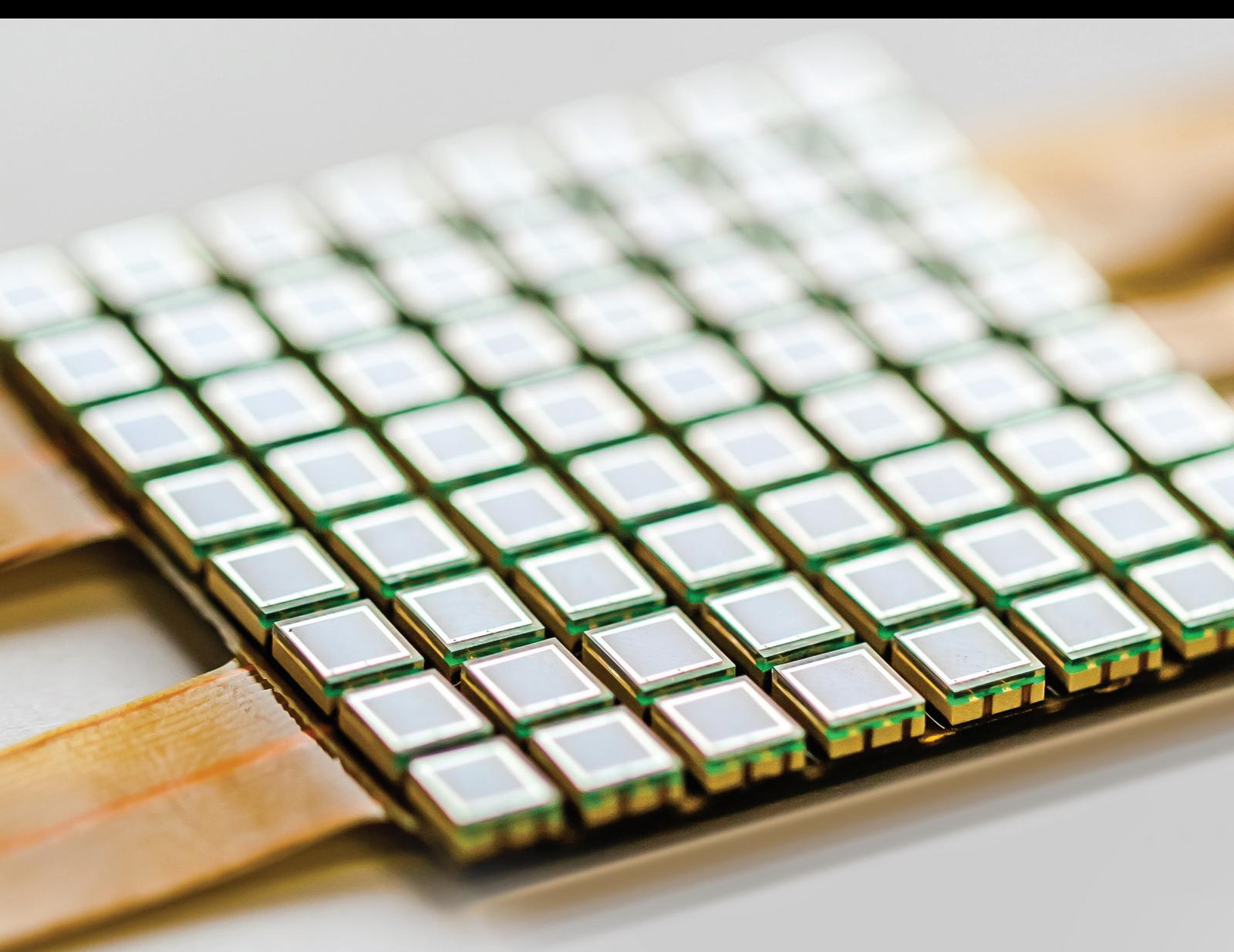


Multispectral, Hyperspectral, and Polarimetric Imaging Technology

Guest Editors: Yongqiang Zhao, Jonathan Cheung-Wai Chan, Seong G. Kong,
Yanli Qiao, Chao-Cheng Wu, and Chein-I. Chang



Multispectral, Hyperspectral, and Polarimetric Imaging Technology

Multispectral, Hyperspectral, and Polarimetric Imaging Technology

Guest Editors: Yongqiang Zhao, Jonathan Cheung-Wai Chan,
Seong G. Kong, Yanli Qiao, Chao-Cheng Wu,
and Chein-I. Chang



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	Wei Gao, Japan	Alain Pauly, France
Francisco J. Arregui, Spain	Michele Giordano, Italy	Giorgio Pennazza, Italy
Francesco Baldini, Italy	Banshi D. Gupta, India	Michele Penza, Italy
Fernando Benito-Lopez, Ireland	María del Carmen Horrillo, Spain	Andrea Ponzoni, Italy
Romeo Bernini, Italy	Wieslaw Jakubik, Poland	Biswajeet Pradhan, Malaysia
Shekhar Bhansali, USA	Hai-Feng Ji, USA	Ioannis Raptis, Greece
Wojtek J. Bock, Canada	Kourosh Kalantar-Zadeh, Australia	Christos Riziotis, Greece
Hubert Brändle, Switzerland	Sher Bahadar Khan, KSA	Maria Luz Rodríguez-Méndez, Spain
Davide Brunelli, Italy	Sang Sub Kim, Republic of Korea	Albert Romano-Rodriguez, Spain
Belen Calvo, Spain	Challa Kumar, USA	Josep Samitier, Spain
Stefania Campopiano, Italy	Hiroki Kuwano, Japan	Giorgio Sberveglieri, Italy
Jian-Nong Cao, Hong Kong	Laura M. Lechuga, Spain	Luca Schenato, Italy
Sara Casciati, Italy	Chengkuo Lee, Singapore	Michael J. Schöning, Germany
Chi Chiu Chan, Singapore	Chenzhong Li, USA	Andreas Schütze, Germany
Nick Chaniotakis, Greece	Eduard Llobet, Spain	Woosuck Shin, Japan
Nicola Cioffi, Italy	Yu-Lung Lo, Taiwan	Pietro Siciliano, Italy
Elisabetta Comini, Italy	Oleg Lupan, Moldova	Vincenzo Spagnolo, Italy
Marco Consales, Italy	Eugenio Martinelli, Italy	Weilian Su, USA
Jesus Corres, Spain	Jose R. Martinez-de Dios, Spain	Tong Sun, United Kingdom
Andrea Cusano, Italy	Yasuko Y. Maruo, Japan	Hidekuni Takao, Japan
Dzung V. Dao, Japan	Ignacio R. Matias, Spain	Isao Takayanagi, Japan
Catherine Dehollain, Switzerland	Mike McShane, USA	Pierre Temple-Boyer, France
Manel del Valle, Spain	Igor L. Medintz, USA	Guiyun Tian, UK
Ignacio Del Villar, Spain	Fanli Meng, China	Suna Timur, Turkey
Utkan Demirci, USA	Aldo Minardo, Italy	Hana Vaisocherova, Czech Republic
Junhang Dong, USA	Joan Ramon Morante, Spain	Qihao Weng, USA
Abdelhamid Errachid, France	Lucia Mosiello, Italy	Matthew J. Whelan, USA
Stephane Evoy, Canada	Masayuki Nakamura, Japan	Stanley E. Woodard, USA
Vittorio Ferrari, Italy	Vincenzo Paciello, Italy	Hai Xiao, USA
Luca Francioso, Italy	Marimuthu Palaniswami, Australia	Hyeonseok Yoon, Republic of Korea
Laurent Francis, Belgium	Alberto J. Palma, Spain	
Mohammad Reza Ganjali, Iran	Gyuhae Park, Republic of Korea	

Contents

Multispectral, Hyperspectral, and Polarimetric Imaging Technology

Yongqiang Zhao, Jonathan Cheung-Wai Chan, Seong G. Kong, Yanli Qiao, Chao-Cheng Wu, and Chein-I. Chang

Volume 2016, Article ID 5634542, 2 pages

Stacked Denoise Autoencoder Based Feature Extraction and Classification for Hyperspectral Images

Chen Xing, Li Ma, and Xiaoquan Yang

Volume 2016, Article ID 3632943, 10 pages

Design of an Active Multispectral SWIR Camera System for Skin Detection and Face Verification

Holger Steiner, Sebastian Sporrer, Andreas Kolb, and Norbert Jung

Volume 2016, Article ID 9682453, 16 pages

Multiband Polarization Imaging

Yongqiang Zhao, Qunnie Peng, Chen Yi, and Seong G. Kong

Volume 2016, Article ID 5985673, 10 pages

Research on Multifeature Segmentation Method of Remote Sensing Images Based on Graph Theory

Wenxing Bao and Xiuhong Yao

Volume 2016, Article ID 8750927, 8 pages

Development of a Gastric Cancer Diagnostic Support System with a Pattern Recognition Method Using a Hyperspectral Camera

Hiroyuki Ogihara, Yoshihiko Hamamoto, Yusuke Fujita, Atsushi Goto, Jun Nishikawa, and Isao Sakaida

Volume 2016, Article ID 1803501, 6 pages

A New Methodology for Spectral-Spatial Classification of Hyperspectral Images

Zelang Miao and Wenzhong Shi

Volume 2016, Article ID 1538973, 12 pages

Airborne Polarimetric Remote Sensing for Atmospheric Correction

Tianquan Liang, Xiaobing Sun, Han Wang, Rufang Ti, and Cunming Shu

Volume 2016, Article ID 3569272, 7 pages

Land Use Transformation Rule Analysis in Beijing-Tianjin-Tangshan Region Using Remote Sensing and GIS Technology

Shang-min Zhao, Wei-ming Cheng, Hai-jiang Liu, Yao Xia, Hui-xia Chai, Yang Song, Wen-jie Zhang, and Tian You

Volume 2016, Article ID 6756295, 10 pages

Undersampled Hyperspectral Image Reconstruction Based on Surfacelet Transform

Lei Liu, Jingwen Yan, Di Guo, Yunsong Liu, and Xiaobo Qu

Volume 2015, Article ID 256391, 11 pages

Use of GMM and SCMS for Accurate Road Centerline Extraction from the Classified Image

Zelang Miao, Bin Wang, Wenzhong Shi, Hao Wu, and Yiliang Wan

Volume 2015, Article ID 784504, 13 pages

Editorial

Multispectral, Hyperspectral, and Polarimetric Imaging Technology

Yongqiang Zhao,¹ Jonathan Cheung-Wai Chan,² Seong G. Kong,³ Yanli Qiao,⁴ Chao-Cheng Wu,⁵ and Chein-I. Chang⁶

¹School of Automation, Northwestern Polytechnical University, Xian, Shanxi 710072, China

²Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium

³Department of Computer Engineering, Sejong University, Seoul 143-747, Republic of Korea

⁴Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, Anhui 230031, China

⁵Department of Electrical Engineering, National Taipei University of Technology, Taipei 10608, Taiwan

⁶Remote Sensing Signal and Image Processing Laboratory (RSSIPL), University of Maryland Baltimore County (UMBC), Baltimore, MD 21201, USA

Correspondence should be addressed to Yongqiang Zhao; zhaoyq@nwpu.edu.cn

Received 2 July 2015; Accepted 7 July 2015

Copyright © 2016 Yongqiang Zhao 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 special issue presents papers that describe research in multispectral, hyperspectral, and polarimetric image processing and applications. The subject matter covered ranges from a review of multiband polarization imaging to a presentation of hyperspectral image reconstruction.

Additionally to intensity and spectrum, polarization is a powerful light parameter that contains important information about the diffusion and reflection of light. Spectral imaging sensor is the one that captures image data at specific frequencies across the electromagnetic spectrum, which includes multispectral imager and hyperspectral imager. Polarimetric imaging is the technology on quantitative measurements of this parameter by imaging sensors. Multispectral, hyperspectral, and polarimetric imaging has demonstrated strong potential to retrieve critical information from the scene of interest in challenging imaging conditions. With the advance of imaging sensor's resolution and sensitivity, imaging spectrometry and polarimetry have emerged over the past several decades as a powerful tool to enhance the information availability in computer vision, remote sensing, and biomedicine. Applications of multispectral, hyperspectral, and polarimetric imaging include specular highlight removal, information enhancement in the shadow, penetration through scattering media, investigation of the

materials as well as biological tissues, and characterization of atmospheric aerosols and cloud particles.

This special issue contains 10 papers describing research in multispectral, hyperspectral, and polarimetric imaging and applications. The paper by Y. Zhao et al. is a review of multiband polarization imaging. This paper reviews earlier attempts and recent advances in multiband polarization imaging techniques and their applications in various fields such as atmospheric observation, object detection and classification, medical diagnostics, and surveillance. It also introduces the future development trend of multiband polarization imaging techniques. In another paper, T. Liang et al. introduce the airborne polarimetric remote sensing system designed by Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Science. By using the airborne remote sensing images which were acquired on the north coast areas of China during the haze weather, they proved the potential of this system for atmospheric correction. C. Xing et al. describe a hyperspectral image classification method based on deep learning; they utilize a stacked denoise autoencoder (SDAE) method to pretrain the network. Hyperion, AVIRIS, and ROSIS hyperspectral data is used to test the classification performance. H. Ogihara et al. introduce a gastric cancer diagnostic support system using hyperspectral camera. In

this system, the problem of selecting the optimum wavelength and optimizing the cutoff value is very important, and authors solved it by pattern recognition. H. Steiner et al. design an active multispectral SWIR camera system that acquires four-band multispectral image stacks in real-time; they also set up a database containing RGB and multispectral SWIR face images for open access. W. Bao and X. Yao represent a remote sensing images segmentation method by combining the spectrum, shape, and texture features based on graph theory. Z. Miao and W. Shi introduce a new hyperspectral image classification method based on spectral-spatial features. Z. Miao et al. propose the method of accurate road centerline extraction from the classified image. S. Zhao et al. analyze the land use transformation rule in Beijing-Tianjin-Tangshan Region, an important industrial base in China, by using remote sensing and GIS technology. They acquire the land use distribution status at 2000, 2005, and 2010 in Beijing-Tianjin-Tangshan Region using remote sensing images, field survey data, images in Google Earth, and visual interpretation methods. Then, the land use transformation rules from 2000 to 2010 are achieved using GIS (geographic information system) technology and land use distribution status. L. Liu et al. propose an undersampled hyperspectral image reconstruction based on surfacelet transform. They use surfacelet to sparsify the hyperspectral images. Besides, a Gram-Schmidt orthogonalization is used in CS random encoding matrix, two-dimensional and three-dimensional orthogonal compressed sensing random encoding, and a patch-based compressed sensing encoding scheme is designed.

Whether your interest is in the multispectral, hyperspectral, and polarimetric imaging systems, the image processing methods, or the applications, we hope you enjoy this special issue.

Acknowledgment

We extend our appreciation to all the authors and reviewers who have worked hard to produce a high-quality issue covering a wide range of topics that represent some of the interesting work being conducted currently in the field of multispectral, hyperspectral, and polarimetric imaging and applications.

*Yongqiang Zhao
Jonathan Cheung-Wai Chan
Seong G. Kong
Yanli Qiao
Chao-Cheng Wu
Chein-I. Chang*

Research Article

Stacked Denoise Autoencoder Based Feature Extraction and Classification for Hyperspectral Images

Chen Xing,¹ Li Ma,¹ and Xiaoquan Yang²

¹*Faculty of Mechanical and Electronic Information, China University of Geosciences, Wuhan, Hubei 430074, China*

²*Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*

Correspondence should be addressed to Li Ma; maryparisster@gmail.com

Received 25 December 2014; Revised 11 May 2015; Accepted 21 June 2015

Academic Editor: Jonathan C.-W. Chan

Copyright © 2016 Chen Xing 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.

Deep learning methods have been successfully applied to learn feature representations for high-dimensional data, where the learned features are able to reveal the nonlinear properties exhibited in the data. In this paper, deep learning method is exploited for feature extraction of hyperspectral data, and the extracted features can provide good discriminability for classification task. Training a deep network for feature extraction and classification includes unsupervised pretraining and supervised fine-tuning. We utilized stacked denoise autoencoder (SDAE) method to pretrain the network, which is robust to noise. In the top layer of the network, logistic regression (LR) approach is utilized to perform supervised fine-tuning and classification. Since sparsity of features might improve the separation capability, we utilized rectified linear unit (ReLU) as activation function in SDAE to extract high level and sparse features. Experimental results using Hyperion, AVIRIS, and ROSIS hyperspectral data demonstrated that the SDAE pretraining in conjunction with the LR fine-tuning and classification (SDAE_LR) can achieve higher accuracies than the popular support vector machine (SVM) classifier.

1. Introduction

Hyperspectral remote sensing images are becoming increasingly available and potentially provide greatly improved discriminant capability for land cover classification. Popular classification methods like k -nearest-neighbor [1], support vector machine [2], and semisupervised classifiers [3] have been successfully applied to hyperspectral images. Besides, some feature matching methods in the computer vision area can also be generalized for spectral classification [4, 5].

Feature extraction is very important for classification of hyperspectral data, and the learned features may increase the separation between spectrally similar classes, resulting in improved classification performance. Commonly used linear feature extraction methods such as principal component analysis (PCA) and linear discriminant analysis (LDA) are simple and easily implemented. However, these methods fail to model the nonlinear structures of data. Manifold learning methods, which are proposed for nonlinear feature extraction, are able to characterize the nonlinear relationships

between data points [1, 6, 7]. However, they can only process a limited number of data points due to their high computational complexity. Deep learning methods, which can also learn the nonlinear features, are capable of processing large scale data set. Therefore, we utilized deep learning for feature extraction of hyperspectral data in this paper.

Deep learning is proposed to train a deep neural network for feature extraction and classification. The training process includes two steps: unsupervised layer-wise pretraining and supervised fine-tuning. The layer-wise pretraining [8] can alleviate the difficulty of training a deep network, since the learned network weights which encode the data structure are used as the initial weights of the whole deep network. The supervised fine-tuning that is performed by logistic regression (LR) approach aims to further adjust the network weights by minimizing the classification errors of the labeled data points. Training the network can achieve both high level features and classification simultaneously. Popular deep learning methods include autoencoders (AE) [9], denoised autoencoders (DAE) [10], convolutional neural networks

(CNN) [11], deep belief networks (DBN) [12], and convolutional restricted Boltzmann machines (CRBM) [13]. In the field of hyperspectral data analysis, Chen utilized AE for data classification [14], and Zhang utilized CNN for feature extraction [15].

In this paper, we focus on the stacked DAE (SDAE) method [16], since DAE is very robust to noise, and SDAE can obtain higher level features. Moreover, since sparsity of features might improve the separation capability, we utilized rectified linear unit (ReLU) as activation function in SDAE to extract high level and sparse features. After the layer-wise pretraining by SDAE, LR layer is used for fine-tuning the network and performing classification. The features of the deep network that are obtained by SDAE pretraining and LR fine-tuning are called tuned-SDAE features, and the classification approach that utilizes LR classifier on the tuned-SDAE features is hereafter denoted as SDAE_LR in this paper.

The organization of the paper is as follows. Section 2 describes the DAE, SDAE, and SDAE_LR approaches. Section 3 discussed the experimental results. Conclusions are summarized in Section 4.

2. Methodology

Given a neural network, AE [14] trains the network by constraining the output values to be equal to the input values, which also indicates that the output layer has equally many nodes as the input layer. The reconstruction error between the input and the output of network is used to adjust the weights of each layer. Therefore, the features learned by AE can well represent the input data. Moreover, the training of AE is unsupervised, since it does not require label information. DAE is developed from AE but is more robust, since DAE assumes that the input data contain noise and is suitable to learn features from noisy data. As a result, the generalization ability of DAE is better than that of AE. Moreover, DAE can be stacked to obtain high level features, resulting in SDAE approach. The training of SDAE network is layer-wise, since each DAE with one hidden layer is trained independently. After training the SDAE network, the decoding layers are removed and the encoding layers that produce features are retained. For classification task, a logistic regression (LR) layer is added as output layer. Moreover, LR is also used to fine-tune the network. Therefore, the features are learned by SDAE pretraining in conjunction with LR fine-tuning.

2.1. Denoise Autoencoder (DAE). DAE contains three layers: input layer, hidden layer, and output layer, where the hidden

layer and output layer are also called encoding layer and decoding layer, respectively. Suppose the original data is $\mathbf{x} \in R^d$, where d is the dimension of data. DAE firstly produces a vector $\tilde{\mathbf{x}}$ by setting some of the elements to zero or adding the Gaussian noise to \mathbf{x} . DAE uses $\tilde{\mathbf{x}}$ as input data. The number of units in the input layer is d , which is equal to the dimension of the input data $\tilde{\mathbf{x}}$. The encoding of DAE is obtained by a nonlinear transformation function:

$$\mathbf{y} = f_e(\mathbf{W}\tilde{\mathbf{x}} + \mathbf{b}), \quad (1)$$

where $\mathbf{y} \in R^h$ denotes the output of the hidden layer and can also be called feature representation or code, h is the number of units in the hidden layer, $\mathbf{W} \in R^{h \times d}$ is the input-to-hidden weights, \mathbf{b} denotes the bias, $\mathbf{W}\tilde{\mathbf{x}} + \mathbf{b}$ stands for the input of the hidden layer, and $f_e()$ is called activation function of the hidden layer. We chose ReLU function [17] as the activation function in this study, which is formulated as

$$f_e(\mathbf{W}\tilde{\mathbf{x}} + \mathbf{b}) = \max(0, \mathbf{W}\tilde{\mathbf{x}} + \mathbf{b}). \quad (2)$$

If the value of $\mathbf{W}\tilde{\mathbf{x}} + \mathbf{b}$ is smaller than zero, the output of the hidden layer will be zero. Therefore, ReLU activation function is able to produce a sparse feature representation, which may have better separation capability. Moreover, ReLU can train the neural network for large scale data faster and more effectively than the other activation functions.

The decoding or reconstruction of DAE is obtained by using a mapping function $f_d()$:

$$\mathbf{z} = f_d(\mathbf{W}'\mathbf{y} + \mathbf{b}'), \quad (3)$$

where $\mathbf{z} \in R^d$ is the output of DAE, which is also the reconstruction of original data \mathbf{x} . The output layer has the same number of nodes as the input layer. $\mathbf{W}' = \mathbf{W}^T$ is referred to as tied weights. If \mathbf{x} is ranged from 0 to 1, we choose softplus function as the decoding function $f_d()$; otherwise we preprocess \mathbf{x} by zero-phase component analysis (ZCA) whitening and use a linear function as the decoding function:

$$f_d(\mathbf{a}) = \begin{cases} \log(1 + e^{\mathbf{a}}), & \mathbf{x} \in [0, 1] \\ \mathbf{a}, & \text{otherwise,} \end{cases} \quad (4)$$

where $\mathbf{a} = \mathbf{W}'\mathbf{y} + \mathbf{b}'$. DAE aims to train the network by requiring the output data \mathbf{z} to reconstruct the input data \mathbf{x} , which is also called reconstruction-oriented training. Therefore, the reconstruction error should be used as the objective function or cost function, which is defined as follows:

$$\text{Cost} = \begin{cases} -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^d [\mathbf{x}_j^{(i)} \log(\mathbf{z}_j^{(i)}) + (1 - \mathbf{x}_j^{(i)}) \log(1 - \mathbf{z}_j^{(i)})] + \frac{\lambda}{2} \|\mathbf{W}\|^2, & \mathbf{x} \in [0, 1], \\ \frac{1}{m} \sum_{i=1}^m \|\mathbf{x}^{(i)} - \mathbf{z}^{(i)}\|^2 + \frac{\lambda}{2} \|\mathbf{W}\|^2, & \text{otherwise,} \end{cases} \quad (5)$$

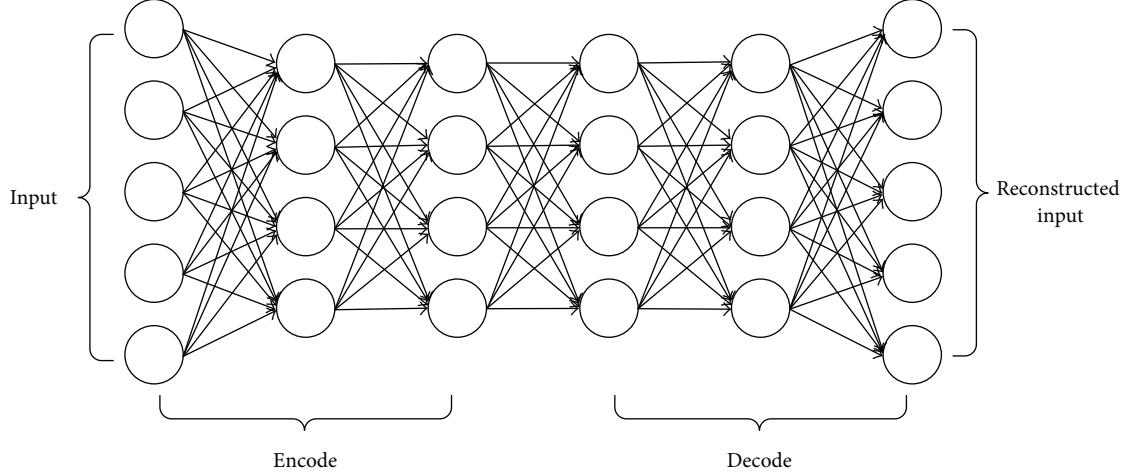


FIGURE 1: The SDAE network is stacked by two DAE structures.

where cross-entropy function is used when the value of input \mathbf{x} is ranged from 0 to 1; the square error function is used otherwise. $\mathbf{x}_j^{(i)}$ denotes j th element of the i th sample and $\|\mathbf{W}\|^2$ is L2-regularization term, which is also called weight decay term. Parameter λ controls the importance of the regularization term. This optimization problem is solved by using minibatch stochastic gradient descent (MSGD) algorithm [18], and m in (5) denotes the size of the minibatch.

