;
- (5) judge if the terminal condition is satisfied. If it is Yes, computed image is the reconstructed HR image. If it is No, we add the error to the estimated HR image and reobtain an updated HR image for the next iteration.



FIGURE 2: The tested image 1 with 191-band.



FIGURE 3: The tested image 2 with 220-band.

3. Experimental Results

In the experimental results, we adopt a well-known dissimilarity criterion called peak signal-to-noise ratio (PSNR) in decibels (dB), which is calculated as

$$\text{MSE}(\text{org}, \text{rec}) = \sum_{i=1}^{\text{width}} \sum_{j=1}^{\text{height}} \frac{(\text{org}(i, j) - \text{rec}(i, j))^2}{\text{width} \times \text{height}}, \quad (8)$$

$$\text{PSNR}(\text{org}, \text{rec}) = 10 \log_{10} \frac{255^2}{\text{MSE}(\text{org}, \text{rec})},$$

where org and rec are the original and reconstructed images, respectively. We measured the objective performance in terms of PSNR, which is a widely adopted criterion in the literature.

The proposed method was tested with 191-band airborne multispectral scanner data set [16] and 220-band spectral image acquired with the AVIRIS data set (shown in Figures 2 and 3). For the subjective performance evaluation in terms of visual effect, we show part of perceived image quality in Figures 4 and 5. As can be seen in Figure 3, result images show that the proposed method yields a better visual quality with details which has sharper edge and better object boundary.

Table 1 shows objective performance comparison in PSNR metric. Our proposed method yields better performance than the other existing methods.

4. Conclusion

We propose an advanced nonlocal means and iterative back projection-based superresolution algorithm for hyperspectral imagery. We introduce an improved version of nonlocal

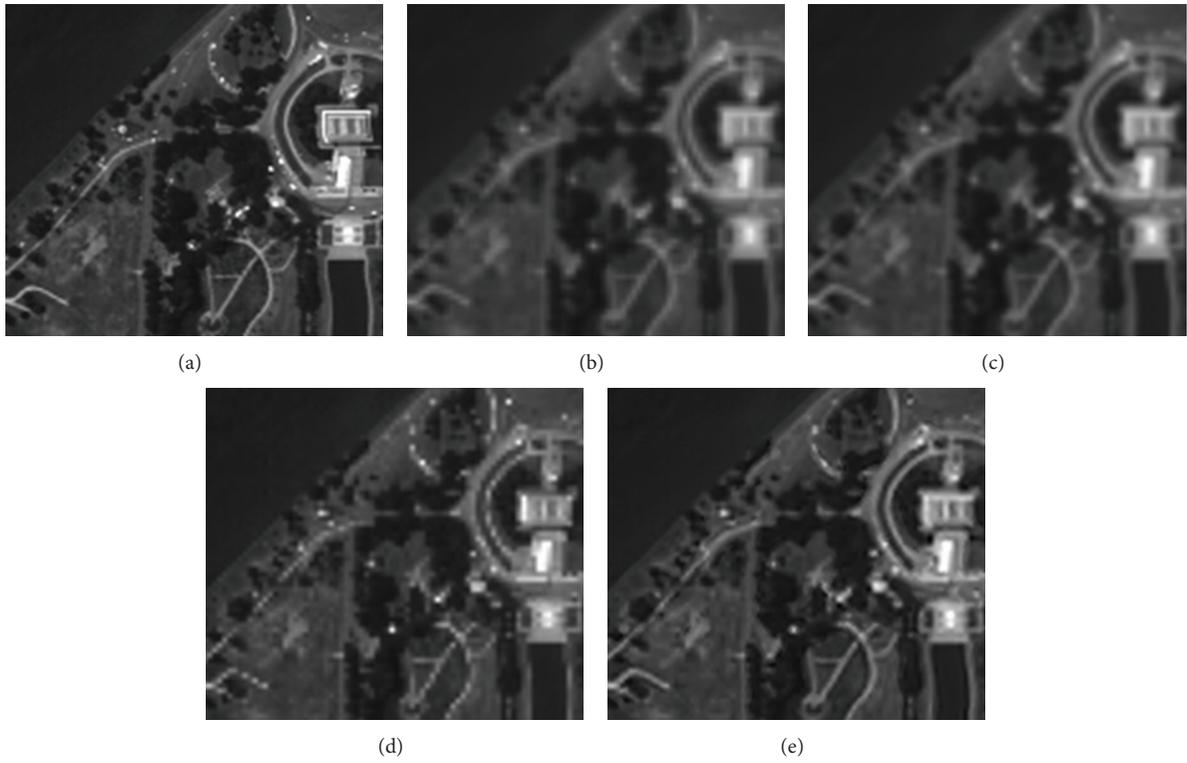


FIGURE 4: (a) Original part image. The perceived image quality comparison using various deinterlacing methods: (b) bilinear, (c) bicubic, (d) POCS, and (e) proposed method.