2.2. Stacked Denoise Autoencoder (SDAE). DAE can be stacked to build deep network which has more than one hidden layer [16]. Figure 1 shows a typical instance of SDAE structure, which includes two encoding layers and two decoding layers. In the encoding part, the output of the first encoding layer acted as the input data of the second encoding layer. Supposing there are L hidden layers in the encoding part, we have the activation function of the k th encoding layer:

$$\mathbf{y}^{(k+1)} = f_e(\mathbf{W}^{(k+1)}\mathbf{y}^{(k)} + \mathbf{b}^{(k+1)}), \quad k = 0, \dots, L-1, \quad (6)$$

where the input $\mathbf{y}^{(0)}$ is the original data \mathbf{x} . The output $\mathbf{y}^{(L)}$ of the last encoding layer is the high level features extracted by the SDAE network. In the decoding part, the output of the first decoding layer is regarded as the input of the second decoding layer. The decoding function of the k th decode layer is

$$\mathbf{z}^{(k+1)} = f_d(\mathbf{W}^{(L-k)T}\mathbf{z}^{(k)} + \mathbf{b}'^{(k+1)}), \quad k = 0, \dots, L-1, \quad (7)$$

where the input $\mathbf{z}^{(0)}$ of the first decoding layer is the output $\mathbf{y}^{(L)}$ of the last encoding layer. The output $\mathbf{z}^{(L)}$ of the last decoding layer is the reconstruction of the original data \mathbf{x} .

The training process of SDAE is provided as follows.

Step 1. Choose input data, which can be randomly selected from the hyperspectral images.

Step 2. Train the first DAE, which includes the first encoding layer and the last decoding layer. Obtain the network weights $\mathbf{W}^{(1)}$ and $\mathbf{b}^{(1)}$ and the features $\mathbf{y}^{(1)}$ which are the output of the first encoding layer.

Step 3. Use $\mathbf{y}^{(k)}$ as the input data of the $(k+1)$ th encoding layer. Train the $(k+1)$ th DAE and obtain $\mathbf{W}^{(k+1)}$ and $\mathbf{b}^{(k+1)}$ and the features $\mathbf{y}^{(k+1)}$, where $k = 1, \dots, L-1$ and L is the number of hidden layers in the network.

It can be seen that each DAE is trained independently, and therefore the training of SDAE is called layer-wise training. Moreover, the trained network weights by SDAE acted as the initial weights in the following LR fine-tuning phase. Therefore, SDAE pretrains the network.

2.3. SDAE_LR. SDAE_LR includes SDAE pretraining and LR fine-tuning. SDAE trains the network weights and obtains features by the reconstruction-oriented learning, and the learned weights acted as the initial weights of the network. Further, LR is used to fine-tune the network weights and obtain the fine-tuned features. It is worth noting that SDAE is unsupervised, while LR is supervised and only the data with labeled information can be used in LR stage. The SDAE_LR network is illustrated in Figure 2, which shows a two-category classification problem (there are two output values). We can see that the decoding part of SDAE is removed and the encoding part of SDAE is retained to produce the initial features. In addition, the output layer of the whole network, which is also called LR layer, is added. The following sigmoid function is used as activation function of LR layer:

$$h(\mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{W}\mathbf{x} - \mathbf{b})}, \quad (8)$$

where \mathbf{x} is the output $\mathbf{y}^{(L)}$ of the last encoding layer. It is also the deep features that are pretrained by SDAE method. The

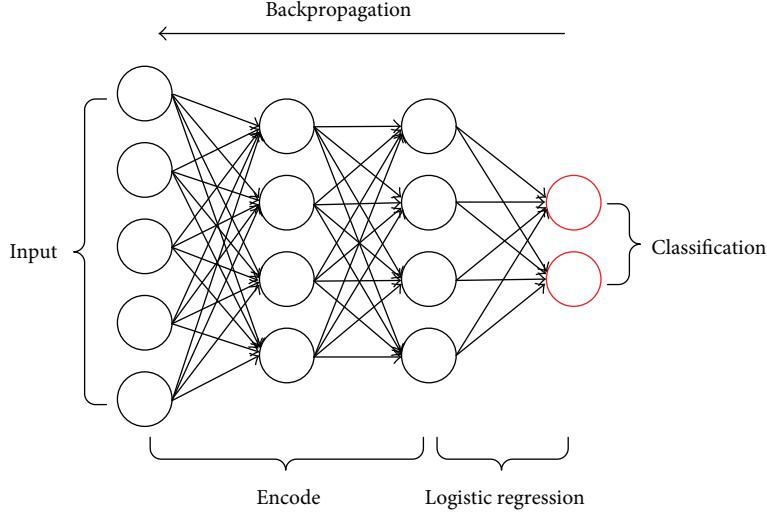


FIGURE 2: SDAE_LR structure includes the encoding part of SDAE for feature extraction and LR for fine-tuning and classification.

output of sigmoid function is between 0 and 1, which denotes the classification results.

Labels are associated with the training data points, and therefore we can use the errors between the predicted classification results and the true labels to fine-tune the whole network weights. The cost function is defined as the following cross-entropy function:

$$\text{Cost} = -\frac{1}{m} \left[\sum_{i=1}^m l^{(i)} \log(h(\mathbf{x}^{(i)})) + (1-l^{(i)}) \log(1-h(\mathbf{x}^{(i)})) \right], \quad (9)$$

where $l^{(i)}$ denotes the label of the sample $\mathbf{x}^{(i)}$. Minimizing the cost function, we can update the network weights. This optimization problem is also solved by MSGD method.

The steps of SDAE_LR network training are as follows.

Step 1. SDAE is utilized to train the initial network weights, described in Section 2.2.

Step 2. Initial weights of the LR layer are randomly set.

Step 3. Training data are used as input data, and their predicted classification results are produced with the initial weights of the whole network.

Step 4. Network weights are iteratively tuned by minimizing the cost function in (9) using MSGD optimization method.

After the network training, we can calculate the features of any input data, which are the output of the last encoding layer. We call the features learned by SDAE pretraining and LR fine-tuned-SDAE feature. It is worth noting that LR classifier is a part of the network. The output of LR layer, which is also the output of the whole network, denotes the classification results. Therefore, SDAE_LR obtains feature

extraction and classification simultaneously. In addition, besides LR, other supervised classifiers like support vector machine (SVM) can also be combined with the tuned-SDAE features.

3. Experimental Results and Analysis

3.1. Data Description. Three hyperspectral images were used for experiments. One was collected over Indian Pine (INP) in 1992. The spatial resolution of this image is 20 m; the available band for analysis of the image is 200 after removal of noisy and water absorption bands. One was acquired by Hyperion instrument over the Okavango Delta, Botswana (BOT), in May 2001. The 224-band Hyperion data have 10 nm spectral resolution over the range of 400 nm–2500 nm. The last high spatial resolution hyperspectral image was collected by reflective optics system imaging spectrometer (ROSIS) over the University of Pavia (PU), Italy. This data set has 103 dimensions of a spectral range from 430 nm to 860 nm, and its spatial resolution is 1.3 m. Both BOT and PU data contain 9 land cover types, and INP has 13 land cover types. Figure 3 shows the RGB images and the ground referenced information with class legends of BOT, PU, and INP images. Table 1 lists the class names and number of the three data sets.

3.2. Network Configuration. We firstly normalized the data in the range between 0 and 1 and then randomly selected 20 thousand data points from BOT, PU, and INP images, which were used for unsupervised pretraining of SDAE. In supervised LR training stage, we randomly divided the labeled data into training data, validation data, and testing data, with a ratio of 5 : 2 : 3. The training data are used in LR for fine-tuning, the validation data are for parameter tuning and termination of the iteration in MSGD method, and the testing data are for evaluating the algorithm.

Network configuration contains three parameters, which are the number of hidden layers, the number of units

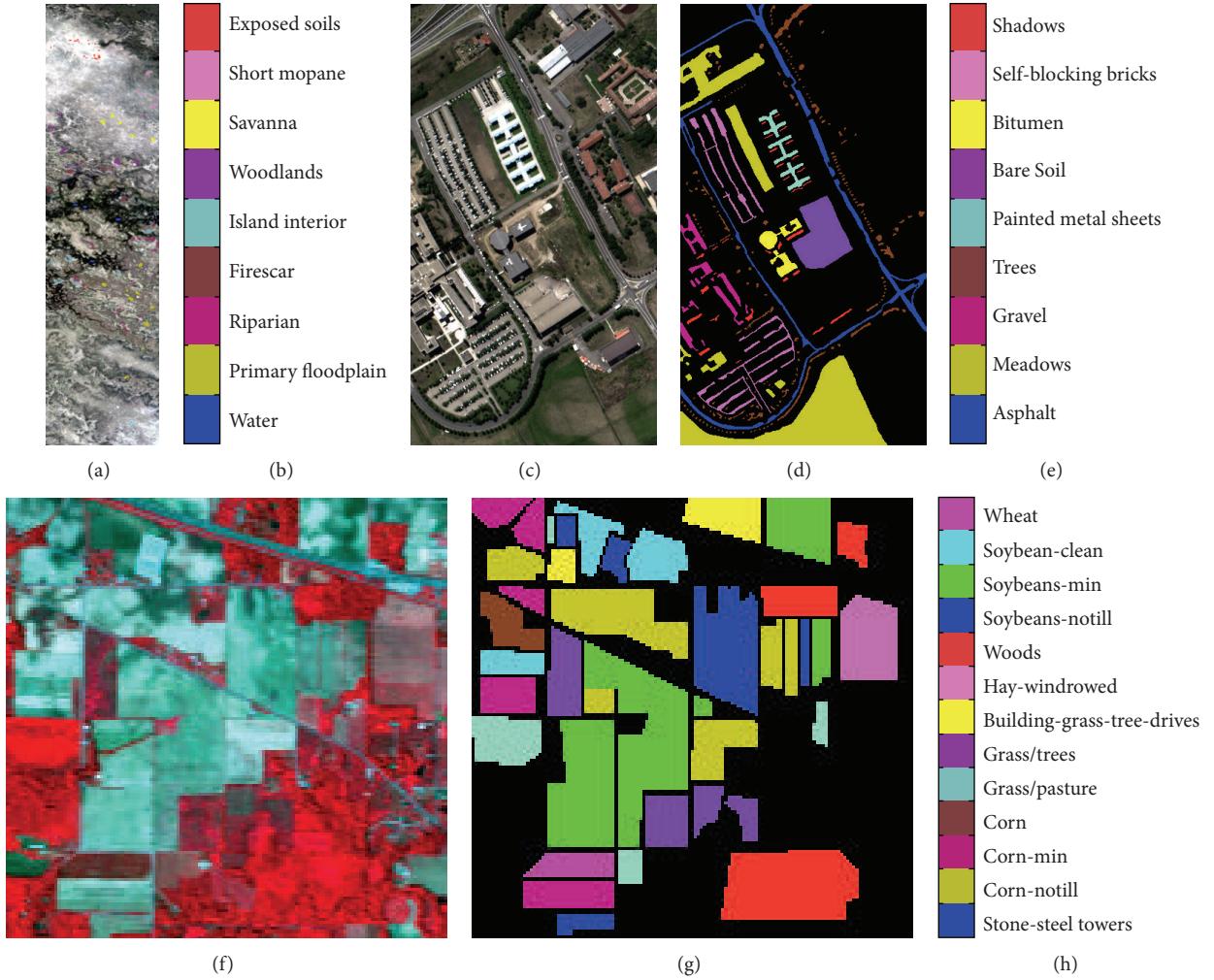


FIGURE 3: Three band false color composite and ground references. (a) False color composite of BOT image with ground reference. (b) Class legend of BOT image. (c) False color composite of PU image. (d) Ground reference of PU image. (e) Class legend of PU image. (f) IND PINE scene. (g) Ground reference of IND PINE image. (h) Class legend of IND PINE image.

TABLE 1: Class information of three datasets and the number of labeled samples in each class.

BOT		INP		PU	
ID	Class Name	ID	Class Name	ID	Class Name
1	Water (158)	1	Stone-steel Towers (95)	1	Asphalt (6631)
2	Floodplain (228)	2	Corn-notill (1434)	2	Meadows (18649)
3	Riparian (237)	3	Corn-min (834)	3	Gravel (2099)
4	Firescar (178)	4	Corn (234)	4	Trees (3064)
5	Island Interior (183)	5	Grass/Pasture (497)	5	Painted metal Sheets (1435)
6	Woodlands (199)	6	Grass/Trees (747)	6	Bare Soil (5029)
7	Savanna (162)	7	Building-Grass-Tree-Drives (380)	7	Bitumen (1330)
8	Mopane (124)	8	Hay-windrowed (489)	8	Self-Blocking Bricks (3682)
9	Exposed Soils (111)	9	Woods (1294)	9	Shadows (947)
		10	Soybeans-notill (968)		
		11	Soybeans-min (2468)		
		12	Soybean-clean (614)		
		13	Wheat (212)		

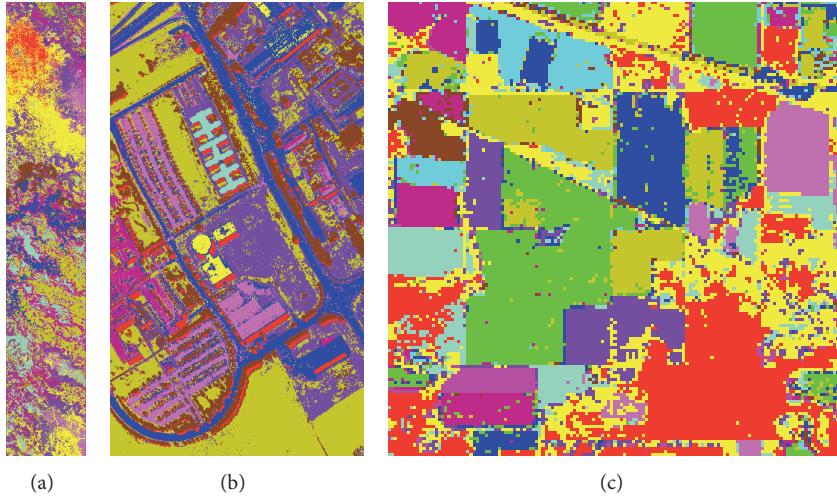


FIGURE 4: Classification results of the whole image on BOT (a), PU (b), and INP (c) data set.

in hidden layer, and the standard deviation of Gaussian noise. The number of hidden layers is selected in the range from 1 to 5, the number of units is chosen from [10, 30, 60, 100, 200, 300, 400], and the standard deviation of Gaussian noise is selected from [0.2, 0.4, 0.6, 0.8]. The optimal selection of these parameters is obtained according to the optimal classification results on the validation data. For BOT, PU, and INP data, the optimal number of layers is 4, 3, and 3, respectively; the best options for the number of units are 100, 300, and 200, respectively; the optimal selections of the standard deviation of Gaussian noise are 0.6, 0.6, and 0.2, respectively. In addition, network training includes two parameters: the epochs of pretraining and fine-tuning are set to be 200 and 1500, and the learning rates of pretraining and fine-tuning are selected as 0.01 and 0.1 empirically.

We used Theano for conducting the SDAE_LR classification. Theano is a Python library that can define, optimize, and evaluate mathematical expressions involving multidimensional arrays efficiently and can use GPU to speed up the calculation.

3.3. SDAE_LR Classification Performance. SDAE_LR method is compared with SVM classifier in this section, where SVM classifiers with linear and RBF kernels on the original data were conducted, which are denoted as LSVM and RSVM, respectively. The parameters in RSVM classifier are tuned by cross-validation method, and the penalty parameter in LSVM is set to be 2. The comparison results using overall accuracies (OA%) are shown in Table 2. It can be seen that the SDAE_LR outperformed LSVM for all the three data sets and obtained higher accuracies than RSVM on PU and INP data. It demonstrates that the features learned by the SDAE pretraining and LR fine-tuning can effectively increase the separation between classes. Figure 4 shows the classification results of the whole images using SDAE_LR for the three images. The acceptable results demonstrate good generalization ability of the SDAE_LR approach.

TABLE 2: Comparison of SDAE_LR and SVM classifier (OA%).

Data	LSVM	RSVM	SDAE_LR
BOT	92.88	96.88	95.53
PU	80.11	93.62	95.97
INP	76.15	90.63	92.06

TABLE 3: Comparison of computational time of SDAE_LR and SVM classifier (seconds).

Data	LSVM	RSVM	SDAE_LR
BOT	0.2632	142.1	94.68
PU	3.782	>12 h	1495
INP	2.727	5814	387.7

Using a machine with Intel Xeon CPU I7-4770, GPU NVIDIA Q4000, and 8 G RAM, the computational time of the three classifiers on BOT, PU, and INP data is shown in Table 3, where the LSVM and RSVM are implemented using CPU and SDAE_LR utilized GPU for computation. LSVM costs least time and RSVM is the most time-consuming because of the parameter tuning. We did not provide the exact time for RSVM on PU data since it is longer than 12 hours. The proposed SDAE_LR is much faster than RSVM, since it is implemented using Theano which accelerates the computation significantly. It is worth noting that the SDAE pretraining is fast and the LR fine-tuning costs time, because the former is layer-wise training and the latter propagates errors through the whole network.

3.4. Comparison of Different Feature Extraction Methods. Features of SDAE_LR network are obtained by SDAE pre-training and LR fine-tuning, which is called tuned-SDAE features. We compare the proposed method with four popular feature extraction methods, including PCA, Laplacian Eigenmaps (LE), locally linear embedding (LLE), and LDA.

TABLE 4: OA% using LSVM on different features.

Data	Raw	PCA	LE	LLE	LDA	Tuned-SDAE
BOT	92.88	88.62	93.70	94.03	89.28	97.02
PU	80.12	77.9	82.46	83.34	78.44	96.59
INP	76.15	66.56	75.06	71.3	74.32	91.92

TABLE 5: OA% using RSVM on different features.

Data	Raw	PCA	LE	LLE	LDA	Tuned-SDAE
BOT	96.88	90.83	92.27	94.14	96.57	95.69
PU	93.62	96.03	85.37	84.65	93.04	96.52
INP	90.63	90.52	78.44	73.86	86.1	92.47

The first three methods are unsupervised methods and LDA is supervised. In addition, PCA and LDA are linear methods, while LE and LLE are nonlinear methods. We set the number of features to be 50 for PCA, LE, and LLE empirically. The tuned-SDAE features are obtained by using the same network configuration described in Section 3.2.

After feature extraction by PCA, LE, LLE, LDA, and SDAE_LR, we used SVM classifiers (LSVM and RSVM) for classification. In addition, we also conducted SVMs on the raw hyperspectral data. Tables 4 and 5 show the overall accuracies of these methods using LSVM and RSVM, respectively. Several observations can be obtained: (1) for different feature extraction methods, tuned-SDAE performed the best. It significantly outperformed the others with the LSVM classifier for all the three data sets. When the RSVM classification was employed, the tuned-SDAE features also obtained the highest accuracies on most of the data sets; (2) compared to the SVM classification on the raw hyperspectral data, the four feature extraction methods (PCA, LE, LLE, and LDA) may not improve the accuracies, while the proposed tuned-SDAE features can consistently obtain better performance on most data sets; (3) in the four feature extraction methods (PCA, LE, LLE, and LDA), we cannot find one method that is consistently better than the others. The features obtained by SDAE_LR produced stable and good performances on all the data sets; (4) RSVM performed better than LSVM on the raw data and the features extracted by PCA, LE, LLE, and LDA, while RSVM and LSVM provided similar results on the tuned-SDAE features.

From the last column of Tables 2, 4, and 5, we can also observe that, with the tuned-SDAE features, different classifiers (LR, LSVM, and RSVM) resulted in similar performances. Within the three classifiers, LR is simplest since it is a part of the network, and the output of the network is the LR classification results.

Computational times of different feature extraction methods on the three data sets are listed in Table 6. Since the computational complexity of LE and LLE is $O(dN^2)$, where d is the number of dimension and N is the number of points, LE and LLE cannot process the large scale data sets. For PU data, we randomly selected 5000 data points for LE and LLE, and the features of the remaining data points are calculated by a kernel-based generalization method [1]. We can see that PCA

TABLE 6: Comparison of computational time of different feature extraction methods (seconds).

Data	PCA	LE	LLE	LDA	Tuned-SDAE
BOT	1.775	5.206	5.596	0.2864	94.68
PU	2.45	1022	70.14	0.3953	1495
INP	0.1918	9362	564.7	0.2788	387.7

and LDA are very fast. For BOT data, LE and LLE cost little time, while for INP and PU data, LE is very time-consuming and LLE also costs time, since the numbers of processed data points of INP and PU are much larger than BOT data. Feature extraction of SDAE_LR also requires times, especially for PU data where 20 thousand data points are used in LR fine-tuning stage.

3.5. Analysis of SDAE_LR Network Configuration. Firstly, we provided sensitivity analysis of three parameters in network configuration (number of hidden layers, number of units in each hidden layer, and standard deviation of Gaussian noise). Secondly, we demonstrated the effect of ReLU activation function. Thirdly, we tested the classification performances relative to different rates of training data.

Figure 5 shows the results of parameter analysis. When one parameter was tested, the values of other parameters were set to be values described in Section 3.2. (1) For the layers of the deep network, we tested five different values (1, 2, 3, 4, and 5), and the classification results are shown in Figure 5(a). For INP and PU data, the best number of layer is 3; for BOT data, the optimal selection is 4. Results on BOT and PU data are not sensitive to these parameters when the number of layer is larger than 2, while results on INP data indicate that only values of 2, 3, and 4 produced satisfactory performance. (2) For the number of units in each hidden layer, we evaluated seven different values (10, 30, 60, 100, 200, 300, and 400). As is shown in Figure 5(b), the best numbers of unit are 100, 300, and 200 for BOT, PU, and INP data, respectively. For INP data, small values like 10 deteriorate the classification performance. However, SDAE_LR is not very sensitive to this parameter in a large range (number of units > 100). (3) For the standard deviation of Gaussian noise, we tested four different values (0.2, 0.4, 0.6, and 0.8). The classification results with respect to this parameter is shown in Figure 5(c). The optimal values are 0.6, 0.6, and 0.2 for BOT, PU, and INP data, respectively. It can be seen that SDAE_LR is not very sensitive to this parameter.

Selection of activation function of the network is very important, and we chose ReLU function as activation function in this paper, since it is able to produce sparse features. To demonstrate the effectiveness of the sparsity, we compared two activation functions: ReLU function and sigmoid function, where the latter cannot obtain sparse features. The extracted features of SDAE_LR are the outputs of the last hidden layer, and therefore the dimensionality of features is equal to the number of units in the hidden layer. We define sparsity rate as the ratio of the number of zeros in the feature to the dimensionality of the feature. A high sparsity rate means there are many zeros in the feature and the feature is

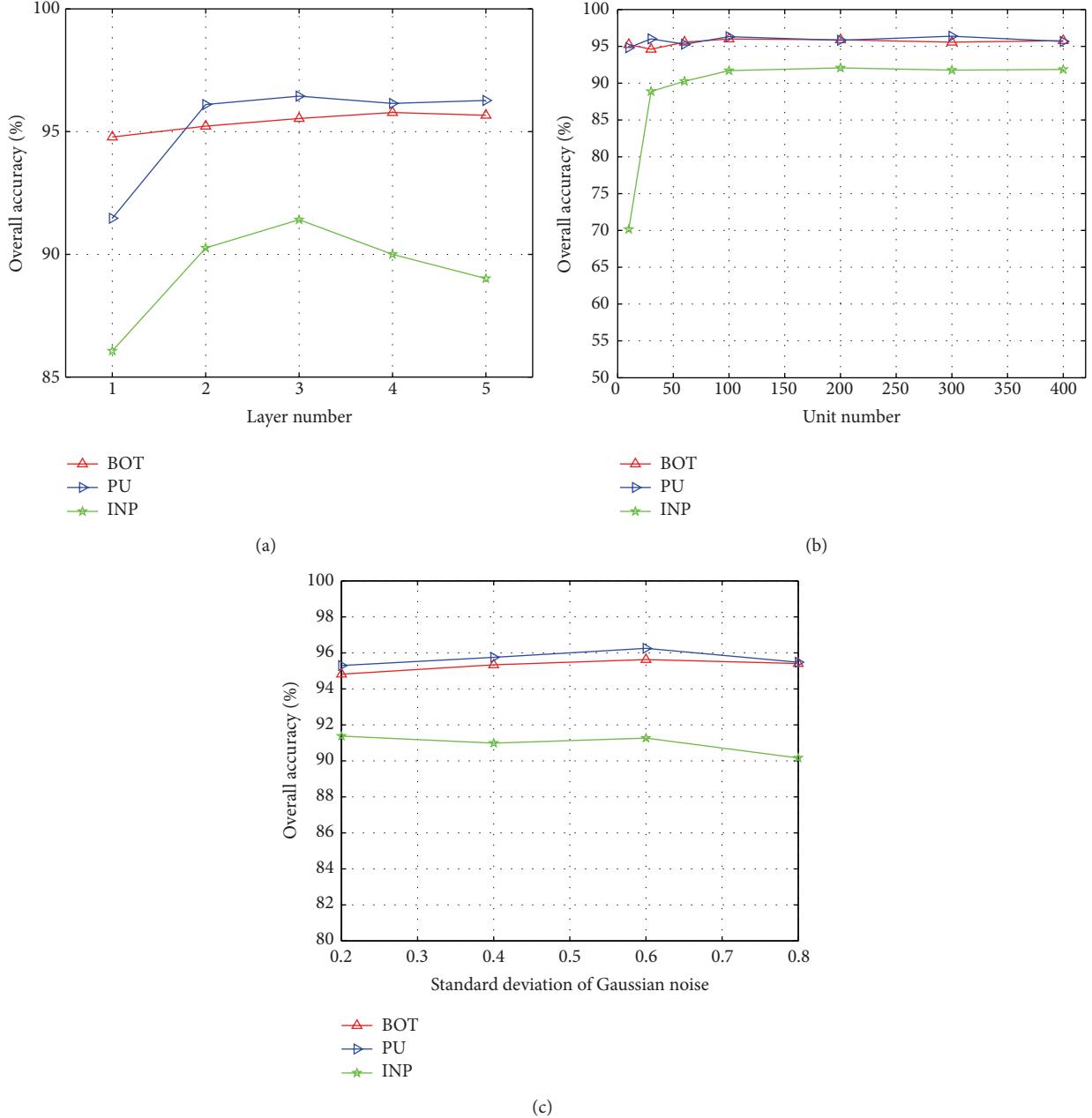


FIGURE 5: Parameter analysis of SDAE_LR approach. (a) For the parameter of the number of hidden layers. (b) For the parameter of the number of units in hidden layer. (c) For the parameter of standard deviation of Gaussian noise.

highly sparse. Figure 6 plots the sparsity rates versus different unit numbers of hidden layer using the ReLU activation function. With different number of units, the sparsity rate is high, and the number of nonzero values in the feature is small. Take PU data for example; when the number of unit is 400, the sparsity rate is 0.9626. It means the number of zeros in the feature is 385, and the feature only contains 15 nonzero values. Table 7 shows the OA using SDAE_LR with ReLU function and sigmoid function. It can be seen that ReLU function outperformed sigmoid function on all the three data sets, which demonstrates the efficiency of the sparse features using ReLU function.

TABLE 7: OA% with different activation functions.

Data	ReLU	Sigmoid
BOT	95.53	93.44
PU	95.97	76.12
INP	92.06	88.95

The number of training data also affects the network training, since LR fine-tuning is supervised and only training data can be used to further adjust the network weights. Figure 7 shows the SDAE_LR performance with respect to

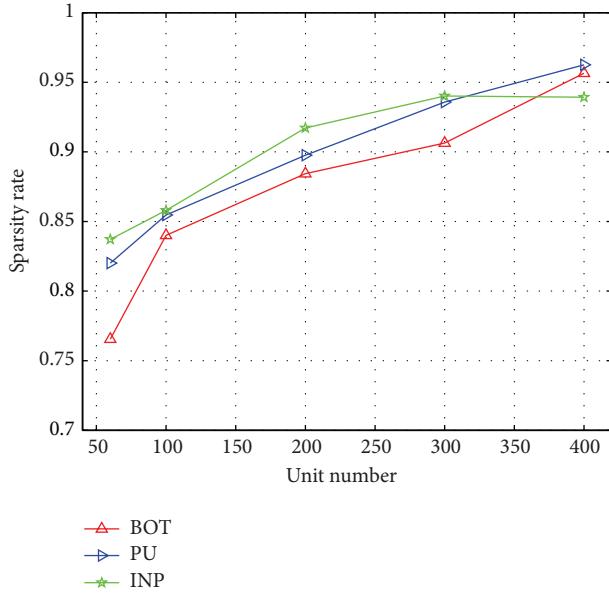


FIGURE 6: Sparsity rate of the network with different unit number of hidden layer.

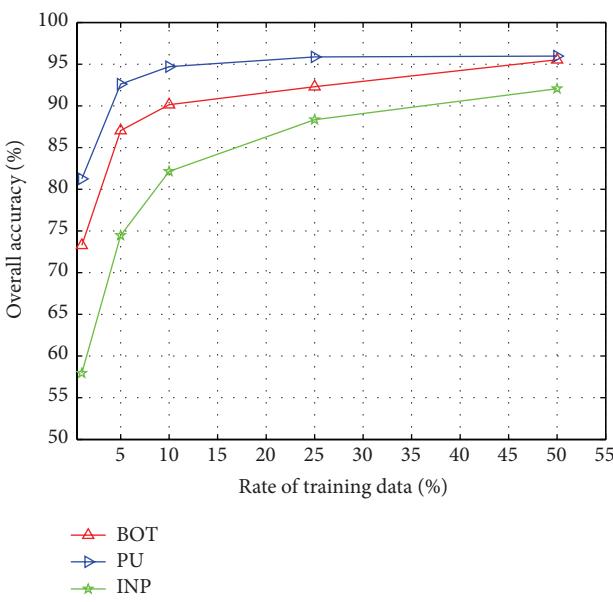


FIGURE 7: SDAE_LR classification performance with respect to the rates of training data.

different rates of training data (1%, 5%, 10%, 25%, and 50%). In general, high training data rates resulted in high accuracies, since LR performs supervised fine-tuning and classification.

4. Conclusion

Deep learning by SDAE_LR is proposed for hyperspectral feature extraction and classification, where SDAE pretrains the network in an unsupervised manner, and LR fine-tunes

the whole network and also performs classification. The features are learned by SDAE pretraining and LR fine-tuning. In the network, ReLU activation function was exploited to achieve the sparse features, which may improve the separation capability of the features. In experiments, SDAE_LR outperformed the popular SVM classifier with linear and RBF kernels. The tuned-SDAE features also provide better classification accuracies than several popular feature extraction methods, which demonstrates the good discriminant ability of extracted features. In SDAE, the utilized ReLU function performed better than sigmoid function, indicating the effect of the sparsity of features.

In SDAE_LR method, we only utilized spectral features of data. Plenty of spatial information of hyperspectral images can also be extracted and exploited [2, 19], such as the texture feature, morphological feature, the spatial coordinate information, and the relations between spatial adjacent pixels. Our further work is to combine spatial information in the SDAE_LR framework to further improve the classification performance.

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 National Natural Science Foundations of China (61102104, 81201067), the Fundamental Research Funds for National University, China University of Geosciences (Wuhan) (CUG120408, CUG120119), and Institute of Automation, Chinese Academy of Sciences.

References

- [1] L. Ma, M. M. Crawford, and J. Tian, “Local manifold learning-based k-nearest-neighbor for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4099–4109, 2010.
- [2] X. Huang and L. Zhang, “An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 257–272, 2013.
- [3] L. Ma, M. M. Crawford, X. Yang, and Y. Guo, “Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2832–2844, 2015.
- [4] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, “Robust L_2 estimation of transformation for non-rigid registration,” *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1115–1129, 2015.
- [5] J. Y. Ma, J. Zhao, Y. Ma, and J. W. Tian, “Non-rigid visible and infrared face registration via regularized gaussian fields criterion,” *Pattern Recognition*, vol. 48, no. 3, pp. 772–784, 2015.
- [6] L. J. P. van der Maaten, E. O. Postma, and H. J. van den Herik, “Dimensionality reduction: a comparative review,” 2008, http://www.iai.uni-bonn.de/~jz/dimensionality_reduction_a_comparative_review.pdf.

- [7] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, “Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding,” *Pattern Recognition*, vol. 48, no. 10, pp. 3102–3112, 2014.
- [8] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, U. D. Montral, and M. Qubec, “Greedy layer-wise training of deep networks,” in *Advances in Neural Information Processing Systems*, vol. 19, pp. 153–160, 2007.
- [9] C. C. Tan and C. Eswaran, “Reconstruction of handwritten digit images using autoencoder neural networks,” in *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 465–469, May 2008.
- [10] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103, 2008.
- [11] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [12] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [13] M. Norouzi, M. Ranjbar, and G. Mori, “Stacks of convolutional restricted boltzmann machines for shift-invariant feature learning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2735–2742, June 2009.
- [14] Y. S. Chen, Z. H. Lin, X. Zhao, G. Wang, and Y. F. Gu, “Deep learning-based classification of hyperspectral data,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.
- [15] F. Zhang, B. Du, and L. Zhang, “Saliency-guided unsupervised feature learning for scene classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2175–2184, 2015.
- [16] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion,” *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.
- [17] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, vol. 15, pp. 315–323, 2011.
- [18] L. Bottou, “Online algorithms and stochastic approximations,” in *Online Learning and Neural Networks*, D. Saad, Ed., pp. 9–42, Cambridge University Press, Cambridge, UK, 1998.
- [19] L. Zhang, L. Zhang, D. Tao, and X. Huang, “On combining multiple features for hyperspectral remote sensing image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, pp. 879–893, 2012.

Research Article

Design of an Active Multispectral SWIR Camera System for Skin Detection and Face Verification

Holger Steiner,¹ Sebastian Sporrer,¹ Andreas Kolb,² and Norbert Jung¹

¹*Safety and Security Research Institute (ISF), Bonn-Rhein-Sieg University of Applied Sciences, Grantham-Allee 20, 53757 Sankt Augustin, Germany*

²*Computer Graphics and Multimedia Systems Group, University of Siegen, 57076 Siegen, Germany*

Correspondence should be addressed to Holger Steiner; holger.steiner@h-brs.de

Received 23 December 2014; Accepted 5 March 2015

Academic Editor: Yongqiang Zhao

Copyright © 2016 Holger Steiner 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.

Biometric face recognition is becoming more frequently used in different application scenarios. However, spoofing attacks with facial disguises are still a serious problem for state of the art face recognition algorithms. This work proposes an approach to face verification based on spectral signatures of material surfaces in the short wave infrared (SWIR) range. They allow distinguishing authentic human skin reliably from other materials, independent of the skin type. We present the design of an active SWIR imaging system that acquires four-band multispectral image stacks in real-time. The system uses pulsed small band illumination, which allows for fast image acquisition and high spectral resolution and renders it widely independent of ambient light. After extracting the spectral signatures from the acquired images, detected faces can be verified or rejected by classifying the material as “skin” or “no-skin.” The approach is extensively evaluated with respect to both acquisition and classification performance. In addition, we present a database containing RGB and multispectral SWIR face images, as well as spectrometer measurements of a variety of subjects, which is used to evaluate our approach and will be made available to the research community by the time this work is published.

1. Introduction

Face recognition is a very important aspect for biometric systems and a very active research topic [1]. The human face has advantages over other biometric traits, as it can easily be captured in a nonintrusive way from a distance [2]. Consequently, biometric face recognition systems are becoming more frequently used, for example, at airports in the form of automated border control systems, for access control systems at critical infrastructure, or even for user log-on and authentication in computers or modern smartphones. However, despite the significant progress in the field, face recognition still faces serious problems in real-world scenarios when dealing with changing illumination conditions, poses, and facial expressions, as well as facial disguises (“fakes”), such as masks [3].

To overcome the problem of changing illumination conditions, the use of infrared imagery has been proposed in the recent years. Frontal illumination of faces with near infrared

light that is invisible to the human eye helps to reduce the influence of ambient light significantly without distracting or blinding the subjects [4].

For the detection of fakes, also referred to as liveness detection, at least three forms of spoofing have to be considered: photographs, prerecorded or live video (e.g., shown on a mobile device), and partial or complete facial disguises such as masks. The impact of such attacks on face recognition has been researched in several studies, for example, in the context of the research project TABULA RASA [5]. Although some countermeasures for such attacks have been proposed [6–8], especially the attacks with facial disguises and masks, they are still a problem for state of the art face recognition systems.

Masks can be manufactured using very different materials with varying textures and surface properties, for example, paper, latex, rubber, plastics, or silicon. Due to the variations found in human skin color and texture, distinguishing any possible material from genuine human skin using only the visual domain is a very difficult task [9].

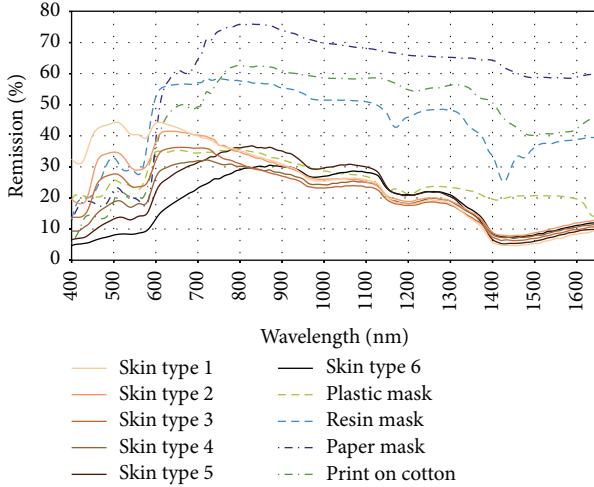


FIGURE 1: Spectral remission intensities of skin and different fake materials.

To overcome these problems, the use of infrared imaging has been proposed in prior work. Jacquez et al. [10] have shown that human skin has very specific remission characteristics in the infrared spectral range: the spectral remission of skin above 1200 nm is widely independent of the skin type and mainly influenced by the absorption spectrum of water. In addition, the spectral remission of most other materials differs strongly from that of skin: Figure 1 shows the remission intensities of human skin in the visual and infrared spectral range up to 1700 nm for six different skin types, denoted as skin types 1 (very light colored) to 6 (very dark colored) after Fitzpatrick [11], compared to the remission spectra of materials that might be used to create facial disguises.

In the literature, the infrared spectrum below 1400 nm is commonly referred to as the near infrared (NIR) band and the spectrum between 1400 nm and 3000 nm as the short wave infrared (SWIR) band. This work focuses on the spectral range of 900 nm up to 1700 nm. When describing this wavelength range, most researchers use only the term SWIR in order to distinguish it from work limited to the NIR range below 1000 nm. This paper will adopt this simplification and also use only the term SWIR in the following to describe this wavelength range. The existing approaches that make use of the SWIR spectral range can be classified into four groups: multispectral image acquisition using multiple cameras with band pass filters [9, 12], hyperspectral imagers [13], single cameras using filter-wheels with band pass filters for sequential multispectral image acquisition [14], and, more recently, single cameras with Bayer-like band pass filter patterns applied directly on the sensor [15]. All of these systems are passive (filter-based) and require sufficient illumination by daylight or external lighting. They will be discussed in detail in Section 2.

In our previous work, we presented an active multispectral point sensor for contactless skin detection which can be used for both safety and security applications, as well as a “proof of concept” of an active multispectral imaging system

[16, 17]. Both the sensor and the imaging system acquire a “spectral signature” of object surfaces: a specific combination of remission intensities in distinct, narrow wavebands that is used for the classification of the object’s surface material.

The contributions of this work are twofold.

(1) Based on our prior work, we present an improved system design of an active multispectral camera system optimized for face verification. The system acquires four-band multispectral image stacks in the SWIR range in real-time. The main improvements are

- (i) optimized illumination homogeneity,
- (ii) extensive camera system calibration,
- (iii) compensation of motion artifacts,
- (iv) advanced classification methods,
- (v) an elaborate evaluation regarding both skin detection and face verification.

(2) We present data from a study with more than 130 participants (at the time of writing) that combines spectral measurements at several points on faces and limbs with pictures taken with both an RGB camera and the presented multispectral camera system. A subset of this database, reduced by the images of participants that did not agree to publication, will be made available to the research community on our website (<http://isf.h-brs.de/>) by the time this work is published. We expect the published database to contain spectrometer data from at least 120 participants and image data from at least 50 participants.

The remainder of this paper is organized as follows: Section 2 gives an overview of the related work. Section 3 presents the design of the proposed camera system with a focus on hardware. Sections 4 and 5 describe the methods applied for image preprocessing and analysis. In Section 6, the camera system and the proposed skin and fake detection method are evaluated. For this purpose, a database of spectrometer measurements, as well as multispectral SWIR and RGB images, is presented. Section 7 concludes the paper.

2. Related Work

In the following, we will focus on work that is directly related to our approach, that is, based on the SWIR spectral range. A more general, comprehensive overview of methods for face recognition in the infrared spectrum, including the thermal infrared range, can be found in [3].

Taking advantage of the specific remission characteristics of human skin in the SWIR spectral range for its detection is not a new idea, but this approach has (to the best of our knowledge) only rarely been researched in the literature.

In 2000, Pavlidis and Symosek [9] demonstrated that the SWIR range has many advantages for face detection in general and for disguise detection in specific. They proposed a dual band camera system, consisting of two coregistered cameras, with one camera having a spectral sensitivity below

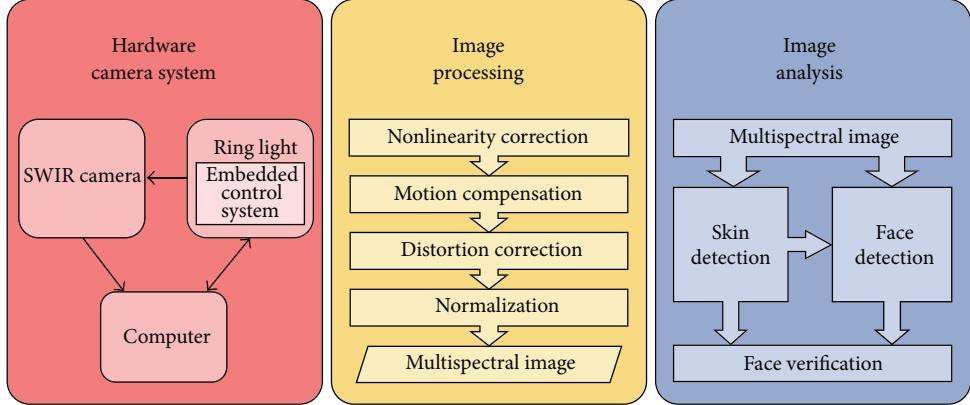


FIGURE 2: Building blocks of the proposed approach.

1400 nm (ideally 800 nm to 1400 nm) and the second camera having a spectral sensitivity above 1400 nm (ideally 1400 nm up to 2200 nm). Their system can work with either sunlight or artificial illumination and it uses a fusion algorithm based on weighted differences to detect skin in the acquired images. Depending on the spectral distribution of the illumination source, the weighting factors have to be adapted, as the system is not independent of ambient light. The authors conclude that their system achieves very good face and disguise detection capabilities compared to systems in the visual spectrum, only limited when it comes to the detection of surgical face alterations, where they see an advantage of systems using the thermal infrared range. In a later publication [18], they presented an extension of the system with a third camera for the visual spectrum and a more advanced face detection approach that included multiband eye and eyebrow detection. Their system uses beam splitters to allow all cameras to view the scene from the same vantage point in order to avoid problems with image registration.

At the U.S. Air Force Institute of Technology, Nunez and Mendenhall [12, 13] researched the use of hyperspectral SWIR imagery to detect skin for remote sensing applications. The authors acquired images in 81 narrow spectral bands between 900 nm and 1744 nm with a hyperspectral camera and introduced a detailed reflectance model of human skin based on this data. For real-time and in the field use, the authors propose a multicamera system to acquire images in distinct narrow wavebands using different band pass filters on each camera. To avoid problems with image registration, this system uses dichroic mirrors to split up the beam so that all cameras share one single lens and view the scene from the same vantage point.

More recently, Bourlai et al. [14] presented a multispectral SWIR image acquisition system using a single camera with an attached rotating filter wheel. The filter wheel is equipped with five band pass filters with a full width at half maximum (FWHM) of 100 nm around the peak wavelengths 1150 nm, 1250 nm, 1350 nm, 1450 nm, and 1550 nm. By synchronizing the camera's integration time to the filter wheel, the system can capture all five waveband images within 260 ms (i.e., at a rate of ≈ 3.8 frames per second (FPS)).

Bertozzi et al. [15] propose a camera with a broadband sensor for both the visual and SWIR spectral range (i.e., 400 nm to 1700 nm) that is equipped with a Bayer-like mosaic filter pattern directly on top of the pixel array. One clear filter (full bandwidth) is combined with three high pass filters with cut-off wavelengths of 540 nm, 1000 nm, and 1350 nm. By subtracting the acquired values of neighboring pixels with different filters, multispectral images in the four wavebands of approximately 400–600 nm, 600–1000 nm, 1000–1300 nm, and 1300–1700 nm can be calculated.

Due to the passive (filter-based) system design, the spectral distribution of the ambient illumination has a strong influence on the multispectral images acquired by any of these systems. In contrast to this, the approach proposed in this work uses active small band illumination instead of filters and is widely independent of ambient light. It combines a comparably high acquisition speed with high spectral resolution and robust detection.

3. Camera System Design

The approach described in this work is composed of three major building blocks illustrated in Figure 2, which we explain in sequential order. This section describes the design goals and decisions for the camera system with a focus on the hardware. Section 4 presents the low-level image processing methods, while Section 5 will focus on higher level image processing and analysis.

3.1. Design Goals. In general, face detection approaches in the context of biometric applications have strong requirements with respect to robustness and speed of the detection. Here, robustness includes both accurate detection under varying external conditions such as lighting and a reliable exclusion of spoofing attacks.

Even though we do not tackle any specific application scenario, we formulate the following, rather generic design goals that allow the realization of various applications.

- (i) The imaging system should be independent of ambient light. The spectral distribution or any flickering of

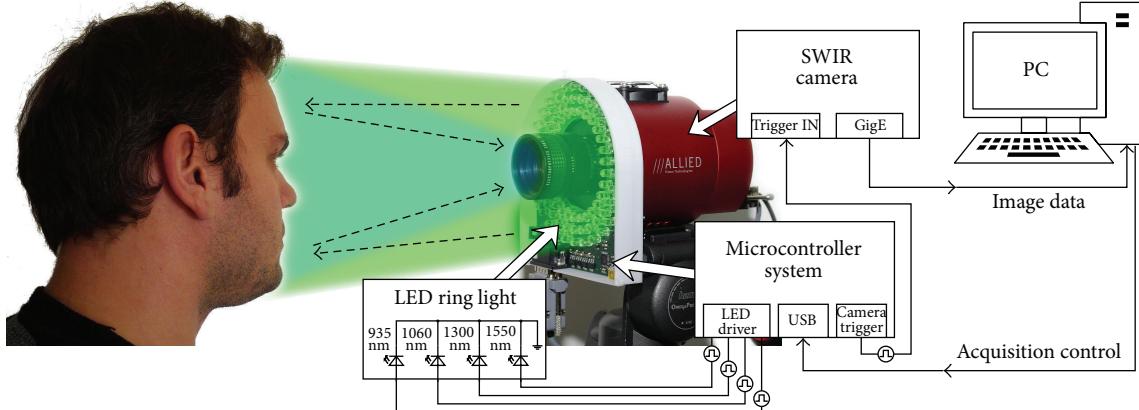


FIGURE 3: Schematic of the camera system setup.

- the light source must not distort the extracted spectral signatures.
- (ii) The acquisition time of a complete multispectral image stack should be as short as possible.
 - (iii) Moving objects must not lead to false classifications.
 - (iv) Face and disguise detection must work independent of a subject's skin type, age, or gender.
 - (v) The operation range should be oriented at typical cooperative user scenarios with short ranges of several meters (as opposed to long range imaging scenarios with distances of more than 100 meters [19]).
 - (vi) The system should require only one single camera. This avoids the need to align the optical path of multiple cameras or to apply complex image registration methods and reduces the costs of the imaging system, as SWIR cameras are still very expensive.

None of the existing approaches described in Section 2 can reach all of these goals.

3.2. System Setup. Based on the specified design goals, we propose a system setup consisting of a single SWIR camera sensitive to a spectral range of 900–1700 nm with an attached LED ring light that illuminates the face of a subject in four distinct narrow wavebands within this spectral range (one at a time), as illustrated in Figure 3. A microcontroller system, which is embedded into the ring light module, triggers short pulses in alternating distinct wavebands and signals the camera to start and stop the exposure of a new image synchronized to the light pulse. The camera transmits the acquired images to a connected computer via Gigabit Ethernet, which in turn is connected to the microcontroller system via USB in order to configure and start the acquisition. We also developed a special software tool that allows a user to control the image acquisition and to perform all related image processing and analysis tasks with a graphical user interface.

3.3. Design of the LED Ring Light. Using LEDs to implement the illumination module is an obvious choice, as they produce

TABLE 1: Number (n), peak wavelength (λ_p), FWHM ($\Delta\lambda_{0.5}$), radiated power (Φ_e), and total radiated power ($\sum \Phi_e$) of the LED types used on the ring light.

n	λ_p [nm]	$\Delta\lambda_{0.5}$ [nm]	Φ_e [mW]	$\sum \Phi_e$ [mW]
10	935	65	30	300
30	1060	50	5.5	165
20	1300	70	8.5	170
30	1550	130	5.0	150

rather narrow band illumination and can be pulsed with high intensities and variable frequencies. Based on findings in our previous work [16], we selected four wavebands for our current setup that are well suited for skin detection and designed an LED ring light with 90 LEDs. The number of LEDs for each waveband is shown in Table 1 and was chosen with regard to both the expected radiated power of each LED and a uniform distribution of the LEDs on the ring light.

A uniform distribution of the LEDs around the camera lens, as well as similar viewing angles and radiant patterns of the different LED types, is very important in order to achieve a homogeneous illumination. Otherwise, the extracted spectral signatures of an object would differ depending on the object's position in relation to the ring light. To avoid this problem, we selected LEDs of the same model and manufacturer (Roithner-Laser ELD-935-525, ELD-1060-525, ELD-1300-535, and ELD-1550-525) and performed optical simulations to find the optimal distribution of the different numbers of LEDs per waveband. For this purpose, we modeled the single LEDs as light sources using the FRED Optical Engineering (Photon Engineering LLC, <http://photonengr.com/>) software by specifying their typical peak wavelengths, spectral and radiant power distributions as defined by their datasheets. FRED performs ray tracing to simulate the propagation of light from each light source to a virtual target plane. It also provides a scripting language and batch processing capabilities to run a series of simulations with different parameters. This way, we compared different placement patterns and varying positions for the LEDs by simulating the resulting intensity distribution for

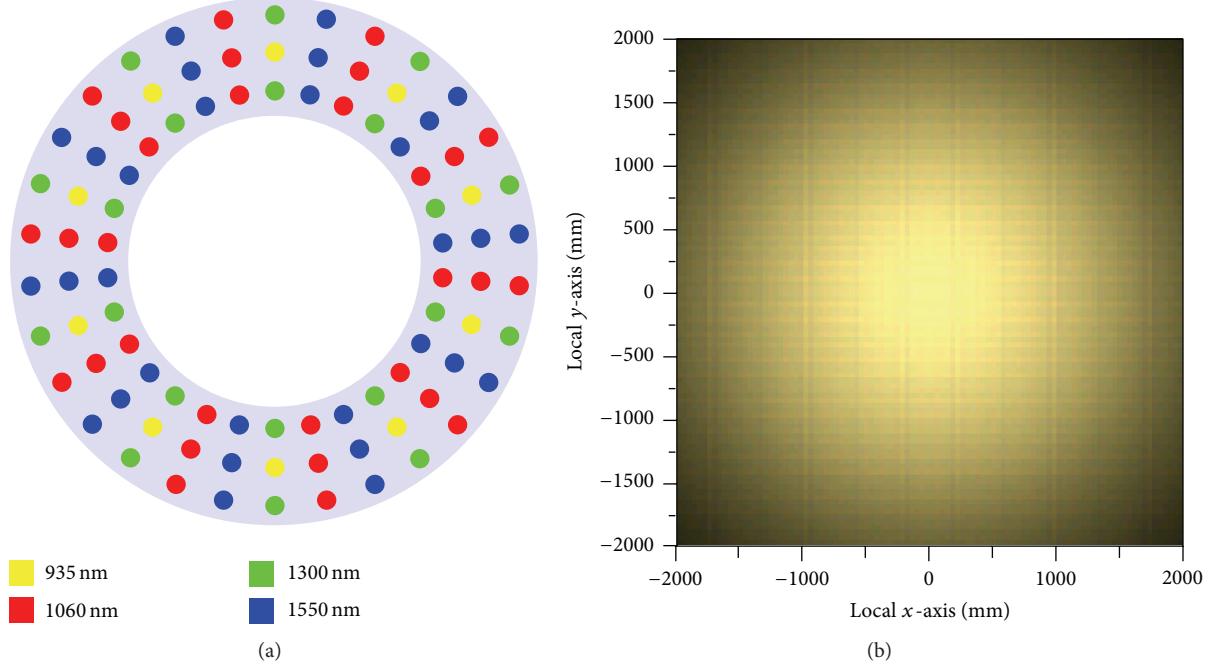


FIGURE 4: Distribution of the different LED types on the ring light (a) and resulting illumination distribution and homogeneity (b) projected on a virtual analysis plane in 5 m distance.

each waveband on the target plane. Ideally, the normalized intensity distributions of all wavebands should be identical, leading to a homogeneous “color” on the target. The best solution we found and the illumination distribution it created are shown in Figure 4. Due to the visualization of the four SWIR wavebands in yellow, red, green, and blue, the resulting (mixed) color is a soft yellow. Inhomogeneities would be noticeable by local changes in the color tone but cannot be observed.