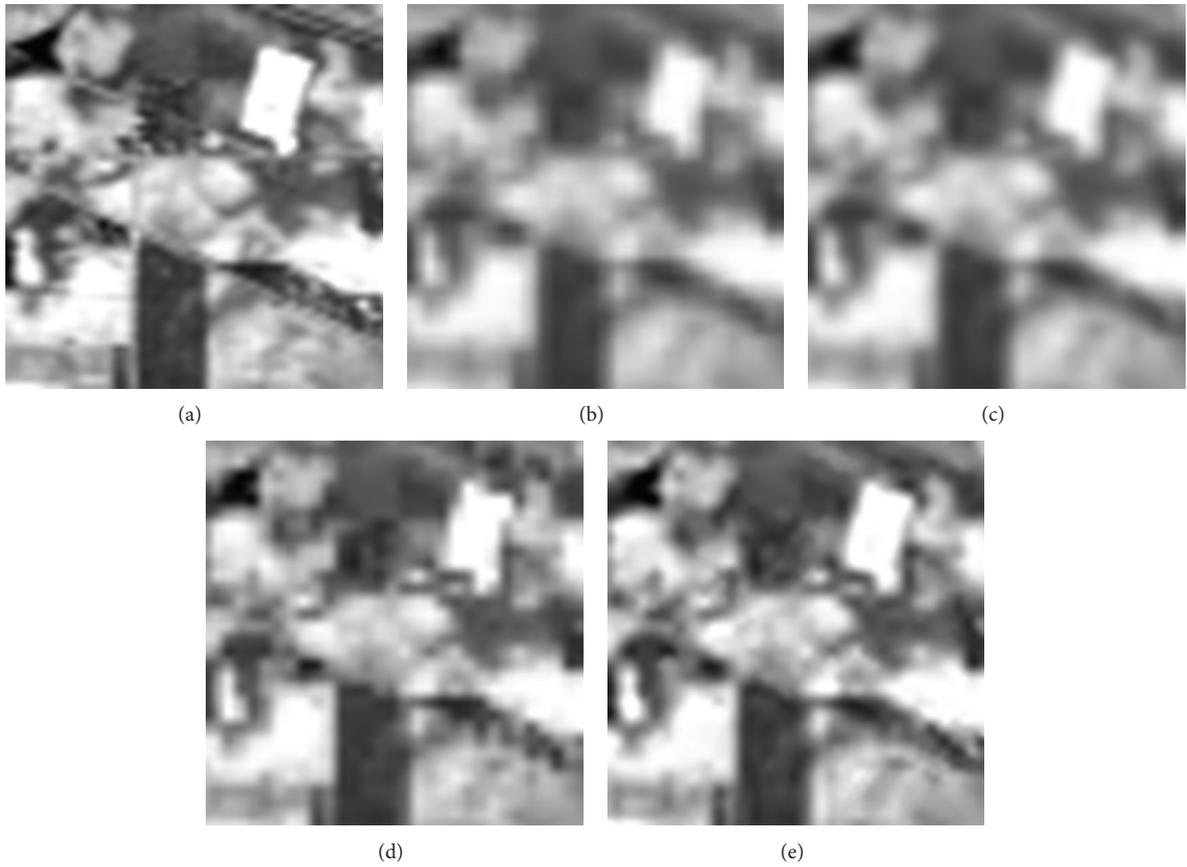


FIGURE 5: (a) Original part image. The perceived image quality comparison using various deinterlacing methods: (b) bilinear, (c) bicubic, (d) POCS, and (e) proposed method.

TABLE 1: Objective performance comparison with metrics of PSNR for different methods.

Method	Bilinear	Bicubic	POCS	Proposed
Image 1	22.67	23.13	23.65	25.23
Image 2	23.65	24.24	24.66	26.58

means filter that automatically selects the edge orientation using the weighted average of a similar neighborhood and reduces the complexity of nonlocal means filter by patch distance. In addition, we combine the advanced nonlocal means with iterative back projection to reduce the artifacts caused by the iterative back projection. The experimental results show that we could obtain better objective and subjective performance compared to conventional methods.

Conflict of Interests

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

Acknowledgment

This work was supported by the Incheon National University Research Grant in 2012.

References

- [1] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1860–1875, 2005.
- [2] T. Huang and R. Tsai, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, T. S. Huang, Ed., vol. 1, JAI, Greenwich, Conn, USA, 1984.
- [3] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, no. 3, pp. 455–469, 1994.
- [4] H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach to joint motion estimation, segmentation, and super resolution," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 479–490, 2007.
- [5] G. Liu and Z. Hu, "MAP based blind super-resolution," in *Proceedings of the International Conference on Industrial Control and Electronics Engineering (ICICEE '12)*, pp. 26–28, Xi'an, China, August 2012.
- [6] Y. Altunbasak, A. J. Patti, and R. M. Mersereau, "Super-resolution still and video reconstruction from MPEG-coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 4, pp. 217–226, 2002.
- [7] J. Ma and J. C.-W. Chan, "Superresolution reconstruction of hyperspectral remote sensing imagery using constrained optimization of POCS," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pp. 7271–7274, July 2012.
- [8] Y. Altunbasak, A. Patti, and R. Mersereau, "POCS based super-resolution image reconstruction using an adaptive regularization parameter," *IJCSI International Journal of Computer Science Issues*, vol. 8, pp. 155–158, 2011.
- [9] F. A. Mianji, Y. Zhang, H. K. Sulehria, A. Babakhani, and M. R. Kardan, "Super-resolution challenges in hyperspectral imagery," *Information Technology Journal*, vol. 7, no. 7, pp. 1030–1036, 2008.
- [10] J. Ma, J. Cheung-Wai Chan, and F. Canters, "An operational superresolution approach for multi-temporal and multi-angle remotely sensed imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 110–124, 2012.
- [11] M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion and transparency," *Journal of Visual Communication and Image Representation*, vol. 4, no. 4, pp. 324–335, 1993.
- [12] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.
- [13] L. P. Yaroslavsky, *Digital Picture Processing, An Introduction*, Springer, Berlin, Germany, 1985.
- [14] C. Sutour, C.-A. Deledalle, and J.-F. Aujol, "Adaptive regularization of the NL-means: application to image and video denoising," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3506–3521, 2014.
- [15] X. Li and Y. Zheng, "Patch-based video processing: a variational Bayesian approach," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 1, pp. 27–40, 2009.
- [16] <https://engineering.purdue.edu/~biehl>.