3.4. Eye Safety Evaluation. Eye safety is a critical aspect of high power SWIR illumination sources, as radiation with a wavelength of up to 1400 nm can still penetrate the human eye and cause thermal damage to the retina. The directive 2006/25/EG of the European Union defines binding permissible limits for illumination systems with pulsed light sources, which should be measured as specified by the applicable standards. For our camera system, this is DIN EN 62471. The directive defines limits for the effective radiance L_R on the retina, which is weighted by a factor depending on the wavelength of the radiation, and the total irradiance E_{IR} on the cornea in a measurement distance of $d = 0.2$ m.

As the necessary measurement setup was not available to us, we analyzed the incident power of the SWIR radiation on the eye of an observer standing in the “sweet spot” of the ring light based on the optical simulation. Assuming a pupil diameter of 7 mm, the maximum incident power at a distance of $d = 0.2$ m is achieved by the 935 nm waveband and reaches a level of $P_{eye} = 0.69$ mW. This corresponds to a total irradiance of $E_{IR} \approx 17.3$ W/m².

TABLE 2: Effective radiance and total irradiance of the ring light’s most critical waveband (935 nm) on the eye of an observer in a distance of $d = 0.2$ m for short and very long observation times.

	L_R [W/(m ² ·sr)]	E_{IR} [W/m ²]
Simulation	—	17.3
Worst case scen.	1,355	25
Limit $t > 1000$ s	≈545,000	100
Limit $t \leq 10$ s	≥ $2.5 \cdot 10^6$	≥3,200

Using a model of our ring light that is simplified in the “safe direction”, we cross-checked this result using the specifications given in the LEDs datasheet. The typical radiant intensity of one 935 nm LED is given as $I_e = 0.1$ W/sr. Now we assume (at worst case) that all LEDs for the 935 nm waveband are continuously powered and directly adjacent, so that the combined radiant intensity of n LEDs can be approximated as $I \approx I_e \cdot n$ and the radiating surface as $A \approx n \cdot A_{LED}$. Now we can calculate L_R and E_{IR} as follows:

$$L_R = \frac{I}{A} \cdot R(\lambda) \quad (1)$$

$$E_{IR} = \frac{I}{d^2}$$

with $R(\lambda)$ being a correction factor according to directive 2006/25/EG and $d = 0.2$ m being the distance of an observer according to DIN EN 62471.

Table 2 shows both our results and the limits defined by the EU directive. As expected, the total irradiance calculated using the simplified “worst case” model is a little higher than

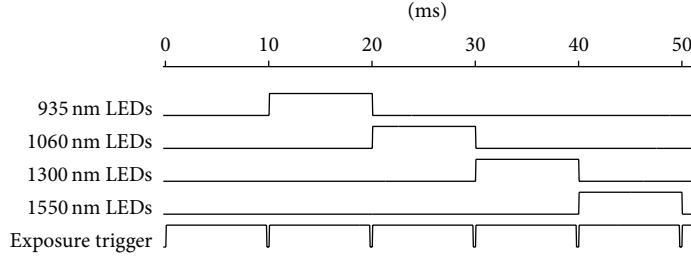


FIGURE 5: Timing diagram of the signals given by the microcontroller system.

the results from simulation, showing its plausibility. Still, the calculated values are by far below the permissible limits, even if the observer stares right into the ring light for a very long time. This leaves some headroom for further increases of the ring light's output power.

3.5. Image Acquisition Principle. In practice, the ring light is working as a pulsed light source. The microcontroller system enables its different wavebands one after the other in a fixed order and simultaneously triggers the camera exposure. To remove the influence of ambient light, in each acquisition cycle an additional camera exposure is triggered without the ring light flashing. This *reference image* is subtracted from each of the other images in preprocessing, so that only light emitted by the ring light in one single waveband remains on these images, which we call *waveband images*. Each set of waveband images and its corresponding reference image are combined in a *multiplespectral image stack*. This method works well for ambient light from continuous light sources, such as daylight. Here, all light sources with intensity variations that are either very slow or very fast compared to one full acquisition cycle can be regarded as continuous. However, “flickering” or pulsed light sources, changing their intensity with frequencies in a magnitude similar to the acquisition frequency, might cause distortions of the spectral signatures. In practice, most flickering light sources are incandescent or fluorescent lamps, flickering at twice the local power line frequency of 50 Hz or 60 Hz, therefore having periods of 10 ms or 8.3 ms, respectively. By using exposure times matching this period or any multiples of it, their influence can easily be reduced to a negligible level.

Our current setup is based on an Allied Vision Goldeye G-032 SWIR camera, which is equipped with an indium gallium arsenide (InGaAs) sensor and features a maximum frame rate of 100 frames per second (FPS) at its full resolution of $636 \cdot 508$ pixels with 14-bit A/D conversion. Due to the camera's very short readout time, it can be operated at this frame rate with an exposure time close enough to 10 ms to remove the effect of flickering lamps. Figure 5 illustrates the chronological order of the signals given by the microcontroller system within one full acquisition cycle of 50 ms, resulting in an effective frame rate of 20 FPS.

4. Image Preprocessing

Each image acquired by the SWIR camera is transmitted to a PC via Gigabit Ethernet. Simultaneously, the microcontroller

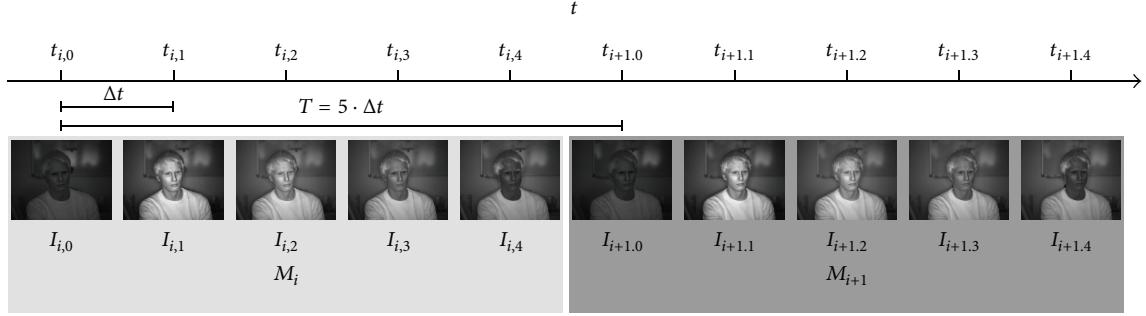
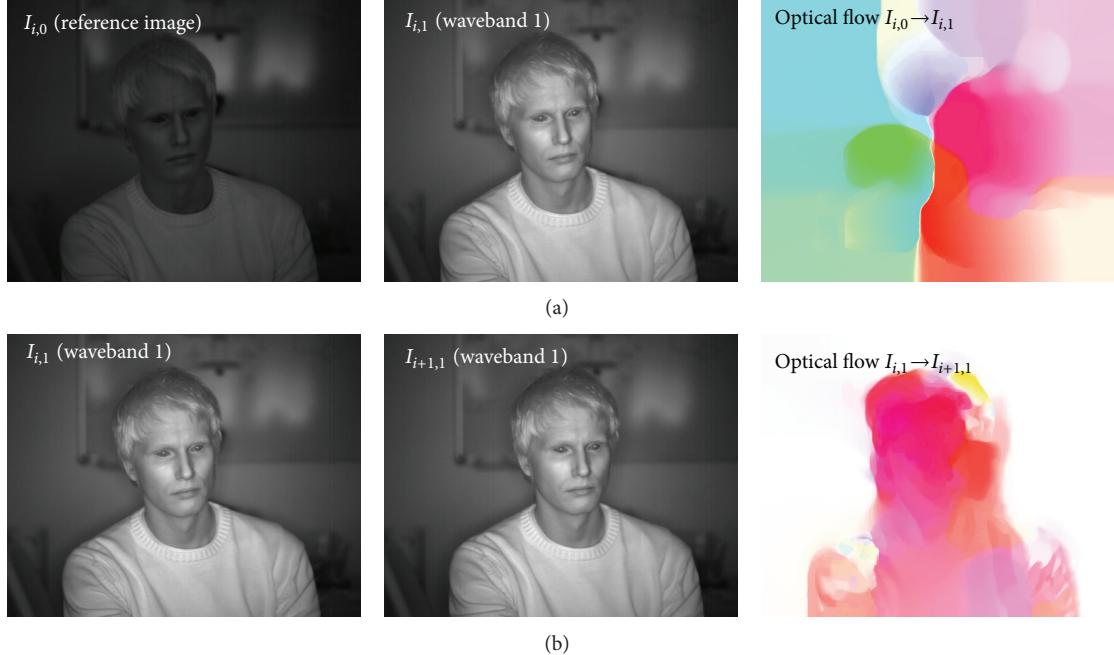
system tells the PC which waveband of the ring light has been active during the exposure via USB connection. Given this information, the software running on the PC performs several preprocessing steps to optimize and match the images in order to compose a multispectral image stack.

4.1. Fixed Pattern Noise Correction. Despite the camera's internal two-point nonlinearity correction (NUC), underexposed images show significant fixed pattern noise depending on the actual pixel intensity. As the system design requires taking one reference image without flashing the ring light, this noise will have an influence on images taken in dark environments. To analyze the sensor's behavior in detail, the sensor area was homogeneously illuminated using an adjustable quartz halogen lamp through an integrating (Ulbricht) sphere and 70 images with increasing brightness were taken. This image data is used as a look up table to apply a multiple-point nonlinearity correction to every single pixel. Figure 9 demonstrates the effectiveness of this method.

4.2. Motion Compensation. In the next step, the waveband images of one acquisition cycle are combined with a multispectral image stack. As the waveband images have been acquired sequentially, the positions of any moving object or person in the scene might have changed between each image of the stack. In practice, this will lead to motion artifacts and potentially cause false classifications due to distorted spectral signatures. This problem is common to all approaches that need to coregister sequentially acquired images, such as filter wheel camera systems [14].

To solve this problem, we propose a frame interpolation method based on motion estimation and compensation techniques to properly align all edges in every image of the stack. For this purpose, optical flow methods have proven to be a very effective, but computationally expensive approach [20]: sufficiently high performance for real-time applications can currently only be achieved by implementations using graphics hardware (GPUs). Hoegg et al. [21] demonstrated that this approach can also be used to compensate motion in coregistered sequential images acquired by a time of flight camera.

However, optical flow cannot be applied on our data directly, as illumination conditions and intensity values of object surfaces might differ strongly between the waveband images. In particular the first step in image merging, the subtraction of the (not actively illuminated) reference image, might cause problems: properly exposed image areas with

FIGURE 6: Image sequence of two successive multispectral image stacks M_i and M_{i+1} with $n = 5$ images.FIGURE 7: Optical flow between waveband image $I_{i,1}$ and its reference image $I_{i,0}$ (a) and between $I_{i,1}$ and its successor $I_{i+1,1}$ (b). Colored pixels represent detected motion.

much detail in the actively illuminated waveband images might be completely dark and without detail in the reference image.

Therefore, we use the following approach to motion compensation: consider a full multispectral image stack M_i , with $i \in \mathbb{N}$ being a sequential number, consisting of n images $I_{i,w}$, acquired at times $t_{i,w}$, $w = 0, \dots, n - 1$. Furthermore, we assume a discrete and equidistant acquisition time $\Delta t = t_{i,w+1} - t_{i,w}$ for each image $I_{i,w}$ and a constant acquisition time $T = t_{i+1,0} - t_{i,0} = n\Delta t$ for the full image stack, as illustrated in Figure 6.

As we cannot successfully apply optical flow directly to the sequence of images, that is, between $I_{i,w}$ and $I_{i,w+1}$ as shown in the upper row of Figure 7, we also consider a subsequent multispectral image stack M_{i+1} and apply optical

flow for corresponding images, that is, between $I_{i,w}$ and $I_{i+1,w}$, $w = 1, \dots, n - 1$ in a bidirectional manner resulting in a set of displacement maps (vector fields). Consider

$$\begin{aligned} \text{Forward flow: } & F_{(i,w) \rightarrow (i+1,w)}, \quad w = 1, \dots, n - 1 \\ \text{Backward flow: } & F_{(i+1,w) \rightarrow (i,w)}, \quad w = 1, \dots, n - 1. \end{aligned} \quad (2)$$

As $I_{i,w}$ and $I_{i+1,w}$ have both been acquired with the same illumination conditions, the results of this operation are much better, as shown in the lower row of Figure 7. Assuming a constant and linear motion between corresponding images $I_{i,w}$ and $I_{i+1,w}$, every vector $F_{(i,w) \rightarrow (i+1,w)}(x, y)$ in the displacement maps describing the movement of pixel (x, y) between $I_{i,w}$ and $I_{i+1,w}$ can be regarded as a linear combination of n identical partial vectors $(1/n)F_{(i,w) \rightarrow (i+1,w)}(x, y)$

describing a pixels movement between $I_{i,w}$ and $I_{i,w+1}$. Based on this assumption, we now apply the forward and backward displacement maps partially to estimate the images $\tilde{I}_{i+1,w}$ at intermediate times $t_{i+1,0}$, resulting in

$$\tilde{I}_{i+1,w} = \frac{(n-w)}{n} F_{(i,w) \rightarrow (i+1,w)} [I_{i,w}] + \frac{w}{n} F_{(i+1,w) \rightarrow (i,w)} [I_{i+1,w}], \quad (3)$$

where $F_{(j,w) \rightarrow (k,w)} [I_{j,w}]$ indicates the application of displacement map $F_{(j,w) \rightarrow (k,w)}$ to image $I_{j,w}$.

Finally, for all $\tilde{I}_{i+1,w}$, $w = 0, \dots, n-1$, the positions of moving objects will match their position in the reference image $I_{i+1,0}$. Thus, any further processing, that is, subtracting $I_{i+1,0}$ from every waveband image $I_{i+1,w}$, $w = 1, \dots, n-1$, and merging the images in one multispectral image stack, can be applied on this motion-corrected waveband images. For this application, the optical flow algorithm by Brox et al. [22], running on a GPU using a CUDA implementation, was found to be the best choice as it delivers very good results combined with acceptable run-times. Results of the motion compensation approach are presented in Section 6.

4.3. Calibration. With the multispectral image stack being properly aligned and the ambient illumination subtracted from all waveband images, lens distortion and differences in the illumination intensities can be corrected as last step in the image preprocessing. For this purpose, three sets of multispectral image stacks are recorded for each lens. A checkerboard calibration pattern is used to calculate a correction matrix for the lens distortion for every waveband individually to compensate for different distortion characteristics due to lateral chromatic aberration of the lens. Additionally, a plain white surface is used to measure both vignetting of the lens and light distribution of the ring light for each waveband and to calculate a respective correction matrix that normalizes the illumination intensity over the image area. Finally, a “white reference” tile with uniform remission characteristics in the SWIR spectral range is used to measure absolute differences in illumination intensities between the wavebands, which are stored as a vector of correction factors for each waveband. This waveband specific correction data is applied on every image of the multispectral image stack after the reference image has been subtracted.

5. Image Analysis

The multispectral image stacks acquired by the camera system are automatically analyzed by software in two steps: first, a skin classification method analyzes the spectral signature of each pixel to detect areas that show human skin. Second, a face detection algorithm searches for faces in the 1060 nm waveband image, as this waveband is very well suited for this purpose: the remission intensity of skin is comparably high, with eyes and mouth appearing darker. Finally, the locations of detected faces are matched against the results of the skin classification in order to verify their authenticity.

5.1. Skin Classification. To optimize both classification accuracy and run-time performance, the skin classification method consists of two algorithms, one for coarse-grained and one for fine-grained classification. Both algorithms perform pixelwise classification using the spectral signatures of the individual pixels as follows:

$$\vec{s}(x, y) = (g_1, \dots, g_{n-1}), \quad (4)$$

with each g_w , $1 \leq w < n$, being the greyscale value of the examined pixel (x, y) in spectral image $I_{i,w}$ of the multispectral image stack M_i , which consists of n spectral images.

For each pixel (x, y) , the first algorithm calculates normalized differences $d[g_a, g_b]$ for all possible combinations of greyscale values g_w within $\vec{s}(x, y)$ as follows:

$$d[g_a, g_b] = \left(\frac{g_a - g_b}{g_a + g_b} \right) \quad (5)$$

with $1 \leq a < n-1$ and $a < b < n$. So for $n = 5$, we get a vector of normalized differences \vec{d} with

$$\vec{d} = (d[g_1, g_2], d[g_1, g_3], d[g_1, g_4], d[g_2, g_3], \\ d[g_2, g_4], d[g_3, g_4]) \quad (6)$$

for each pixel (x, y) . The normalized differences range from $-1 \leq d[g_a, g_b] \leq +1$. In contrast to the values of the spectral signatures, they are independent of the absolute brightness of the analyzed pixel (x, y) , which differs with the measurement distance. This allows for a robust and fast classification of skin-like materials by specifying upper and lower thresholds for each normalized difference. However, this “difference filter” algorithm is not capable of distinguishing skin from materials that are very similar to skin, such as some kinds of silicon used for the creation of masks.

Therefore, a second classification algorithm is applied on the samples classified as “skin-like.” Based on results of our previous work [23], we use support vector machines (SVMs) for this fine-grained classification. The SVMs were trained using normalized difference vectors \vec{d} , which were calculated (as described above) based on spectral signatures extracted from multispectral images of skin, skin-like materials, and other materials acquired with the presented camera system. As shown in Section 6, the SVM classifier performs much better than the difference filter but has a much higher computational complexity. Limiting the SVM classification to those samples that have been positively classified by the difference filter significantly reduces the typical run-time of the skin detection. In addition, outliers and “unknown” material samples (samples that were not included in the training data) are less likely to create false positives when using two different classifiers. All pixels classified as skin are stored in a binary image I_{Skin} with 1 representing skin and 0 representing no-skin.

5.2. Face Detection. In the second step of the image analysis, we apply state of the art face detection algorithms on the 1060 nm waveband image to detect faces. We tested both



FIGURE 8: The multispectral image stack acquired by the camera system and an RGB counterpart.

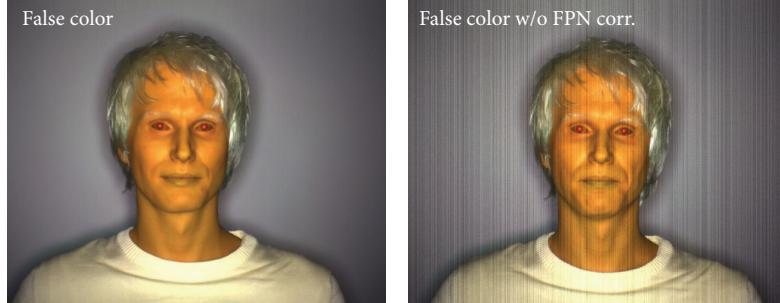


FIGURE 9: False color representation of the waveband images 1060 nm, 1300 nm, and 1550 nm with and without fixed pattern noise (FPN) correction.

the proprietary FaceVACS software from Cognitec Systems GmbH and an open source implementation of a local binary pattern histogram (LBPH) based face recognizer for this purpose. The result of the skin classification can optionally be used to improve the performance of the face detection algorithm by limiting the search for faces to areas in which skin has been detected. To verify the faces found in the image, their locations are matched with the result of the skin detection method. A face is verified as authentic if the ratio of “skin” to “no-skin” pixels within the facial area is above a specified threshold.

6. Results and Discussion

The results of our work are separated into four subsections: first, we present images acquired with our multispectral camera system and the results of the image processing methods. Second, we describe the design of a study with (by the time of writing) more than 130 participants and present the acquired

data, consisting of both spectrometer data and multispectral images. Based on this data, the performance and robustness of the proposed skin detection and classification approach are analyzed. Finally, the performance of the fake detection approach is evaluated.

6.1. Acquisition Quality and Performance. Figure 8 shows an example of the multispectral image stack acquired by our camera system after image processing, consisting of four waveband images and the reference image used to compensate for ambient light, as well as a color image taken with a high quality RGB camera for comparison.

Due to insufficiently corrected axial chromatic aberrations of the camera’s lens leading to a focus shift with increasing wavelengths, it is impossible to have all waveband images perfectly focused at the same time. This effect can only be reduced by stopping down the lens to a smaller aperture. As only the 1060 nm waveband image is used for

TABLE 3: Signal to noise ratio (SNR) of the ring light illumination for different target distances and ambient lighting conditions.

Distance [m]	Ambient light	SNR [dB]			
		935 nm	1060 nm	1300 nm	1550 nm
1.5	Dark	58	66	66	64
	Bright	34	42*	42*	41
2	Dark	53	62	62	60
	Bright	32	39	38	36
3	Dark	47	57	56	54
	Bright	23	33	32	30
4	Dark	44	53	52	51
	Bright	21	31	30	28

Values marked with * have been cut off due to sensor saturation.

face detection, we focus on this waveband image and accept a slight falloff in sharpness on the other waveband images.

6.1.1. Influence of Ambient Light. To evaluate the influence of ambient light on the camera system, a series of images of a reference target positioned in a distance of ≈ 1.5 m was taken with varying illumination conditions. The averaged illumination intensities measured on the reference target are shown in Figure 11. In this measurement, the ambient light is not yet subtracted from the signal pulses. Fluorescent lamps are barely visible for the SWIR camera, while daylight and incandescent lamps might increase the overall brightness significantly. Even without reaching saturation, the sensor shows some nonlinear behavior with increasing brightness levels: the actual signal strength, that is, the difference between the remission intensities with active ring light illumination and ambient light only, decreases by up to $\approx 20\%$ between dark and bright ambient illumination. However, the relative intensity differences between the wavebands stay almost the same and the influence on the normalized differences between the wavebands is only very small as long as the sensor is not saturated. Saturation can be avoided easily by dynamically reducing the exposure time. However, this will also reduce the acquired remission intensity of the SWIR pulses. Therefore, ambient light can be widely neglected but might reduce the maximum operation distance of the camera system.

6.1.2. Operation Range. The maximum operation distance of the camera system depends on several factors. The most important one is the radiated power of the ring light: with increasing distance to a target, the acquired remission intensities (the “signal”) will strongly decrease until they can no longer be distinguished from noise. In addition, as described before, with increasing ambient light the signal strength slightly decreases, while the absolute (shot) noise increases [24]. To evaluate the quality of the signal, we measured both the noise level in the reference image and the signal amplitude for a target at different distances in both dark and bright

environments and calculated the signal to noise ratio (SNR) according to [25] as follows:

$$\text{SNR} = \frac{\mu_{\text{Signal}}}{\sigma_{\text{Ref}}} \quad (7)$$

with μ_{Signal} being the average signal amplitude on the target and σ_{Ref} being the standard deviation within the same area in the reference image. Results are presented in Table 3. In our experiments, a $\text{SNR} \geq 20$ dB was enough to ensure reliable skin classification. Therefore, even in bright daylight conditions (overcast sky at noon), the system can operate at distances of up to at least 4 meters.

Besides the signal to noise ratio, the resolution and field of view of the camera system also put a limit on the operation range. For reliable face detection and recognition, current state of the art algorithms require the image of a face to have an eye-to-eye resolution of ≥ 60 pixels [4] or ≈ 1 pixel/mm. For our camera, we selected a lens with a focal length of 50 mm, which results in an angle of view of $\text{AOV} \approx 18^\circ$ and an operation distance of $d_{\min} \approx 0.7$ m $\leq d \leq d_{\max} \approx 2$ m.

6.1.3. Calibration Results. Figure 9 shows the effectiveness of the fixed pattern noise correction method: it presents a “false color” representation of the upper three wavebands before and after correction. The 1060 nm waveband is mapped to the red (R), the 1300 nm waveband to the green (G), and the 1550 nm waveband to the blue (B) channel.

An evaluation of the illumination intensity and homogeneity of the ring light showed some unexpected results. First, the 935 nm waveband appears much darker than the other wavebands, although the combined radiated power of all 935 nm LEDs is much higher than that of the other wavebands. A likely explanation is the characteristic of the camera’s sensor, which is less sensitive in this waveband. Second, despite coming from the same manufacturer and having similar packages, the different LED types have slightly different radiant patterns. Therefore, in practice, the light distribution is not as good as the simulated distribution. However, both the absolute intensity differences and the inhomogeneity can be corrected by applying the calibration data, as shown in Figure 12.

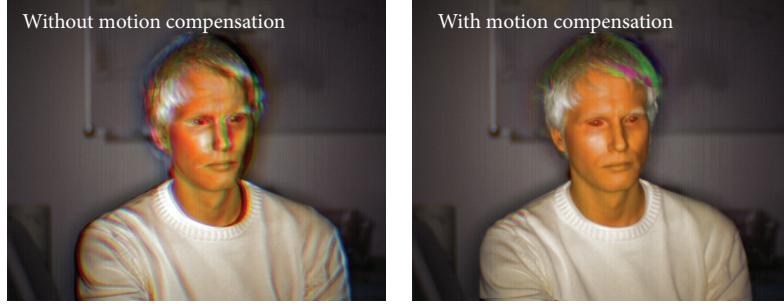


FIGURE 10: Results of the motion compensation method.

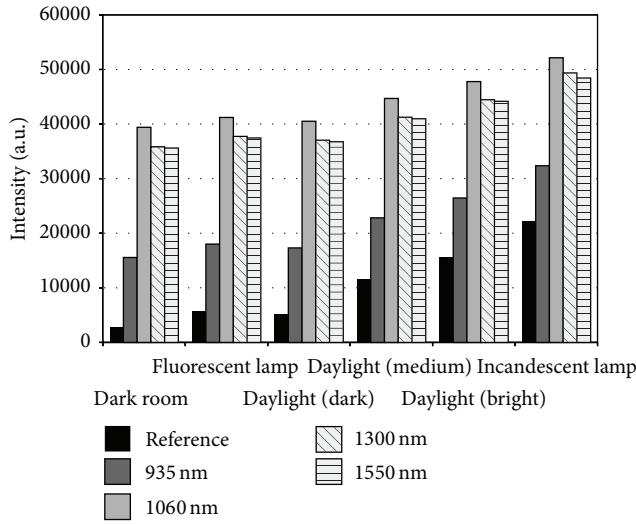


FIGURE 11: Remission intensities of the SWIR light pulses with increasing ambient light.

6.1.4. Motion Compensation. The results of the motion compensation approach are shown in Figure 10, with the original image on the left and the corrected image on the right, both represented as false color images with 3 wavebands. With a GPU-accelerated implementation using CUDA, the method based on the dense optical flow algorithm by Brox et al. [22] currently requires ≈ 110 ms to process the 3 images on our machine (intel Core i7 4771 CPU, nVidia GTX 780 graphics card, Ubuntu Linux 14.04 64 bit, GCC5.3, CUDA 6.5). When motion compensation is applied in real-time on a stream of acquired images, it becomes the bottleneck of the entire image processing chain and limits the frame rate of the camera system to currently ≈ 9 FPS with 3 or ≈ 6.5 FPS with 4 wavebands. Without motion compensation, the performance is only limited by the camera system's maximum frame rate of 25 FPS with 3 or 20 FPS with 4 wavebands.

6.2. Study Design. In order to evaluate the robustness of our approach to skin detection and to gather training data for the classification algorithms, we designed a study to acquire images of a representative number of persons with both our

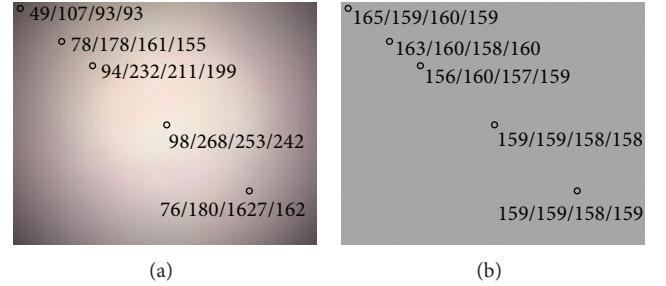


FIGURE 12: Ring light homogeneity before (a) and after (b) correction. Values represent pixel intensities in 935/1060/1300/1550 nm wavebands.

camera system and an RGB camera (Canon EOS 50D), as well as spectrometer (TQ irSys 1.7) data in the spectral range of 660 nm to 1700 nm. By the time of writing, the study is still ongoing. A subset of the resulting database, reduced by the images of participants that do not agree to publication, will be made available to the research community by the time this work is published.

In the following, we present data from 135 participants. Multispectral SWIR images were taken of all 135 persons (76 women, 59 men), while RGB images and spectrometer measurements have only been acquired for 120 of them (73 women, 47 men). As the study was conducted at our university, the most common skin types were 2 and 3 and most of our participants were between 20 and 29 years old with an average of ≈ 28 . The respective frequency distributions are shown in Tables 4 and 5. It has to be noted that several of our participants have been wearing make-up. As this will be a common situation in real-life applications, testing the influence of make-up was part of this study.

For each subject, spectrometer data was acquired at 16 measuring points on face and arms: 5 points on the face (forehead, nose, cheek frontal and sideways, and the chin), 3 at the neck (front, sideways, and back), 2 at the ear, 4 at the arm (front and back of both upper arm and forearm), and 2 at the hand (palm and back). These points have been chosen as they cover all skin regions that are typically expected in the field of view of a camera meant for face detection.

With both the RGB camera and the multispectral camera system, 7 portrait pictures were taken for each subject: three

TABLE 4: Age distribution of participants.

Age	<20	20–29	30–39	40–49	50–59
N	18	73	21	12	10

TABLE 5: Skin type distribution of participants.

Skin type	1	2	3	4	5	6
N	3	44	75	9	3	1

frontal shots with different facial expressions, two shots from an angle of $\pm 45^\circ$, and two profile shots from an angle of $\pm 90^\circ$. Subjects wearing glasses were asked to take them off for these shots. In this case, we added an additional image with glasses on for comparison.

In Figure 13, we present both RGB and (false color) multispectral SWIR portrait images of six participants of our study representing the skin types 1 to 6 after Fitzpatrick [11]. As expected, the obvious differences of the skin color in the RGB images are almost neglectable in the SWIR images.

6.3. Robustness of Skin Detection and Classification. In the following, we will analyze both spectrometer and camera data in detail in order to prove the validity of our approach to skin detection.

6.3.1. Spectrometer Data. For this evaluation, we used spectrometer data from only 8 of the 16 measuring points of 101 subjects, leaving out hands, arms, and ears, resulting in a total of 808 skin samples. We combined these samples with 336 samples of different materials (including different plastics, textiles, metal, and wood) and transformed the spectrometer data by applying a model of the ring light's LEDs in order to simulate the expected spectral signatures \vec{s}' of the camera system. For this purpose, each samples' reflectance spectrum is convoluted with each LED's emission spectrum [26].

We calculated the normalized differences \vec{d}' between all wavebands of the spectral signatures \vec{s}' for all samples and applied a principal component analysis (PCA) on the data set. Figure 14 presents a plot of the two main components, which already separate most of the samples. Using difference filters by specifying minimum and maximum thresholds for each normalized difference in \vec{d}' , all skin samples can be separated perfectly from all material samples, as shown in Table 6.

6.3.2. Camera Data. To analyze the data acquired with the camera system, we extracted the spectral signatures of skin and a variety of other materials from the images taken during the study. Pixels showing skin are stored as positive examples and “no-skin” pixels as negative examples. Similar to the spectrometer data, we applied a PCA on this data set. The two main components are illustrated in Figure 15 and perfectly separate the two classes. However, the difference filter classifier cannot separate all skin samples from all

TABLE 6: Confusion matrix of the difference filter classifier applied to the spectrometer data set.

		Predicted class	
		Skin	Material
Actual class	Skin	808	0
	Material	0	336

TABLE 7: Confusion matrix of the difference filter classifier applied to the camera data set.

		Predicted class	
		Skin	Material
Actual class	Skin	77771	0
	Material	1561	99773

TABLE 8: Confusion matrix of the SVM classifier applied to the camera data set.

		Predicted class	
		Skin	Material
Actual class	Skin	77771	0
	Material	0	101334

material samples, as shown in Table 7: some material samples belonging to “CP-Flesh,” a silicon mixture specifically designed to imitate human skin, show up as false positives. Therefore, we used LibSVM to train a SVM classifier on the data set. To evaluate the SVM’s performance, we applied a tenfold cross validation, with each fold randomly choosing 90% of the samples for training and 10% for testing. The results of the SVM are shown in Table 8: skin and material can be separated perfectly.

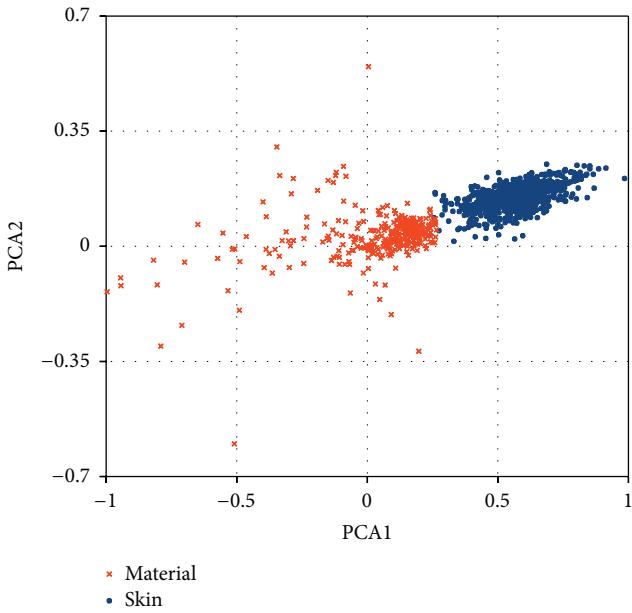
By analyzing the data and reviewing the acquired images, we did not find a significant influence of make-up on the skin classification results. Therefore, we asked one subject to use very large amounts of make-up and powder and acquired additional images. We found that only very thick layers of powder, which are clearly visible in both the RGB and the SWIR images, could influence the spectral signatures enough to lead to false negative results. Therefore, our approach to skin detection proves to be robust against different skin types, typical make-up, and varying measurement conditions.

6.4. Evaluation of Face Verification. To analyze the face verification performance of the presented camera system, we first evaluated the usability and quality of the acquired images for face detection. Then, we tested the skin classification performance of our approach on different fakes and compared the results to the acceptance rate of state of the art face recognition software.

6.4.1. Usability of SWIR Images. To evaluate the usability of the SWIR images, we trained both the proprietary state of the

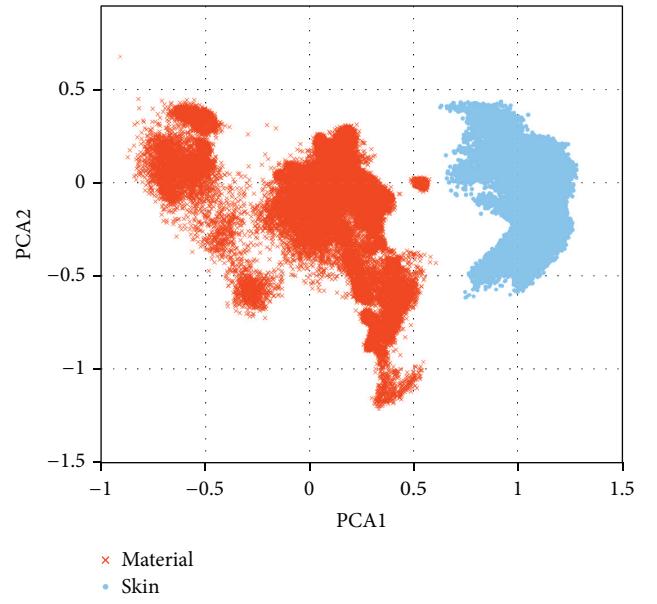


FIGURE 13: RGB and SWIR portrait images of skin types 1 to 6 according to Fitzpatrick [11].

FIGURE 14: Two main components (PCA1/PCA2) of a principal component analysis applied to normalized differences \vec{d} of the spectrometer data.

art FaceVACS and the openCV implementation of the LBPH face recognizer with the RGB face images acquired in the context of our study. Then we fed the algorithms with the face images acquired with the multispectral camera system and tried to identify and verify the faces using only the 1060 nm waveband image.

FaceVACS identified all faces correctly. Furthermore, it verified 73% of all faces with a probability score of $PS \geq 99.9\%$ and 92% with $PS \geq 98\%$. Only $< 4\%$ of all faces were verified with a probability score of $PS < 90\%$, with $PS_{min} = 76\%$ being the minimum. These rare examples of low probability have been investigated in detail and might be caused by strong highlights in the eyes (reflections from the ring light) or

FIGURE 15: Two main components (PCA1/PCA2) of a principal component analysis applied to normalized differences \vec{d} of the camera data.

differing head poses. However, the acceptance threshold of 70% was met by all test images.

In contrast to this, the LBPH face recognizer did a surprisingly bad job: it identified only 22% of all 1060 nm face images correctly and calculated very low confidence values for those that it actually verified. We compared this result to its performance when trained on additional SWIR images (which were not part of the test samples) and got a much better result of 92% with much better confidence values for the verified test images. We conclude that the classifier used by this face recognizer uses features that are not invariant to absolute greyscale values and excluded this algorithm from the further evaluation.

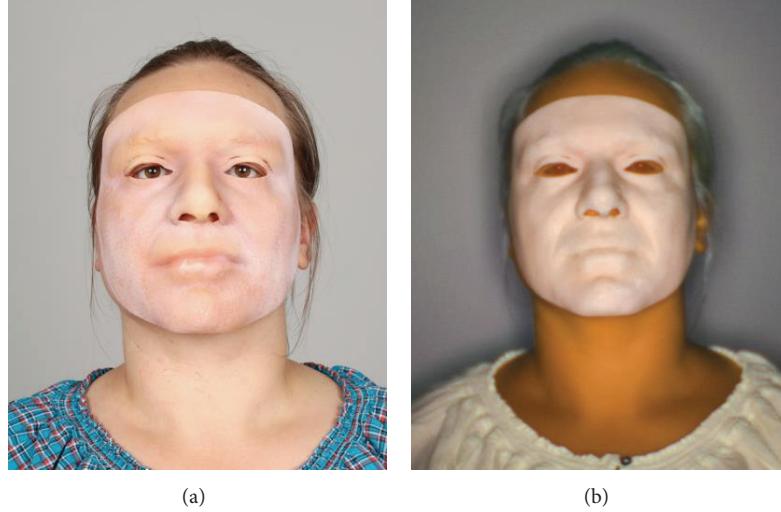


FIGURE 16: RGB (a) and SWIR false color (b) image of a subject wearing a mask created using a 3D printer.

6.4.2. Fake Detection Performance. In the context of a previous research project together with the German Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik, BSI), several photo-fakes and masks, which mimic the face of one of our test subjects, were manufactured in order to test the vulnerability of face recognition systems and to develop respective countermeasures. Different materials have been used for these masks, including special silicon mixtures, plastics, hard resin, textiles, and paper. Make-up and paint have been applied to the masks to make them more realistic. With the genuine face of the subject enrolled in FaceVACS, all fakes and masks achieved a probability of more than 70% in verification when pictured using an RGB camera and were accepted, except for the paper mask. In particular photo-fakes and prints on t-shirts achieved very high scores in FaceVACS due to a missing liveness detection.

Using images acquired with our camera system, most of the fakes achieved much lower scores in FaceVACS even without skin classification, because the colorants used are less visible in the SWIR range. This applies to most of the photo-fakes and prints, as well as plastics and hard resin masks: verification scores drop from 99.9% down to less than 20%. Figure 16 shows RGB and SWIR (false color) images of a subject wearing a mask created using a 3D printer, which is much easier to detect in the SWIR image. Transparent silicon masks, however, are still a significant problem.

Adding spectrometer measurements of all fakes to our database and training a new difference filter classifier showed that none of the skin-like fakes, such as the silicon masks, could be separated from skin easily. The same holds true for the camera data: we added images of the fakes to our data set and applied the difference filter classifier on it. The results are shown in Table 9: with this data set, more than 10% of the material samples are classified as skin, namely, all of the silicon masks. Fortunately, a SVM classifier produces a much better result and achieves a precision of 99.968% in a tenfold

TABLE 9: Confusion matrix of the difference filter classifier applied to the camera data set including fakes.

		Predicted class	
		Skin	Material
Actual class	Skin	146821	0
	Material	22933	227104

TABLE 10: Confusion matrix of the SVM classifier applied to the camera data set including fakes.

		Predicted class	
		Skin	Material
Actual class	Skin	146734	87
	Material	40	249997

cross validation, as shown in Table 10: 87 (0.059%) of the skin samples are rejected, but only 40 (0.016%) of the material samples are classified as skin. As each sample is a single pixel of an image, this error will not have a big influence in reality.

Finally, we tested the classifiers with a new data set. We took images of two subjects with and without silicon masks and applied both the difference filter and the SVM classifier successively on the images. The results of the difference filter are shown in the upper half of Figure 17: the classifier detects all skin pixels correctly but also classifies most of the fake pixels as skin.

A set of both true and false positive samples from the results of the difference filter classifier was annotated with correct classes and used as test set for the SVM classifier. The results are almost perfect, as shown in Table 11: only 16 samples (=pixels) of the fake material are still classified as skin, while no true skin pixels were rejected. These results also hold true in practice, as shown in the lower half of Figure 17: only pixels showing uncovered skin are left in the image,

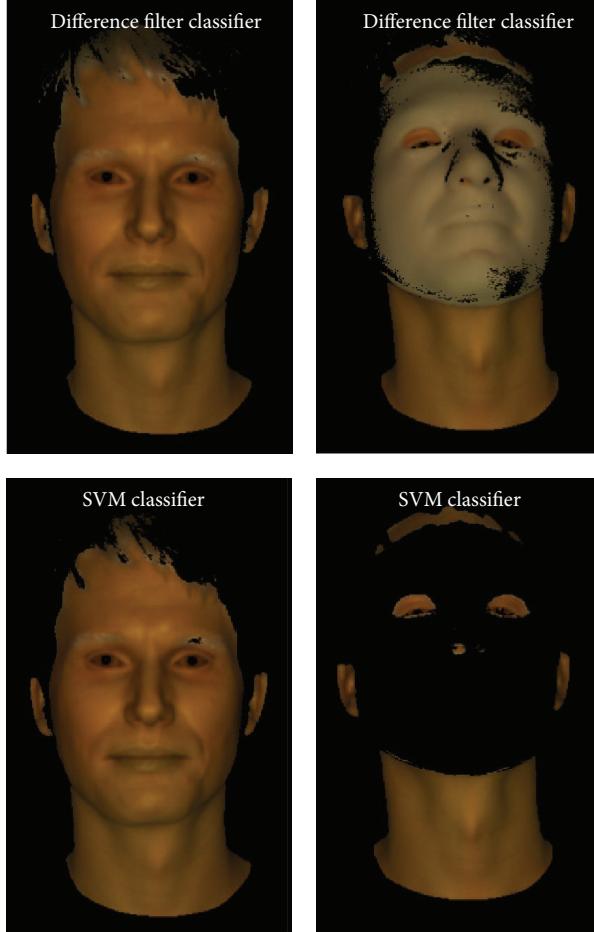


FIGURE 17: Resulting images after difference filter (top) and SVM classifier (bottom) on a face image without (left) and with (right) a silicon mask.

TABLE II: Confusion matrix of the SVM classifier applied to the fake testing data set (prefiltered by the difference filter classifier).

		Predicted class	
		Skin	Material
Actual class	Skin	76783	0
	Material	16	47995

while the mask pixels are rejected. Thus, the detected faces without masks are verified as authentic with a very high ratio of skin to no-skin pixels within the facial area, while the faces with masks are reliably rejected as fakes.

7. Conclusions

We proposed an active multispectral SWIR camera system for real-time face detection and verification. The system acquires four-band multispectral image stacks within an acquisition time of 50 ms. The extraction of spectral signatures from the acquired images allows for reliable skin detection independent of skin type. Our approach requires only one SWIR camera and uses active small band illumination based on

pulsed LEDs, making it widely independent of ambient light. Motion artifacts at moving objects due to sequential acquisition of waveband images are effectively removed by using optical flow based motion compensation techniques. The system can be used for a variety of application scenarios without the need for regular calibration.

For the application of face detection, recognition, and verification, the active frontal SWIR illumination ensures robust face detection and extraction of facial features. Based on the acquired multispectral images, the proposed analysis methods allow detecting spoofing attacks using fakes or facial disguises such as silicon masks, which are still a big problem for state of the art face recognition systems, with significantly improved reliability.

In addition to the camera system, we presented a database of face images from several subjects in different poses and perspectives, acquired with both our camera system and an RGB camera, supplemented by spectrometer data in the wavelength range between 660 nm and 1700 nm. This database will be made available to the research community on our website by the time this work is published.

Conflict of Interests

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

Acknowledgments

This work was funded by the German Federal Ministry of Education and Research as part of the program “FHprofUnt” (FKZ: 03FH044PX3) and supported by the German Research Foundation (DFG) as part of the research training group GRK 1564 “Imaging New Modalities.” The authors also thankfully acknowledge the support of the German Federal Office for Information Security (BSI).

References

- [1] N. Erdogan and S. Marcel, “Spoofing face recognition with 3d masks,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1084–1097, 2014.
- [2] T. Bourlai and B. Cukic, “Multi-spectral face recognition: identification of people in difficult environments,” in *Proceedings of the 10th IEEE International Conference on Intelligence and Security Informatics (ISI ’12)*, pp. 196–201, June 2012.
- [3] R. Shoja Ghiass, O. Arandjelović, A. Bendada, and X. Mal dague, “Infrared face recognition: a comprehensive review of methodologies and databases,” *Pattern Recognition*, vol. 47, no. 9, pp. 2807–2824, 2014.
- [4] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, Springer, 2nd edition, 2011.
- [5] N. Kose and J.-L. Dugelay, “On the vulnerability of face recognition systems to spoofing mask attacks,” in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’13)*, pp. 2357–2361, May 2013.
- [6] K. Kollreider, H. Fronthaler, and J. Bigun, “Non-intrusive liveness detection by face images,” *Image and Vision Computing*, vol. 27, no. 3, pp. 233–244, 2009.

- [7] J. Komulainen, A. Hadid, and M. Pietikainen, "Context based face anti-spoofing," in *Proceedings of the 6th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '13)*, pp. 1–8, September 2013.
- [8] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proceedings of the 6th International Conference on Biometrics (ICB '13)*, pp. 1–6, June 2013.
- [9] I. Pavlidis and P. Symosek, "The imaging issue in an automatic face/disguise detection system," in *Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum (CVBVS '00)*, pp. 15–24, Hilton Head, SC, USA, 2000.
- [10] J. A. Jacquez, J. Huss, W. McKEEhan, J. M. Dimitroff, and H. F. Kuppenheim, "Spectral reflectance of human skin in the region 0.7–2.6 μ m," *Journal of Applied Physiology*, vol. 8, no. 3, pp. 297–299, 1955.
- [11] T. B. Fitzpatrick, "The validity and practicality of sun-reactive skin types I through VI," *Archives of Dermatology*, vol. 124, no. 6, pp. 869–871, 1988.
- [12] A. S. Nunez, *A physical model of human skin and its application for search and rescue [Ph.D. thesis]*, Air Force Institute of Technology, Wright-Patterson Air Force Base, School of Engineering, Wright-Patterson, Ohio, USA, 2009.
- [13] A. S. Nunez and M. J. Mendenhall, "Detection of human skin in near infrared hyperspectral imagery," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '08)*, pp. II621–II624, July 2008.
- [14] T. Bourlai, N. Narang, B. Cukic, and L. Hornak, "On designing a SWIR multi-wavelength facial-based acquisition system," in *Infrared Technology and Applications*, vol. 8353 of *Proceedings of SPIE*, April 2012.
- [15] M. Bertozi, R. Fedriga, A. Miron, and J.-L. Reverchon, "Pedestrian detection in poor visibility conditions: would SWIR help?" in *Image Analysis and Processing—ICIAP 2013*, A. Petrosino, Ed., vol. 8157 of *Lecture Notes in Computer Science*, pp. 229–238, Springer, Berlin, Germany, 2013.
- [16] O. Schwaneberg, H. Steiner, P. H. Bolívar, and N. Jung, "Design of an LED-based sensor system to distinguish human skin from workpieces in safety applications," *Applied Optics*, vol. 51, no. 12, pp. 1865–1871, 2012.
- [17] H. Steiner, O. Schwaneberg, and N. Jung, "Advances in active near-infrared sensor systems for material classification," in *Proceedings of the Imaging Systems and Applications*, Optical Society of America, Monterey, Calif, USA, June 2012.
- [18] J. Dowdall, I. Pavlidis, and G. Bebis, "Face detection in the near-ir spectrum," *Image and Vision Computing*, vol. 21, no. 7, pp. 565–578, 2003.
- [19] B. E. Lemoff, R. B. Martin, M. Sluch, K. M. Kafka, A. Dolby, and R. Ice, "Automated, long-range, night/day, active-SWIR face recognition system," in *40th Infrared Technology and Applications*, vol. 9070 of *Proceedings of SPIE*, pp. 90703I-1–90703I-10, Baltimore, Md, USA, June 2014.
- [20] L. Raket, L. Roholm, A. Bruhn, and J. Weickert, "Motion compensated frame interpolation with a symmetric optical flow constraint," in *Advances in Visual Computing*, vol. 7431 of *Lecture Notes in Computer Science*, pp. 447–457, Springer, Berlin, Germany, 2012.
- [21] T. Hoegg, D. Lefloch, and A. Kolb, "Realtime motion artifact compensation for pmd-tof images," in *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, vol. 8200 of *Lecture Notes in Computer Science*, pp. 273–288, Springer, Berlin, Germany, 2013.
- [22] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Computer Vision—ECCV 2004*, vol. 3024 of *Lecture Notes in Computer Science*, pp. 25–36, Springer, Berlin, Germany, 2004.
- [23] O. Schwaneberg, U. Köckemann, H. Steiner, S. Sporrer, A. Kolb, and N. Jung, "Material classification through distance aware multispectral data fusion," *Measurement Science and Technology*, vol. 24, no. 4, Article ID 045001, 2013.
- [24] J. R. Janesick, *Scientific Charge-Coupled Devices*, vol. 83, SPIE Press, 2001.
- [25] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publications, 1997.
- [26] O. Schwaneberg, *Concept, system design, evaluation and safety requirements for a multispectral sensor [Ph.D. thesis]*, University of Siegen, 2013.

Review Article

Multiband Polarization Imaging

Yongqiang Zhao,¹ Qunnie Peng,¹ Chen Yi,¹ and Seong G. Kong²

¹School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

²Department of Computer Engineering, Sejong University, Seoul 05006, Republic of Korea

Correspondence should be addressed to Yongqiang Zhao; zhaoyq@nwpu.edu.cn and Seong G. Kong; skong@sejong.edu

Received 26 December 2014; Revised 8 April 2015; Accepted 15 April 2015

Academic Editor: Stefania Campopiano

Copyright © 2016 Yongqiang Zhao 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.

Multiband polarization imaging is an emerging sensing method that enables simultaneous acquisition of multiband spectral and multiangle polarization information of an object of interest in the scene. Spectral signatures of the light reflected from a target reveal the characteristics of the material composing the target while polarized light provides useful information on the surface features such as light scattering and specular reflection. In multiband spectral imaging, combined spectral and polarization information offers a comprehensive representation of an object utilizing complementary spectral and polarization information in visual sensing. Multiband polarization imaging has demonstrated a potential in the recognition of targets in challenging operating environments such as low-contrast and hazy conditions. This paper presents the concept and recent advances of multiband polarization imaging techniques, in particular, a bioinspired multiband polarization vision system. Applications of multiband polarization imaging in various fields include atmospheric observation, object detection and classification, medical diagnostics, surveillance, and 3D object reconstruction.

1. Introduction

An electromagnetic wave of light reflected from an object carries the signatures that characterize the object such as light intensity as a function of wavelength or spectrum, transmission directions, and the plane of polarization. Reflectance of absorption patterns of the spectrum describes material compositions of a target object, serving as a useful signature for target detection and classification. As a combined technique of spectroscopy and photography, spectral imaging observes spectral response of an object in different wavelengths at every location in an image plane. Spectral imaging acquires multiple images of a scene in different spectral bands. According to some criteria such as spectral resolution, number of spectral bands, and contiguosity of bands, spectral imaging techniques exist in several terms such as multispectral imaging and hyperspectral imaging, often in the visible and infrared regions of electromagnetic spectrum. Spectroscopic analysis has been used to identify the features inherent to the object of interest through measurement and analysis of electromagnetic spectra produced by reflection or absorption. Polarization is a property of transversal waves that can

oscillate with more than one orientation. Light can be approximated as a plane wave and propagates as a transverse wave, in which both the electric and magnetic fields are perpendicular to the direction of propagation. In linear polarization, the oscillation of these fields takes place in a single direction. A polarizer is an optical filter that transmits only one polarization direction. Imaging in different polarization angles has been employed to capture unique surface features of an object. The amount and orientation information of polarization have been of great significance to enhance discriminating power of vision-based detection and classification systems.

Recently there have been growing interests in multiband polarization imaging techniques to take advantage of both polarization and spectral signatures in various computer vision tasks. Multiband polarization imaging offers several advantages over conventional imaging techniques in sensing and analysis of objects in challenging environments. Multiband polarization imaging measures the intensity of the light reflected from the object in multiple spectral bands and multiple polarization angles to capture comprehensive optical characteristics of an object of interest in the scene. Combined spectral and polarization information

gives a complete representation of an object in terms of the material compositions as well as surface characteristics. In multiband polarization imaging, spatial, spectral, and polarization information are simultaneously acquired [1]. Spatial, spectral, and polarization information reveal the different characteristics of a material. It has been demonstrated that spatial-spectral information (spectral imagery) or spatial-polarization information (polarization imagery) provides improved classification accuracies [2–4]. Multiband polarization imaging provides an effective means to observe surface characteristics as well as material properties of an object in the scene, which may not be readily obtained using conventional imaging techniques. Polarization information finds surface properties of an object such as diffuse or specular reflection, scattering, and refractive index. Spectral signatures of an object measured in different spectral bands are used to characterize the types of materials.

Multiband polarization vision has gained significant interests in both biological and computer vision research communities [5] from the observations that some biological organisms in the nature demonstrate multiband polarization vision capabilities. Mantis shrimp and dragonfly are known to be able to easily detect and catch even transparent prey utilizing combined spectral and polarization information. They detect and recognize hidden or camouflaged objects or easily navigate through the water using the difference in polarization properties of the target and the background [2–4]. A compound eye of such organisms consists of tens of thousands of individual imaging units called ommatidia with effective imaging field-of-view of nearly 360 degrees. Each ommatidium responds to the light in different spectral bands as well as polarization angles. A unique structure of ommatidia in the eyes of mantis shrimp, squid, cuttlefish, and dragonfly is highly sensitive to polarized lights. In addition, the compound vision system demonstrates parallel processing capabilities that exist from retina to all ganglions [6, 7]. According to the studies on such organisms, a group of ommatidia are found in the dorsal rim area (DRA) of a compound eye. The DRAs of different organisms are known to have similar physiological characteristics. Every ommatidium contains two homochromatic photoreceptors that are orthogonal to each other. The photoreceptors in different ommatidia are sensitive to the light of different polarization angles [7–9]. Studies also indicate that the optic nerve systems of those organisms have particular polarization coding ability. From the research on retina in ommatidium of mantis shrimp and squid, the rhabdom and retinular cells have spectral perception ability from 300 nm to 700 nm with a perception bandwidth from 30 nm to 60 nm [10]. Mantis shrimp has 12 photoreceptors, each sampling a narrow set of wavelengths ranging from deep ultraviolet (300 nm) to far red (720 nm). Compared to three photoreceptors of the human visual system, those organisms can sense the light in a wider spectral region than humans do. Each type of photoreceptors is sensitive to a specific color. Mantis shrimp's color vision system is based on temporal signaling combined with scanning eye movements, enabling color recognition rather than discrimination without brain-power-heavy comparisons [11]. This scheme probably gives the

predatory shrimp a speed advantage in distinguishing prey with different color from cluttered background under changing light conditions [12]. In addition, the ommatidium in the hemisphere of the eye can identify the luminance, which acts as panchromatic imaging. Based on such physiological structures and functions, the organisms demonstrate multiband polarization vision in high resolution [13]. Multiband polarization vision system can be implemented based on the model of a cluster of ommatidia, where each ommatidium senses spectral or polarization information of the scene radiance in a particular spectral band or polarization angle.

Applications of multiband polarization imaging in object classification and clustering have demonstrated the effectiveness of multiband polarization imaging techniques. The variations of polarization parameters with wavebands are closely linked to the physiochemical characteristics of materials. Different inherent properties can enhance object discrimination and classification even when no obvious intensity difference exists. The measurement of Fresnel reflection coefficients in multiple bands can quantitatively assess the conductive characteristics to retrieve the dielectric constants, which provides good detectability of conductors and insulators. Roughness and surface orientation can be reflected in the spectral polarization parameters, which is critical for inhomogeneous objects identification [14]. Multiband polarization imaging has demonstrated enhanced target detection and navigation in military applications [15, 16]. Medical diagnosis with cytometry imaging and tissue assessment [17, 18] and the other applications including the measurement of aerosol density in the atmosphere [19, 20], geological exploration of glacier and vegetation distribution [21, 22], image dehazing [23–25], land cover classification [26], pathologic diagnosis of epithelial tissues [27], and visualization of an object hidden in the shadow [28] are application examples to name a few. Section 4 introduces some of popular application examples such as atmospheric observation, earth remote sensing, medical diagnosis, surveillance and reconnaissance, dehazing, and 3D reconstruction of specular objects.

2. Multiband Polarization Imaging Principles

Multiband polarization imaging techniques capture spectral and polarization information inherent to the properties of a target. We analyze multiband polarization image data to distinguish the differences in spectropolarimetric information among different objects. Spectral imaging attempts to measure intensity distribution of the light over different wavelengths while polarization imaging finds intensity of the image in different polarization directions. Figure 1 illustrates a hierarchy of multiband polarization imaging techniques. Two major steps to ensure precise multiband polarization imaging are spectral tuning and polarization adjustment. Multiband polarization imaging techniques obtain spectral information based on dispersion, channel tuning, or interference. Polarization information can be measured by rotating polarizer or micropolarization array. Images of different spectral bands and polarization angles are captured in sequence, which cannot satisfy time-sensitive imaging requirement.

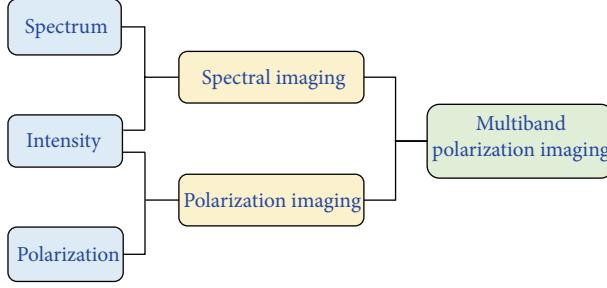


FIGURE 1: A hierarchy of multiband polarization imaging.

2.1. Spectral Tuning. Spectral imaging combines photography and spectroscopy to generate image data whose picture element (pixel) is associated with a spectral signature (spectrum). The spectral information provided by this pixel is valuable in the discrimination, detection, and classification of elements and structures within the image [29]. Each pixel in a spectral image is typically composed of narrow spectral bands of the electromagnetic spectrum. A spectral image I constitute a 3D data cube in two spatial coordinates (x, y) and one spectral dimension λ

$$I(x, y, \lambda_i), \quad i = 1, 2, \dots, L, \quad (1)$$

where L denotes the number of spectral bands. Spectral imaging has the ability to exploit multiple regions of the electromagnetic spectrum to probe and analyze the composition of a material. The materials comprising various objects in a scene reflect, absorb, and emit electromagnetic radiation in amounts that vary with the wavelength. If the radiation arriving at the sensor is measured over a spectral range, the resulting spectral signature can be used to uniquely characterize and identify any given material. Spectral feature of a pixel in a spectral image is compared to a database of known materials to determine the type of the material of the pixel.

In general, spectral components can be extracted using filters or devices with a function of spectral separation. Conventional spectral tuning methods include prism, mechanical filter wheel, diffraction grating, interference Fourier transform, and tomography [30, 31]. A flexible and programmable filtering elements such as acoustooptic tunable filter (AOTF) and liquid crystal tunable filter (LCTF) have been popular in spectral imaging. Multiband polarization imagers [2] collect spectral information using conventional methods such as diffraction grating, interference Fourier transform, and tunable filters and measure polarization information using optical polarizers. A multiband polarization imager using a single camera equipped with an LCTF and a polarizer wheel was proposed to capture simultaneously spectral and polarization information [21]. This imager obtains spectral information by sequentially tuning the center frequency of a bandpass characteristic of LCTF in different spectral bands and a set of component images of different polarization angles by rotating the polarizer in a filter wheel using a stepper motor. Such multiband polarization imaging techniques that capture spectral and polarization information

separately and individually in sequence tend to be time consuming and require complicated optical, mechanical, and electronic devices. A multiband polarization imager was built for navigation with a constant-gain omnidirectional mirror, an UV camera, and a color camera [32]. The goal was to measure the polarization pattern of the sky. The images in different polarization angles are captured in sequence, with large distortion due to the omnidirectional mirror. Such types of polarization imagers may not capture moving objects due to slow response time. In the beam splitting method, the camera can only measure one-third of the input energy with a difficulty to acquire multiband polarization information.

Compared with conventional optical imaging methods, spectral imaging techniques have access to spectral resolution, spatial resolution, radiative resolution, and time resolution of a target. The four resolutions enable reflecting the intrinsic spectral signatures of a target and distinguish the spectral differences among variable objects. Various spectral tuning methods being used in multiband polarization imaging systems are listed as Table 1.

2.2. Polarization Adjustment. Polarization is an important physical quantity describing the physicochemical properties of an object during the interaction of reflection, scattering, and transmission with solar radiation. Polarization imaging acquires multiple images in different polarization angles with abundant spectral information to improve the discrimination capability. Polarization imaging techniques have demonstrated such advantages as high signal-to-noise ratio and strong contrast over conventional optical imaging methods [33, 34]. Polarization imaging techniques have been successfully utilized in remote sensing and computer vision fields [34, 35].

Polarization features of an object are usually represented using the Jones vector, Stokes vector, and Muller matrix [56]. Strategies such as rotating polarizer, splitter, and division of FPA have been utilized to modulate polarization states [57]. Instantaneous acquisition of the Stokes parameters is of great importance in polarization imaging research. Multiband polarization imagery represents spatial, spectral, and polarization information of an object using four Stokes parameters describing the state of polarization in multiple bands. Suppose an object was images at four different polarization directions, 0, 45, 90, and 135 degrees at a certain wavelength λ . Let $I_0(x, y)$, $I_{45}(x, y)$, $I_{90}(x, y)$, and $I_{135}(x, y)$ denote the four

TABLE 1: Spectral tuning methods in multiband polarization imaging systems.

	Design features	Basic principles	Features
Dispersion	Prism interferometer	Light dispersion	Easy to realize Low spatial resolution Low spectral resolution
Interference	Interferometer	Fourier transforms Spectral pixel interferogram	Without calibration High imaging efficiency Low spatial resolution
Spectral filtering	Optical filters	AOTF: light diffraction LCTF: electrical tuning	Easy to realize Simple and compact Low imaging efficiency
Tomography	Tomography imager	Tomography projection	Complex structure Full-field High spectral resolution

image components measured at different polarization angles by rotating the polarizer to different orientations. Then the Stokes vector at wavelength λ is given by

$$\begin{aligned} S_\lambda(x, y) &= \begin{bmatrix} S_{0,\lambda}(x, y) \\ S_{1,\lambda}(x, y) \\ S_{2,\lambda}(x, y) \\ S_{3,\lambda}(x, y) \end{bmatrix} \\ &= \begin{bmatrix} I_{0,\lambda}(x, y) + I_{90,\lambda}(x, y) \\ I_{0,\lambda}(x, y) - I_{90,\lambda}(x, y) \\ I_{45,\lambda}(x, y) - I_{135,\lambda}(x, y) \\ I_{R,\lambda}(x, y) - I_{L,\lambda}(x, y) \end{bmatrix}, \end{aligned} \quad (2)$$

where $S_{0,\lambda}$ is the total intensity of light at wavelength λ , $S_{1,\lambda}$ denotes the difference between the light intensity at 0° and 90° , $S_{2,\lambda}$ represents the difference between 45° and 135° linear components, and $S_{3,\lambda}$ is the circular right to left polarization state. Circular polarization component $S_{3,\lambda}$ of a ground scene tends to be small, and therefore is often negligible [57]. When a linear polarization analysis is taken into account, the degree of linear polarization (DoLP) can be represented by a fraction of intensity attributed to the polarized light state as [34]

$$\text{DoLP}_\lambda(x, y) = \frac{\sqrt{S_{1,\lambda}^2(x, y) + S_{2,\lambda}^2(x, y)}}{S_{0,\lambda}(x, y)}. \quad (3)$$

Polarization angle (Orient) indicates the angle of major axis of polarization ellipse with respect to the reference direction (x -axis)

$$\text{Orient}_\lambda(x, y) = \frac{1}{2} \tan^{-1} \left(\frac{S_{2,\lambda}(x, y)}{S_{1,\lambda}(x, y)} \right). \quad (4)$$

Joint spatial, spectral, and polarization information can be represented by seven independent variables: spatial coordinates (x, y), wavelength (λ), and polarization angles (S_0, S_1, S_2, S_3). Based on the mathematic description of polarization, multiples images are required to characterize the polarization states of scenes. With the advance of photodetectors, there are a number of strategies to capture polarization signatures [21, 57]. Table 2 shows strategies of polarization adjustment.

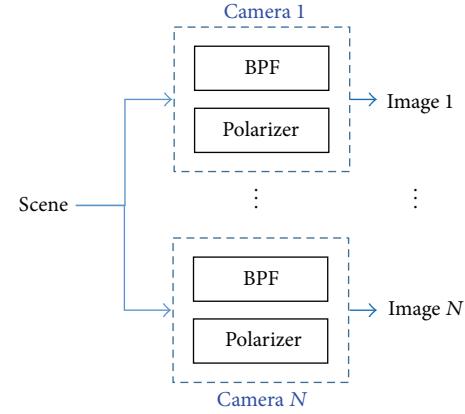


FIGURE 2: Acquisition of multiband spectral and polarization information of the scene using a camera array.

3. Multiband Polarization Imaging Systems

Inspired by the vision mechanism of some biological organisms, we developed a multiband polarization imaging system. The multiband polarization imager consists of a 3-by-3 array of CCD cameras, where each camera is equipped with either a color filter or an optical polarizer. An individual camera was set to capture an image in a certain spectral band or a polarization angle. This imager is to capture multiple images of the same scene in different spectral bands and polarization angles simultaneously for tasks such as target detection, so a large amount of geometrical distortion is not acceptable. Figure 2 shows a schematic diagram of the proposed multiband polarization imager for the acquisition of multiband spectral and polarization information from the scene. We use an optical polarizer to simulate the photoreceptor that is sensitive to the polarized light. The rhabdom and retinular cells having spectral sensing ability are simulated by a spectral bandpass filter to obtain multiband spectral image data. A 9-way gigabit ethernet card processes multiband spectral and polarization image data in parallel. This imager is configured to have five cameras with five color filters in the spectral bands of red, green, blue, yellow, and orange and four cameras with four optical polarizers of $0, 45, 90$, and 135 degrees.

TABLE 2: Polarization adjustment strategies.

	Design features	Integration issues	Imaging issues
Rotating polarizer	Simple and robust Time consuming Suitable for static scenes	Inexpensive Easy to implement	Misregistration Low efficiency
Beam splitter	Large size Light splitting Simultaneous acquisition	Expensive Complex integration High mechanical flexibility	Easy to align High efficiency Low contrast
Division of aperture	Small size Simultaneous acquisition	Expensive Complex optical elements	Easy to align Low spatial resolution
Division of FPA	Small size Compact Simultaneous acquisition	Expensive Complex fabrication High mechanical flexibility	High efficiency Imaging blur Low spatial resolution



FIGURE 3: Hardware configuration of a prototype multiband polarization imager. (a) Physical appearance of a 3-by-3 camera array and (b) arrangement of component cameras.

Figure 3 shows physical appearance of a prototype multi-band polarization imaging system and the configuration of component cameras. For the sake of simplicity in implementation, nine ($N = 9$) component CCD cameras are arranged in a 3-by-3 rectangular array rather than a hexagonal array as in the compound eye of an insect eye. Each camera is equipped with a color filter or a polarizer. Table 3 shows the arrangement of nine individual component cameras. To measure the polarization information with less noise or error, polarizers with four polarization angles of 0° , 45° , 90° , and 135° are used, mounted on the camera positions of $(1, 2)$, $(2, 1)$, $(2, 3)$, and $(3, 2)$. The spectral bandpass filters of red (600–700 nm), orange (590–610 nm), yellow (570–590 nm), green (490–570 nm), and blue (450–490 nm) measure the spectral information in five bands in the visible spectrum, mounted on the camera positions of $(1, 1)$, $(1, 3)$, $(2, 2)$, $(3, 1)$, and $(3, 3)$ respectively. An industrial CCD camera (Basler Ace 1300, 30 gm) of dimension $42 \text{ mm} \times 29 \text{ mm} \times 29 \text{ mm}$ ($W \times D \times H$) was used for each component camera. A large amount of image data coming from nine CCD cameras is processed and transmitted in parallel using 9-channel gigabit ethernet. An optical lens with the focal length of 16 mm, viewing angles of 38° (diagonal), 30.8° (horizontal), and 23.4° (vertical) was

TABLE 3: Placement of color filters and polarizers in a multiband polarization imager.

Camera number	Position	Polarization	Spectral band (nm)
1	(1, 1)	—	600–700 (red)
2	(1, 2)	0°	—
3	(1, 3)	—	590–610 (orange)
4	(2, 1)	45°	—
5	(2, 2)	—	570–590 (yellow)
6	(2, 3)	90°	—
7	(3, 1)	—	490–570 (green)
8	(3, 2)	135°	—
9	(3, 3)	—	450–490 (blue)

used in all the cameras. For a cooling purpose, CCD cameras are arranged with a gap of 22 mm.

Unlike sequential imaging process of conventional multi-band polarization imagers, the proposed imager can capture simultaneously a set of images of the scene in multiple spectral bands as well as different polarization angles. Spectral and polarization information is extracted using color filters and

TABLE 4: Application examples and mainstream sensors/techniques of multiband polarization imaging techniques.

Application examples	Mainstream sensors/techniques
Atmospheric observation	Electromechanical rotation Dispersion, Spectral filtering [19, 20, 36, 37] Multi-CCDs coordination [38]
Earth remote sensing	Interference, FPA-integrated CCDs Spectro-polarimetric retrieval [21, 22, 34]
Medical diagnosis	Multimodal sensors with tomography Endoscopy with LCTF-based CCDs [39–42]
Surveillance and reconnaissance	CCDs with polarizer/aperture Multi-data fusion [15, 43–47]
Image dehazing	Spectral filtering CCDs with polarizer/splitter [23, 48–51]
3D reconstruction	Single CCD/multi-CCDs with polarizer Multiview/binocular image fusion [52–55]

optical polarizers. Multiband spectral and polarization information can be measured in the common FOV region that is viewed by all the component cameras. Due to the different viewing angles of each component camera and the resulting mismatch in the FOV, the spectral and polarization information is not complete in nonoverlapping, boundary region. This will reduce the effective FOV of the multiband polarization imager. To expand the FOV, missing spectral and polarization information in the boundary region must be recovered. An attempt has been made to estimate the missing spectral and polarization information in the expanded FOV using the low-rank matrix recovery method [43, 44] that exploits the redundancy and correlation of the measured data.

4. Applications of Multiband Polarization Imaging Techniques

Multiband polarization imaging offers simultaneous acquisition of spectral and polarimetric signatures of an object for the detection of targets camouflaged or hidden in cluttered background. This technique has demonstrated improved object detection capabilities in a wide variety of applications in optical metrology that range from atmospheric science and remote sensing to 3D reconstruction. Table 4 lists some of popular application examples and the corresponding mainstream sensors and techniques used.

4.1. Atmospheric Observation. Multiband polarization imaging can be used to acquire spectral and polarization features of suspended atmospheric particles, which is useful in the correction of atmospheric distortions, climate investigation, and astronomical sensing [19]. Due to inherent polarization

effect caused by the atmospheric scattering and absorption of the light, the polarization states of aerosol particles in multiple spectral bands help to discriminate distribution, category, height, density, and size of suspended particles. Spectrum of metastable atomic oxygen in the upper atmosphere can be detected, which enables accurate measurements of the atmospheric properties, such as velocity, humidity, and temperature [36].

Early atmospheric exploration using multiband polarization imaging techniques starts in 1980s. The Polarization and Directionality of the Earth's Reflectances (POLDER) instrument has been used to obtain spectral and polarization information of aerosol to detect the atmospheric distribution, with the wavebands centered at 443 nm, 670 nm, and 865 nm [22]. Sano et al. [20] validated the accuracy of suspended particles distribution by extending the observation to six spectral bands. Airborne multiband polarimetric investigation [58] was conducted in the spectral range from 400 nm to 1000 nm. The polarization properties of aerosols with wavebands were analyzed. Li et al. [37] exploited ground-based multiband polarization data to retrieve the optical parameters of atmospheric aerosols, such as the optical density, size, polarization phase, and negative refractive index. Gartley et al. [59] explored multiband polarization detection by extending the wavebands from 200 nm to 250 nm, which enables achieving effective measurements of spectral aerosol absorption. Marbach et al. [38] incorporated multiview and multiangular schemes into multiband polarization imaging to resolve the directional anisotropy and the microphysical properties of aerosols. Redding et al. [60] developed an experimental multiband polarization apparatus. This setup extracts the polarization ratio and spectral features to identify the types and aggregations of aerosol particles. Distribution and density of aerosols between the two different weather conditions were successfully discriminated.

4.2. Earth Remote Sensing. Spectropolarimetric signatures of radiation energy reflected by ground targets have been applied in earth remote sensing. The joint spatial, spectral, and polarization information are more discriminative and accurate to describe the geometry and physicochemical properties of terrain targets, outperforming traditional sensing methods that capture intensity information only. The salient differences reflected in spectral polarization properties among various types of plants can be utilized to estimate the growth and the biomass of crops, plants species, and distribution and coverage area of forests [21, 22]. In earth remote sensing, exploiting spectropolarimetric differences reflected by various components like rocks and plants enables vegetation distributions. Microwave reflective and emissive characteristics of snow, water surface retrieved by the combination of spatial, spectral, and polarimetric detection, have potential applications in remote sensing, such as snowfall and rainfall parameters, sea salinity, wave height, marine pollution, oil spills, and coastline identification [61].

Recent research on red tide detection with spectral polarized radiance measurements highlights the performance of multiband polarization imaging techniques, which gives incomparable detection accuracy than conventional

detection methods. By investigating spectral variation and polarization effect of phytoplankton, the red tide species discrimination was implemented [62]. Multiband polarization imaging techniques are accessible for explorations of ice caves and glaciers [63]. Geological evolution can be recurred via detecting the polarized signatures of reflectance and fluorescence spectra. With the difference of spectral polarized information reflected by sand, soil, and water radiation, multiband polarization imaging techniques reveal significant information to observe landforms and geology. Mineral distribution and composition are approachable to be detected, with the inherent spectral polarized features. Maturity and dryness can be easily discriminated. Since multiband polarization imaging provides informative ground characteristics and thereby can be a useful indicator for evaluating the state of land desertification and erosion [64].

4.3. Medical Diagnosis. Pathological changes modify birefringence and structure of the tissue, which can be measured in terms of the polarization and spectral changes of scattered light. Multiband polarization imaging has been a powerful diagnostic tool to measure the features for quantitative pathology analysis. Absorption spectra in multiple wavebands are associated with molecular aggregations and provide information on the structure of biological tissues. Reflective index affected by the pathology of cell kernel and collagen can be retrieved by polarization analysis. Therefore, the compositional characteristics of organic tissues can be detected via spectral polarimetric diagnosis, such as the size and category of cell kernel and collagen [17, 18, 39, 65].

Widely applied in medical diagnosis, multiband polarization imaging strategy is an emerging technology for cytometry imaging, due to the operational flexibility of fast spectra collections and blood density estimation [40, 41]. Complementary to traditional X-ray-based diagnosis, multiband polarization detection optimizes the assessment of bone tissue. As a nondestructive measurement of biochemical properties, multiband polarization imaging is sensitive to local changes in mineral maturity regarding the spectral polarized signatures. Those attributes make multiband polarization imaging as a potential indicator for clinical diagnosis, like fracture risk and bone damage assessment [42]. The severity of natural caries lesions on occlusal surfaces can also be resolved using the joint spectra and polarization tomography. The integrated reflectivity is acquired to monitor the mineral loss and dental decay in the occlusal pit and fissures [66]. Multiband polarization imaging is effective to diagnose organic characteristics of skin pathology. Efforts have been applied in the skin melanin cancerous melanotic nevus detection and chilblain tissue diagnosis [67].

4.4. Surveillance and Reconnaissance. Multiband polarization imaging detection proves to have prominent performance in military strikes. The identification and tracking of missiles is accomplished using spectral polarization analysis of exhaust plume and fume. A strong polarization effect caused by fume in multiple bands is likely to expose the missile routes [45, 46]. Conventional photoelectric detection

often fails to distinguish man-made objects with low contrast. However, obvious differences in spectral and polarization features exist between military targets and natural objects. Multiband polarization imaging seeks to identify the geometric description of anomalous materials in cluttered background, even hidden or camouflage targets like tanks and armored vehicles under the trees [3, 4, 15, 43, 44, 47]. Multiband polarization imaging detection visually enhances the contrast of hidden targets compared to intensity-only images. With a high refractive index of the artifacts, spectral and polarization signatures from multiband polarization images fusion can be applied to ground reconnaissance and battle damage assessment [68]. On the other hand, multiband polarization detection impels the design of new type of camouflage coating materials [35]. By modulating the modality and roughness, the artifacts emit homothetic polarization spectra with natural background, which make it difficult to discriminate.

Outperforming conventional optical detections, multiband polarization imaging has particular merits in surveillance and reconnaissance. Polarization images in different spectral bands are exploited to retrieve the sea regime, which ensures sailing safety [58, 69]. Aircraft aviation often encounters difficulties in overcast conditions, due to the severe erosion and turbulence caused by the ice crystal in cirrus. Such atmospheric disturbance is disruptive to the precision of navigation. However, crystal aerosols are strongly polarized of incident radiation, and the joint spectropolarimetric signatures are flexible for the detection of cirrus aggregation, which guarantees aircraft navigation and guidance [70, 71].

4.5. Image Dehazing. Outdoor imaging in poor weather conditions such as haze and mist remains a challenging task in practice. Captured images are pervasively plagued with nontrivial degradations caused by atmospheric particles. The airlight is partially polarized and dominates in the measured radiation. Taking into account the polarization parameters of atmospheric scattering and the refinement of transmission in three visible bands, the haze effect can be effectively eliminated using multiband polarization dehazing methods [23, 48, 49]. Image details and color fidelity of the scene are remarkably improved, which is closely comparable to the original haze-free scene. Compared to prior-based image dehazing methods, multiband polarization techniques provide more accurate visual effects, due to the unique physics-based analysis of intrinsic properties of targets [50, 51]. Multiband polarization scheme has performed physics-based effectiveness in descattering and visibility enhancement in turbid media, even liquids and solids. This unified spectral polarization imaging has remarkable advantages in specific applications, such as underwater inspection, torpedo navigation, and ecology evaluation [72, 73]. Overlapping cast shadow removal can be implemented based on spectral polarization analysis [74, 75]. Unexpected highlight and flare suppression is efficiently actualized by multiband polarimetric observation.

4.6. 3D Reconstruction of Specular Objects. A three-dimensional (3D) reconstruction of less-textured objects with specular surfaces can be a challenging task. Due to the lack of

features, classical photometric 3D reconstruction algorithms may fail to acquire dense disparity since these methods rely on precise alignments, like binocular stereo vision. A 3D shape reconstruction with multiband polarization imaging has demonstrated effectiveness in the inspection of specular objects, disregarding the texture information [52, 76]. Reflection of unpolarized light becomes partially polarized according to the dielectric index of the surface. Using the Fresnel laws, the geometrical parameters such as zenith angle and azimuth angle can be retrieved from the degree of polarization and the angle of polarization, respectively. Therefore, the normal vector is integrated to reconstruct the depth information. Refractive index differences in multiple spectra can be employed to resolve the ambiguity on the zenith angle [53, 77]. Multiband polarization technique can be employed as an alternative 3D inspection method of specular objects [54, 55]. Specular reflection component can be eliminated using polarization signatures and spectra coherence constraints, while maintaining high visual quality [78].

5. Conclusion

Multiband polarization imaging is an emerging photoelectric detection technique with the ability to simultaneously acquire spectral and polarization information from an object of interest. This paper summarizes the principles and recent advances of multiband polarization imaging techniques. A multiband polarization imaging system implemented using a 2D camera array is introduced, which is inspired by a combined spectral and polarization vision mechanism of biological organisms. A number of application examples of multiband polarization imaging techniques are presented that include atmospheric observation, earth remote sensing, medical diagnosis, surveillance and reconnaissance, image dehazing, and 3D reconstruction of specular objects.

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 National Natural Science Foundation of China (61371152, 61071172, and 61374162), New Century Excellent Talents Award Program from Ministry of Education of China (NCET-12-0464), China Postdoctoral Science Foundation funded project (2013M532080), the Ministry of Education Scientific Research Foundation for Overseas Study, and the Fundamental Research Funds for the Central Universities (3102015ZY045).

References

- [1] B. M. Flusche, M. G. Gartley, and J. R. Schott, "Assessing the impact of spectral and polarimetric data fusion via simulation to support multimodal sensor system design requirements," *Journal of Applied Remote Sensing*, vol. 4, no. 1, Article ID 043562, 15 pages, 2010.
- [2] Y. Zhao, Q. Pan, C. Yi, and S. G. Kong, *Multiband Polarization Imaging and Applications*, Springer International Publishing, 2015.
- [3] Y. Zhao, L. Zhang, D. Zhang, and Q. Pan, "Object separation by polarimetric and spectral imagery fusion," *Computer Vision and Image Understanding*, vol. 113, no. 8, pp. 855–866, 2009.
- [4] Y.-Q. Zhao, P. Gong, and Q. Pan, "Object detection by spectropolarimetric imagery fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 3337–3345, 2008.
- [5] L. P. Lee and R. Szema, "Inspirations from biological optics for advanced photonic systems," *Science*, vol. 310, no. 5751, pp. 1148–1150, 2005.
- [6] T. W. Cronin, N. Shashar, R. L. Caldwell, J. Marshall, A. G. Cheroske, and T.-H. Chiou, "Polarization vision and its role in biological signaling," *Integrative and Comparative Biology*, vol. 43, no. 4, pp. 549–558, 2003.
- [7] T. W. Cronin and J. Marshall, "Patterns and properties of polarized light in air and water," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1565, pp. 619–626, 2011.
- [8] T.-H. Chiou, A. R. Place, R. L. Caldwell, N. J. Marshall, and T. W. Cronin, "A novel function for a carotenoid: astaxanthin used as a polarizer for visual signalling in a mantis shrimp," *Journal of Experimental Biology*, vol. 215, no. 4, pp. 584–589, 2012.
- [9] V. Pignatelli, S. E. Temple, T.-H. Chiou, N. W. Roberts, S. P. Collin, and N. J. Marshall, "Behavioural relevance of polarization sensitivity as a target detection mechanism in cephalopods and fishes," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1565, pp. 734–741, 2011.
- [10] J. Marshall and S. Kleinlogel, *Processing of Visual Information in Mantis Shrimps*, University of Queensland, Queensland, Australia, 2007.
- [11] J. Morrison, "Mantis shrimp's super colour vision debunked," *Nature News*, 2014.
- [12] H. H. Thoen, M. J. How, T.-H. Chiou, and J. Marshall, "A different form of color vision in mantis shrimp," *Science*, vol. 343, no. 6169, pp. 411–413, 2014.
- [13] T. W. Cronin, N. J. Marshall, and R. L. Caldwell, "Spectral tuning and the visual ecology of mantis shrimps," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 355, no. 1401, pp. 1263–1267, 2000.
- [14] Y. Zhao, G. Zhang, F. Jie, S. Gao, C. Chen, and Q. Pan, "Unsupervised classification of spectropolarimetric data by region-based evidence fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 755–759, 2011.
- [15] J. Duan, Q. Fu, C. Mo, Y. Zhu, and D. Liu, "Review of polarization imaging for international military application," in *International Symposium on Photoelectronic Detection and Imaging: Imaging Sensors and Applications*, vol. 8908 of *Proceedings of SPIE*, International Society for Optics and Photonics, Beijing, China, August 2013.
- [16] B. D. Bartlett and A. Schlamm, "Anomaly detection with varied ground sample distance utilizing spectropolarimetric imagery collected using a liquid crystal tunable filter," *Optical Engineering*, vol. 50, no. 8, Article ID 081207, 2011.
- [17] Z. Yongqiang, Z. Lei, and P. Quan, "Spectropolarimetric imaging for pathological analysis of skin," *Applied Optics*, vol. 48, no. 10, pp. D236–D246, 2009.
- [18] Y. Zhao, Q. Zhang, and J. Yang, "High-resolution multiband polarization epithelial tissue imaging method by sparse representation and fusion," *Applied Optics*, vol. 51, no. 4, pp. A27–A35, 2012.

- [19] K. Arai and Y. Nishimura, "Aerosol parameter estimation through atmospheric polarization observation with the different observation angles," in *Polarization Science and Remote Sensing IV*, vol. 7461 of *Proceedings of SPIE*, San Diego, Calif, USA, August 2009.
- [20] I. Sano, S. Mukai, and T. Takashima, "Multispectral polarization measurements of atmospheric aerosols," *Advances in Space Research*, vol. 19, no. 9, pp. 1379–1382, 1997.
- [21] K. Homma, H. Shingu, H. Yamamoto, H. Kurosaki, and M. Shibayama, "Application of an imaging spectropolarimeter to agro-environmental sciences," in *Sensors, Systems and Next-Generation Satellites VII*, vol. 5234 of *Proceedings of SPIE*, pp. 638–647, September 2003.
- [22] H. Guo and T. Yu, "A review of atmospheric aerosol research by using polarization remote sensing," *Spectroscopy and Spectral Analysis*, vol. 34, no. 7, pp. 1873–1880, 2014.
- [23] J. Mudge and M. Virgen, "Real time polarimetric dehazing," *Applied Optics*, vol. 52, no. 9, pp. 1932–1938, 2013.
- [24] S. Fang, X. S. Xia, X. Huo, and C. Chen, "Image dehazing using polarization effects of objects and airlight," *Optics Express*, vol. 22, no. 16, pp. 19523–19537, 2014.
- [25] T. Treibitz and Y. Y. Schechner, "Active polarization descattering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 385–399, 2009.
- [26] Y.-Q. Zhao, L. Zhang, and S. G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 747–756, 2011.
- [27] Y. Zhao, Q. Zhang, and J. Yang, "High-resolution multiband polarization epithelial tissue imaging method by sparse representation and fusion," *Applied Optics*, vol. 51, no. 4, pp. 27–35, 2012.
- [28] I. Ibrahim, P. W. Yuen, K. Hong et al., "Illumination invariance and shadow compensation via spectro-polarimetry technique," *Optical Engineering*, vol. 51, no. 10, Article ID 107004, 2012.
- [29] D. A. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, 2002.
- [30] X. Meng, J. Li, D. Liu, and R. Zhu, "Fourier transform imaging spectropolarimeter using simultaneous polarization modulation," *Optics Letters*, vol. 38, no. 5, pp. 778–780, 2013.
- [31] J. Li, B. Gao, C. Qi, J. Zhu, and X. Hou, "Tests of a compact static Fourier-transform imaging spectropolarimeter," *Optics Express*, vol. 22, no. 11, pp. 13014–13021, 2014.
- [32] N. Carey and W. Stürzl, "An insect-inspired omnidirectional vision system including UV-sensitivity and polarisation," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV '11)*, pp. 312–319, November 2011.
- [33] R. Azzam and N. Bashara, *Ellipsometry and Polarized Light*, North-Holland Publishing, Elsevier Science Publishing, 1987.
- [34] J. S. Tyo, D. L. Goldstein, D. B. Chenault, and J. A. Shaw, "Review of passive imaging polarimetry for remote sensing applications," *Applied Optics*, vol. 45, no. 22, pp. 5453–5469, 2006.
- [35] X. Sun, Y. Qiao, and J. Hong, "Review of polarization remote sensing techniques and applications in the visible and infrared," *Journal of Atmospheric and Environmental Optics*, vol. 3, article 5, 2010.
- [36] D. Tanré, F. M. Bräon, J. L. Deuzä et al., "Remote sensing of aerosols by using polarized, directional and spectral measurements within the A-Train: the PARASOL mission," *Atmospheric Measurement Techniques*, vol. 4, no. 7, pp. 1383–1395, 2011.
- [37] Z. Li, P. Goloub, C. Devaux et al., "Aerosol polarized phase function and single-scattering albedo retrieved from ground-based measurements," *Atmospheric Research*, vol. 71, no. 4, pp. 233–241, 2004.
- [38] T. Marbach, P. Phillips, A. Lacan, and P. Schlüssel, "The multi-viewing, -channel, -polarisation imager (3MI) of the EUMET-SAT polar system—second generation (EPS-SG) dedicated to aerosol characterisation," in *Sensors, Systems, and Next-Generation Satellites XVII*, vol. 8889 of *Proceedings of SPIE*, International Society for Optics and Photonics, October 2013.
- [39] L. Ying, L. Wen, and L. Cheng, "Research on the fluorescence polarization characteristic of alcoholism blood," *Acta Optica Sinica*, vol. 7, pp. 2065–2068, 2010.
- [40] J. A. Levitt, D. R. Matthews, S. M. Ameer-Beg, and K. Suhling, "Fluorescence lifetime and polarization-resolved imaging in cell biology," *Current Opinion in Biotechnology*, vol. 20, no. 1, pp. 28–36, 2009.
- [41] A. J. Makowski, C. A. Patil, A. Mahadevan-Jansen, and J. S. Nyman, "Polarization control of Raman spectroscopy optimizes the assessment of bone tissue," *Journal of Biomedical Optics*, vol. 18, no. 5, Article ID 055005, 2013.
- [42] D. Fried, P. Ngaotheppitak, C. Darling, and C. Ho, "Polarization sensitive optical coherence tomography for quantifying the severity of natural caries lesions on occlusal surfaces," in *13th Lasers in Dentistry*, vol. 6425 of *Proceedings of SPIE*, International Society for Optics and Photonics, San Jose, Calif, USA, March 2007.
- [43] W. Hubbard, G. Bishop, T. Gowen, D. Hayter, and G. Innes, "Multispectral-polarimetric sensing for detection of difficult targets," in *Electro-Optical and Infrared Systems: Technology and Applications V*, vol. 7113 of *Proceedings of SPIE*, International Society for Optics and Photonics, Wales, UK, October 2008.
- [44] Q. Wang, J. Wang, D. Zhao, L. Ma, Z. Chen, and Z. Li, "Recognition of camouflage targets with hyper-spectral polarization imaging system," in *International Symposium on Photoelectronic Detection and Imaging 2013: Imaging Spectrometer Technologies and Applications*, vol. 8910 of *Proceedings of SPIE*, International Society for Optics and Photonics, Beijing, China, August 2013.
- [45] S. Kaplan, G. Simon, S. Woods, A. Carter, and T. Jung, "Calibration of IR test chambers with the missile defense transfer radiometer," in *18th Technologies for Synthetic Environments: Hardware-in-the-Loop*, vol. 8707 of *Proceedings of SPIE*, International Society for Optics and Photonics, Baltimore, Md, USA, April–May 2013.
- [46] M. Flusche, G. Gartley, and R. Schott, "Exploiting spectral and polarimetric data fusion to enhance target detection performance," in *Imaging Spectrometry XV*, vol. 7812 of *Proceedings of the SPIE*, 78120C, p. 10, International Society for Optics and Photonics, August 2010.
- [47] B. D. Bartlett and A. Schlamm, "Anomaly detection with varied ground sample distance utilizing spectropolarimetric imagery collected using a liquid crystal tunable filter," *Optical Engineering*, vol. 50, no. 8, Article ID 081207, 9 pages, 2011.
- [48] E. Namer, S. Schwartz, and Y. Y. Schechner, "Skyless polarimetric calibration and visibility enhancement," *Optics Express*, vol. 17, no. 2, pp. 472–493, 2009.
- [49] S. Fang, X. Xia, X. Huo, and C. Chen, "Image dehazing using polarization effects of objects and airlight," *Optics Express*, vol. 22, no. 16, pp. 19523–19537, 2014.
- [50] A. El-Saba, M. S. Alam, and W. A. Sakla, "Pattern recognition via multispectral, hyperspectral, and polarization-based imaging," in *Automatic Target Recognition XX; Acquisition, Tracking*

- Pointing, and Laser Systems Technologies XXIV; and Optical Pattern Recognition XXI*, vol. 7696 of *Proceedings of SPIE*, April 2010.
- [51] C. Chen, Y.-Q. Zhao, D. Liu, L. Luo, and Q. Pan, "Robust materials classification based on multispectral polarimetric BRDF imagery," in *International Symposium on Photoelectronic Detection and Imaging 2009: Advances in Imaging Detectors and Applications*, vol. 7384 of *Proceedings of SPIE*, pp. 1–8, August 2009.
- [52] O. Morel, C. Stoltz, F. Meriaudeau, and P. Gorria, "Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging," *Applied Optics*, vol. 45, no. 17, pp. 4062–4068, 2006.
- [53] P. Boher, T. Leroux, T. Bignon, and V. Collomb-Patton, "Multispectral polarization viewing angle analysis of circular polarized stereoscopic 3D displays," in *21st Stereoscopic Displays and Applications*, vol. 7524 of *Proceedings of SPIE*, San Jose, Calif., USA, February 2010.
- [54] R. Rantson, C. Stoltz, D. Fofi, and F. Mériadeau, "3D reconstruction by polarimetric imaging method based on perspective model," in *Optical Measurement Systems for Industrial Inspection VI*, vol. 7389 of *Proceedings of SPIE*, p. 12, Munich, Germany, June 2009.
- [55] P. Garbat and J. Woźnicki, "Polarization imaging in 3D shape reconstruction," in *Image Processing and Communications Challenges 5*, pp. 205–211, Springer International Publishing, Berlin, Germany, 2014.
- [56] W. Shurcliff, *Polarized light. Production and Use*, Harvard University Press, Cambridge, Mass, USA, 1966.
- [57] L. B. Wolff, "Polarization vision: a new sensory approach to image understanding," *Image and Vision Computing*, vol. 15, no. 2, pp. 81–93, 1997.
- [58] Z. Xu, D. K. P. Yue, L. Shen, and K. J. Voss, "Patterns and statistics of in-water polarization under conditions of linear and nonlinear ocean surface waves," *Journal of Geophysical Research: Oceans*, vol. 116, no. 7, 2011.
- [59] M. G. Gartley, S. D. Brown, A. D. Goodenough, N. J. Sanders, and J. R. Schott, "Polarimetric scene modeling in the thermal infrared," in *Polarization Science and Remote Sensing III*, vol. 6682 of *Proceedings of SPIE*, International Society for Optics and Photonics, San Diego, Calif, USA, September 2007.
- [60] B. Redding, Y. Pan, C. Wang, G. Videen, and H. Cao, "Polarization resolved angular optical scattering of aerosol particles," in *Advanced Environmental, Chemical, and Biological Sensing Technologies XI*, vol. 9106 of *Proceedings of SPIE*, International Society for Optics and Photonics, May 2014.
- [61] J. Shi, "Estimation of snow water equivalence with two Ku-band dual polarization radar," in *Proceedings of the International Geoscience and Remote Sensing Symposium*, 2004.
- [62] L. Storrie, A. Hall, S. Hang et al., "Spectral profiling and imaging (SPI): extending L.I.F.E. technology for the remote exploration of life in ice caves (R.E.L.I.C.) on Earth and Mars," in *14th Instruments, Methods, and Missions for Astrobiology*, vol. 8152 of *Proceedings of SPIE*, International Society for Optics and Photonics, San Diego, Calif, USA, September 2011.
- [63] H. Shingu, H. Kurosaki, K. Homma, T. Suzuki, and H. Yamamoto, "Earth observation system incorporating an LCTF spectropolarimeter," in *6th Sensors, Systems, and Next-Generation Satellites*, vol. 4881 of *Proceedings of SPIE*, pp. 503–510, International Society for Optics and Photonics, Crete, Greece, April 2003.
- [64] R. Tawalbeh, D. Voelz, D. Glenar et al., "Infrared acousto-optic tunable filter point spectrometer for detection of organics on mineral surfaces," *Optical Engineering*, vol. 52, no. 6, Article ID 063604, 2013.
- [65] Y. Zhao, L. Zhang, and Q. Pan, "Spectropolarimetric imaging for anomaly epithelial tissue detection," in *Sequence and Genome Analysis: Methods and Applications (CreateSpace)*, pp. 297–330, 2011.
- [66] Y. Zhao, T. Yang, P. Wei, and Q. Pan, "Spectropolarimetric imaging for skin characteristics analysis," in *Medical Imaging and Informatics*, pp. 322–329, Springer, Berlin, Germany, 2008.
- [67] G. Jie and Y. He, "An experiment for high speed retrieval by ENVISAT-ASAR cross-polarized observations," in *Proceedings of the 32nd IEEE International Geoscience and Remote Sensing Symposium (IGARSS '12)*, pp. 2809–2812, Munich, Germany, July 2012.
- [68] Y. Zhao, G. Zhang, F. Jie, N. Liu, L. Song, and Q. Pan, "Damage assessment using multi-band polarimetric imaging," in *Proceedings of the 15th Chinese Conference on Image and Graphics*, 2010.
- [69] J. E. Vaillancourt and B. C. Matthews, "Submillimeter polarization of Galactic clouds: a comparison of 350 μm and 850 μm data," *The Astrophysical Journal, Supplement Series*, vol. 201, no. 2, article 13, 2012.
- [70] K. Knobelspiesse, B. van Diedenhoven, A. Marshak, S. Dunagan, B. Holben, and I. Slutsker, "Cloud thermodynamic phase detection with polarimetrically sensitive passive sky radiometers," *Atmospheric Measurement Techniques Discussions*, vol. 7, no. 12, pp. 11991–12036, 2014.
- [71] F. J. Crosby, "Stokes vector component versus elementary factor performance in a target detection algorithm," in *Polarization: Measurement, Analysis, and Remote Sensing VI*, vol. 5432 of *Proceedings of SPIE*, pp. 1–11, Orlando, Fla, USA, April 2004.
- [72] D. M. Kocak, F. R. Dalgleish, F. M. Caimi, and Y. Y. Schechner, "A focus on recent developments and trends in underwater imaging," *Marine Technology Society Journal*, vol. 42, no. 1, pp. 52–67, 2008.
- [73] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent surface modeling from a pair of polarization images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 73–82, 2004.
- [74] K. Dana, R. Friedhoff, B. Maxwell, S. Bushell, and C. Smith, "Polarization-based shadow detection," U.S. Patent 8,319,821, 2012.
- [75] S. K. Nayar, X.-S. Fang, and T. Boult, "Separation of reflection components using color and polarization," *International Journal of Computer Vision*, vol. 21, no. 3, pp. 163–186, 1997.
- [76] G. A. Atkinson and E. R. Hancock, "Recovery of surface orientation from diffuse polarization," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1653–1664, 2006.
- [77] S.-S. Lin, K. M. Yemelyanov, E. N. Pugh Jr., and N. Engheta, "Separation and contrast enhancement of overlapping cast shadow components using polarization," *Optics Express*, vol. 14, no. 16, pp. 7099–7108, 2006.
- [78] N. Kong, Y.-W. Tai, and J. S. Shin, "A physically-based approach to reflection separation: from physical modeling to constrained optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 209–221, 2014.

Research Article

Research on Multifeature Segmentation Method of Remote Sensing Images Based on Graph Theory

Wenxing Bao¹ and Xiuhong Yao²

¹School of Computer Science and Engineering, Beifang University of Nationalities, Yinchuan, Ningxia 750021, China

²Department of Information Engineering and Technology, Kashi Normal University, Kashi, Xinjiang 844006, China

Correspondence should be addressed to Wenxing Bao; bwx71@163.com

Received 17 November 2014; Revised 28 January 2015; Accepted 5 March 2015

Academic Editor: Chao-Cheng Wu

Copyright © 2016 W. Bao and X. Yao. 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.

According to the characteristics of high-resolution remote sensing (RS) images, a new multifeature segmentation method of high-resolution remote sensing images combining the spectrum, shape, and texture features based on graph theory is presented in the paper. Firstly, the quadtree segmentation method is used to partition the original image. Secondly, the spectrum, shape, and texture weight components are calculated all based on the constructed graph. The matching degree between pixels and the texture is computed similarity. Finally, the ratio cut standards combination of the spectrum, shape, and texture weight components is used for the final segmentation. The experimental results show that this method can obtain more ideal results and higher segmentation accuracy applied to RS image than those traditional methods.

1. Introduction

In the past decade, many scholars have done the research about graph-based approaches to image segmentation. Graph theory has got a lot of attention because of its representational power and flexibility properties. Bai et al. [1] apply graph cut (GC) theory to the classification of hyperspectral remote sensing images. Fuzzy SVM classifier and the GC-based classification were used in two-step classification strategy in this paper. Felzenszwalb and Huttenlocher [2] define a predicate for measuring the evidence for a boundary between two regions using a graph-based representation. The time complexity of the method runs in $O(m \log m)$ time for m graph edges. The work of Wang and Siskind [3] presents cut ratio as a new cost function of graph methods for segmenting image. This method is useful for some image segmentation applications. Cui and Zhang [4] use Minimum Span Tree optimal theory to realize object based on high-resolution image segmentation. The result proved that this method can obtain high quality segmentation. Kato et al. [5, 6] propose a Markov random field (MRF) image segmentation model based on the integration of colour and texture descriptors. This method can use both synthetic and natural color images.

Another early approach to image segmentation based on graph cut has been proposed [7–15].

The objectives of this paper are to obtain better image segmentation results and relative high segmentation accuracy for high-resolution RS images. For RS image, the algorithm's time complexity and space complexity rate will be high if only graph theory for image segmentation is used. Quarter-tree segmentation method is a fast image segmentation algorithm, but it cannot divide meaningful target area for RS images. If the threshold is set too low, oversegmentation phenomenon will be very serious. If the threshold is set too high, it cannot form a more accurate target edge [1, 13]. Therefore, this paper designs a new method by combining the merits of quarter-tree segmentation and ratio cut (R -cut) algorithm, and the method can be used in high-resolution RS images. This method is effective to reduce the size of the graph vertices, improving the accuracy of image segmentation. It first establishes the mapping relation of RS image and graph and then sets an energy function of the graph according to the remarkable weights. We can solve the energy function to get the minimum which will lead to the result of graph segmentation. At last we mapped the graph segmentation result back to image. Because the construction of graph and

the extraction of remarkable weights can be based on both pixel and image blocks, the methods based on graph theory will be good in image segmentation.

The organization of this paper is as follows. In the second section, quadtree segmentation method and R -cut theory for multifeature segmentation of RS image are described. Results and discussion are given in Section 3. The final section is the conclusions.

2. The R -Cut Theory for Multifeature Segmentation of RS Image

2.1. Quadtree Segmentation Method. A quadtree is a tree data structure in which each internal node has exactly four children. Quadtrees are most often used to zone a two-dimensional space by recursively subdividing it into four quadrants or regions [16]. Quadtree decomposition is currently a valuable method in image processing and computer graphics. The procedure of quadtree segmentation is as follows.

Step 1. The original image (typically $2N \times 2N$) is divided into four same size regions.

Step 2. It is to detect the constant gray level of each region segmentation image.

Step 3. If it cannot meet the request of constant gray level of the image, then each district will be divided into four areas of the same size and go to Step 2.

Step 4. If it meets the request, then stop the iterative process.

2.2. R -Cut Standards. The graph partition problem is defined on data represented in the form of a graph $G = (V, E)$, with V vertices and E edges. Where in form, G corresponds to the image, the vertices correspond to regions, and the edges correspond to adjacent relations between the regions. Ratio cut represents the ratio of the corresponding sums of two different weights of edges along the cut boundary. A minimum ratio cut refers to the smallest cut ratio [3].

The energy function of R -cut standards is as follows:

$$R\text{-cut}(A, B) \triangleq \frac{c_1(A, B)}{c_2(A, B)}, \quad (1)$$

where A and B represent two different image blocks, respectively, and $c_1(A, B)$ and $c_2(A, B)$ calculated the energy of cut sets by using two methods of weight calculation, respectively:

$$\begin{aligned} c_1(A, B) &\triangleq \sum_{u \in A, v \in B, (u, v) \in E} w_1(u, v), \\ c_2(A, B) &\triangleq \sum_{u \in A, v \in B, (u, v) \in E} w_2(u, v), \end{aligned} \quad (2)$$

where u denotes the vertex of A , v denotes the vertex of B , $w_1(u, v)$ and $w_2(u, v)$ denote the weights associated with each edge (u, v) , respectively, $w_1(u, v)$ is the first edge weight, and $w_2(u, v)$ is the second edge weight.

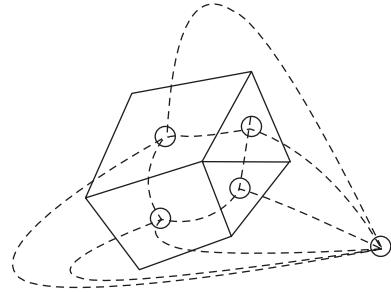


FIGURE 1: Dual graph.

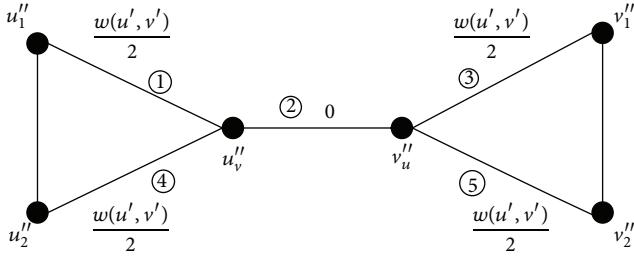
2.3. R -Cut Reduction Algorithm. The R -cut reduction algorithm steps are as follows.

Step 1 (calculate minimum ratio cut). In order to simplify the calculation, we calculate the minimum ratio ring instead of the minimum ratio. The dual graph $G' = (V', E')$ of the graph G is constructed and shown in Figure 1. There is a one-to-one correspondence between the minimum ratio cut of graph G and the minimum ratio ring of the dual graph G' , therefore, the problem that calculate the minimum ratio cut set $\{e_1, e_2, \dots, e_l\}$ of the original G can be transformed into the minimum ratio ring of the dual graph G' .

Step 2 (calculate the minimum ratio ring). In order to simplify the calculation, the negative simple ring is computed instead of the minimum ratio ring. The $G'(b)$ representations of the dual graph G' with the linear conversion w'_{11} , where w'_{11} is obtained by doing the conversion: $w'_{11} = w'_1 - bw'_2$ and functions w'_1 and w'_2 are the edge weights function of graph G' . The conversion of weights function will not change the minimum ratio ring loop of the dual graph G' , so it will not change the minimum ratio cut of the original G . Only the minimum loop cost of $G'(b^*)$ is nil consideration, and the graph G' has a minimum ratio of the loop which contains the loop ratio b^* . The graph G' has a minimum ratio of the loop which contains the loop ratio b^* , if and only if the minimum loop cost of $G'(b^*)$ is equal to zero. The relationship between b^* and b is as follows.

If $G'(b)$ has a negative cost loop, then $b^* < b$. While if $G'(b)$ does not have negative cost loop, then $b^* \geq b$. Let r_{\min} and r_{\max} , respectively, be the minimum and maximum loop ratio of d , so $b_{\min} = r_{\min}$ and $b_{\max} = r_{\max}$. Then, $b_{\min} \leq b^* \leq b_{\max}$. Let $b = b_{\min} + b_{\max}/2$; if $G'(b)$ has a negative cost loop, then b_{\max} value is set to b ; otherwise the b_{\min} value is set to b , continuing the repeated calculation, until we cannot find a negative cost of simple loop concerning the one corresponding to b ; and now b is the minimum loop ratio b^* and at the same time the negative cost simple loop of $G'(b)$ is the negative cost simple loop which we want to find.

Step 3 (calculate the minimum cost perfect matching). In order to reduce the calculation, the negative simple ring is computed instead of the minimum cost perfect matching. Construct a new graph $G'' = (V'', E'')$ from the graph G' which is obtained by the previous step. Graph G' contains

FIGURE 2: Transform to G'' .

a negative cost loop, if and only if G'' has the minimum cost perfect matching. From graph G' to graph G'' , the specific conversion principles are as follows.

- (i) For each vertex u'_1 of graph G' , graph G'' contains two vertices u''_1, u''_2 and an edge (u''_1, u''_2) of which weight value is equal to zero.
- (ii) For each edge (u', v') of graph G' , graph G'' contains two corresponding vertices u''_v and v''_u and five corresponding edges. Figure 2 shows the weight value of the five edges.

According to G'' which is obtained through the above three steps, we can calculate the $c_1(A, B)$ and $c_2(A, B)$ and obtain the minimum ratio cut according to formula (1).

2.4. Weight Calculation. In this paper, a multifeature segmentation method which takes into account the spectrum, shape, and texture features of RS image is applied.

The weight component based on the spectrum is defined as W_{ij}^{spectrum} , the weight component based on the shape is defined as W_{ij}^{shape} , the weight component based on the texture is defined as W_{ij}^{texture} , and W_{ij} is the combination of the above three aspects of information [8]:

$$W_{ij} = W_{ij}^{\text{spectrum}} \times W_{ij}^{\text{shape}} \times W_{ij}^{\text{texture}}. \quad (3)$$

The weight component based on the spectrum W_{ij}^{spectrum} is

$$W_{ij}^{\text{spectrum}} = \exp\left(\frac{-\chi^2(C_i, C_j)}{\sigma_{\text{color}}}\right). \quad (4)$$

In formula (4), σ_{color} denotes the standard deviation of the pixel color. χ^2 can be described as follows:

$$\chi^2(C_i, C_j) = \frac{1}{2} \sum_{k=1}^k \frac{[C_i(k) - C_j(k)]^2}{C_i(k) + C_j(k)}, \quad (5)$$

where k is the number of filter's types. If i and j correspond to two pixels, χ^2 is used to record the spectral similarity between the pixels i and j . If i and j are corresponding to the two blocks, χ^2 is used to record the spectral similarity between the blocks i and j .

The weight component based on the shape W_{ij}^{shape} is

$$W_{ij}^{\text{shape}} = 1 - \max_{x \in M_{ij}} p_{\text{con}}(x), \quad (6)$$

where $\max_{x \in M_{ij}} p_{\text{con}}(x)$ is used to mark the matching degree between the two pixels or blocks i and j , which is obtained by calculating the maximum value of all the probability coefficient along the line C of the set of pixels M_{ij} after connecting i and j in a straight line. If this line exactly intersects with a profile, then $\max_{x \in M_{ij}} p_{\text{con}}(x)$ is large, the weight is small, and i and j may belong to two classes; on the contrary, if the line is parallel to the profile, then $\max_{x \in M_{ij}} p_{\text{con}}(x)$ is small, the weight is greater, and i and j may belong to the same class.

The weight component based on the texture W_{ij}^{texture} is

$$W_{ij}^{\text{texture}} = \exp\left(\frac{-\chi^2(h_i, h_j)}{\sigma_{\text{texture}}}\right), \quad (7)$$

where h_i and h_j are the histogram obtained by doing texture operator transform for i and j , respectively, and σ_{texture} denotes the standard deviation of the texture of object. χ^2 is used to record the texture similarity between i and j . If the difference between h_i and h_j is too large, the values of χ^2 will be large, and W_{ij}^{texture} is very small. So i and j do not belong to the same class [1–4, 8].

The algorithm flowchart is shown in Figure 3.

3. Experimental Results and Analysis

3.1. Evaluation Method of Segmentation Results. Evaluation method of image segmentation is divided into qualitative and quantitative analysis.

This paper makes a quantitative analysis on the segmentation results by using the theory proposed in [14, 17–20]. The specific evaluation index is defined as follows.

3.1.1. Homogeneity. From [14], we handle the standard deviation of all the pixels as a measure of the object homogeneity criterion. The standard deviation of the object can be written as

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (C_i - \bar{C})^2}, \quad (8)$$

where n is the number of all pixels within the object, C_i represents the pixel gray value of pixel i , and \bar{C} represents the gray mean of the object.

3.1.2. Heterogeneity. For each object, we calculate the average difference absolute value of the object with the neighborhood to reflect the degree of difference between the object and the adjacent object [14]. The formula of heterogeneity can be written as

$$\Delta C = \frac{1}{l} \sum_{i=1}^n l_i |\bar{C} - \bar{C}_u|, \quad (9)$$

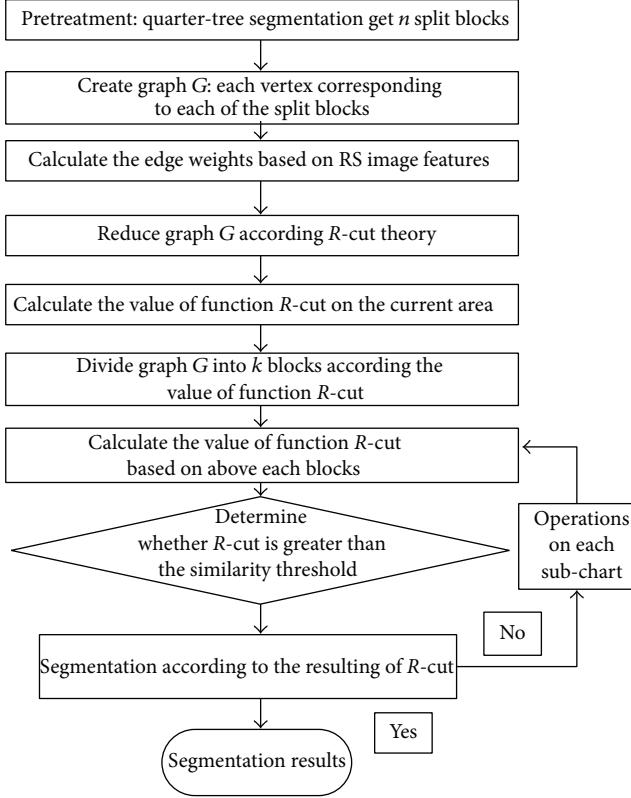


FIGURE 3: The algorithm flowchart.

where l is the boundary length of the current object, l_i is the common edge length of the current object with i adjacent objects, \bar{C} is the gray mean of the current object, \bar{C}_u is the gray mean of i adjacent objects, and n is the number of adjacent objects with the current object.

3.1.3. Segmentation Evaluation Index (SEI). SEI is inversely proportional with its homogeneity and proportional with its heterogeneity [14]. The SEI of the object is defined as follows:

$$SEL = \frac{\Delta C_l}{\sigma}. \quad (10)$$

3.1.4. Probabilistic Rand Index (PRI). The Probabilistic Rand Index (PRI) counts the similarity of pairs of pixels whose labels are consistent between the computed segmentation and the ground truth. The expression of PRI can be defined as

$$\begin{aligned} PR(S_{\text{test}}, \{G_k\}) \\ = \frac{1}{\binom{N}{2}} \sum_{i,j,i < j} [c_{ij}p_{ij} + (1 - c_{ij})(1 - p_{ij})]. \end{aligned} \quad (11)$$

This measure takes values in $[0, 1]$ -0 when two images have no similarities, and $c = 1$ when two images are identical, where S_{test} is the segmentation that is to be compared with the reference segmentation image and $\{G_k\}$ is ground-truth segmentations, where c_{ij} denotes the event of a pair of pixels i and j having the same label and p_{ij} its probability [17–19].

3.1.5. The Variation of Information (VoI). The Variation of Information (VoI) metric defines the distance between two segmentations as the average conditional entropy of one segmentation given the other one and thus roughly measures the amount of randomness in one segmentation which can be explained by the other [20]. The formula of VoI can be written as

$$d_{\text{VI}}(C, C') = H(C) + H(C') - 2I(C, C'), \quad (12)$$

where H and I , respectively, represent the entropies and the mutual information between two clustering of data C and data C' . This measure takes values in $[0, 1]$.

3.1.6. The Global Consistency Error (GCE). The Global Consistency Error (GCE) measures the extent to which one segmentation can be viewed as a refinement of the other. GCE can be defined as

$$\begin{aligned} GCE(S_1, S_2) \\ = \frac{1}{n} \min \left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\}. \end{aligned} \quad (13)$$

S_1 and S_2 are input segmentations images. $E(S_1, S_2, p_i)$ and $E(S_2, S_1, p_i)$ are the local refinement error, respectively. $E(S_1, S_2, p_i)$ is zero precisely when S_1 is a refinement of S_2 as pixel p_i , but not vice versa. [21].

3.2. The Experiment Results. In this section, we apply the proposed algorithms to real high-resolution data by the ALOS high-resolution RS images of Shi Zuishan Industrial Park, Ningxia, China. Its ground spatial resolution is 2.5 m and the size is 512×512 pixels. According to the human visual, field surveys and spectral measurement results, we select five samples, lime pile, cinder heap, house, road, and wasteland. To validate the algorithm, the images were segmented from the spectral, shape, texture, and multifeature segmentation based on graph theory, respectively, and then made a comparison among the four segmentation results. Original image and various algorithms segmentation results are shown in Figures 4 and 5, respectively. From Figure 4, the spectral segmentation based on graph theory has certain limitation; it is prone to split too small for textured areas, but it is less likely to split for the areas of relatively close texture. The shape segmentation based on graph theory is prone to split too small. Although the method of the texture segmentation based on graph theory can get a better segmentation of all types of surface features, it is not obvious to the boundary between the surface features. The multifeature segmentation method of remote sensing images based on graph theory not only can make the measurement, the spectra, and texture information of different objects better, reflect the differences between the different types of surface features, and achieve better segmentation, but also can accurately obtain the boundary between different types of objects; in short, it can ensure the accuracy of subsequent analysis.

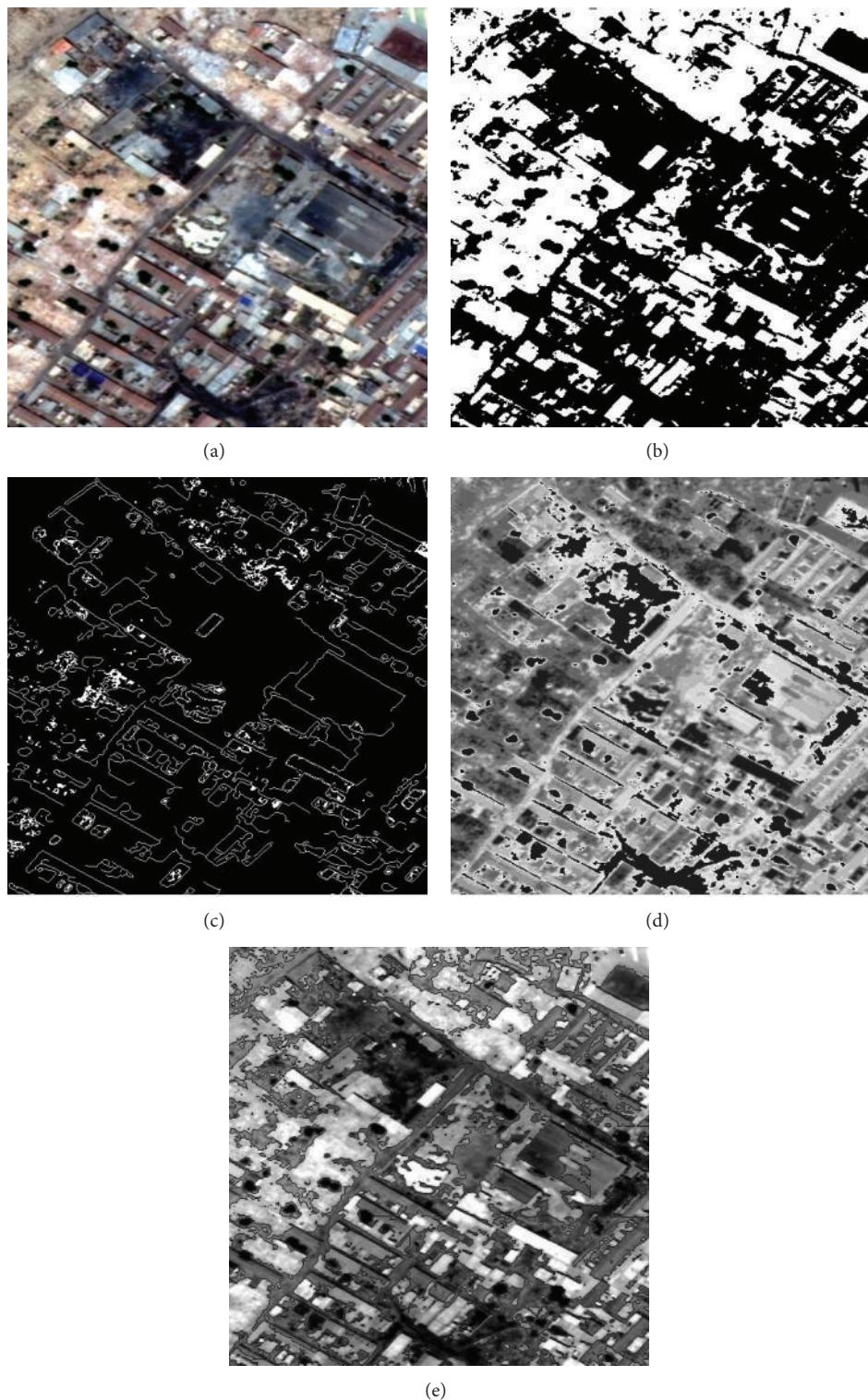


FIGURE 4: Results of various segmentation algorithms based on graph theory (RS image 1). (a) Original image 1. (b) Spectral segmentation. (c) Shape segmentation. (d) Texture segmentation. (e) Multifeature segmentation of proposed algorithms.

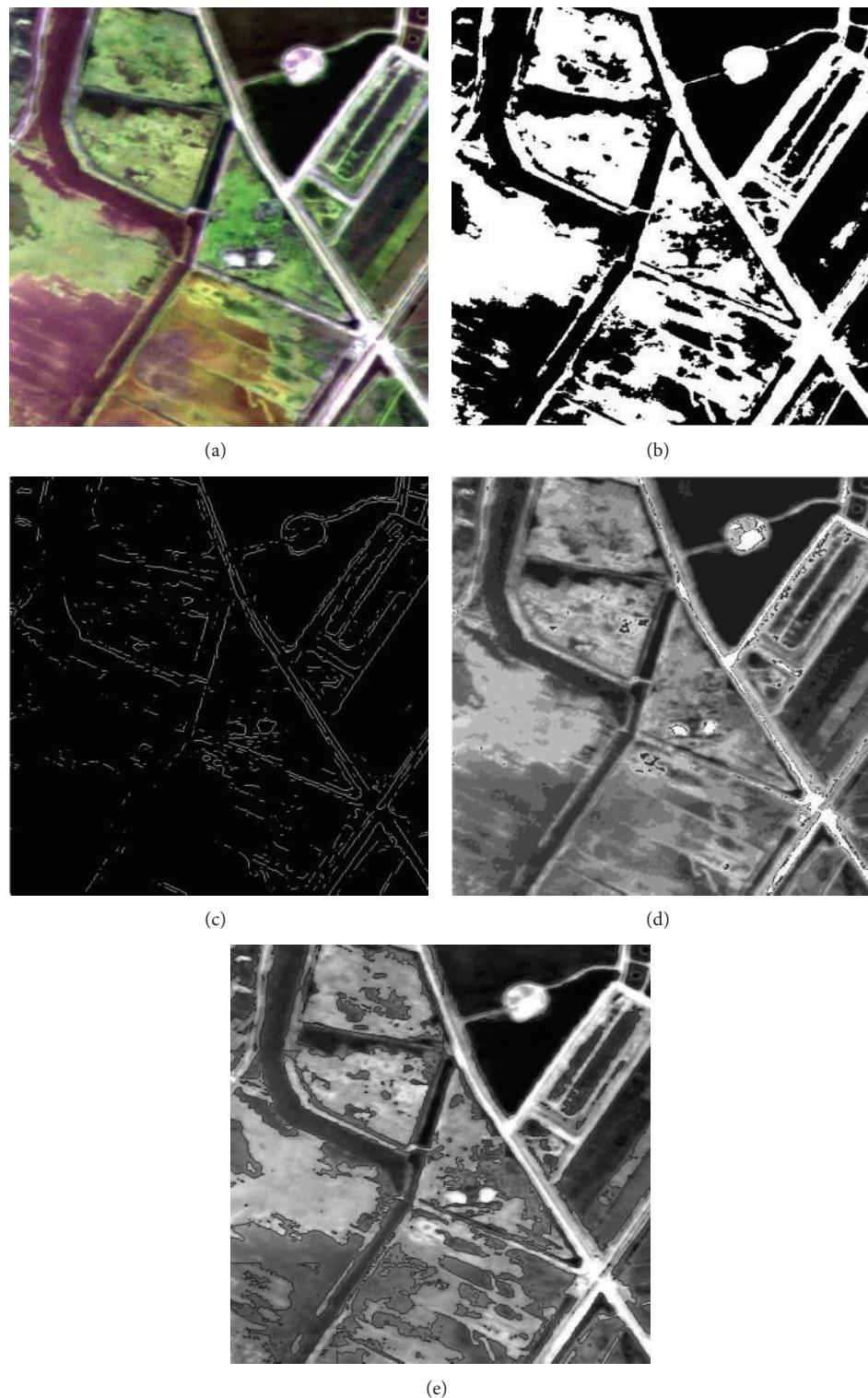


FIGURE 5: Results of various segmentation algorithms based on graph theory (RS image 2). (a) Original image 2. (b) Spectral segmentation. (c) Shape segmentation. (d) Texture segmentation. (e) Multifeature segmentation.

TABLE 1: Evaluation of segmentation scale (RS image 1).

Segmentation algorithm	Homogeneity (σ)		Heterogeneity (ΔC)		Segmentation evaluation index (SEI)	
	Lime heap	Cinder heap	Lime heap	Cinder heap	Lime heap	Cinder heap
Spectral	5.1580	0.1000	97.8235	22.8538	18.9654	228.5380
Shape	5.5125	0.9150	17.5005	3.6434	3.1747	3.9819
Texture	1.0646	4.0370	112.4847	50.4797	105.6591	12.5043
Multifeature	2.1045	0.9877	124.1282	52.5459	58.9822	53.2003

TABLE 2: Segmentation evaluation results over region-based algorithm and boundary-based algorithm.

Segmentation methods	Origin image 1			Origin image 2		
	PRI	GCE	VoI	PRI	GCE	VoI
Quadtree	0.8821	0.8737	12.1816	0.9564	0.9564	13.5365
Watershed	0.8797	0.8778	11.8219	0.9542	0.9568	13.2924
Mean shift	0.8818	0.8703	12.0964	0.9377	0.9607	12.9237
Multiresolution	0.8781	0.8741	11.9189	0.9543	0.9568	13.4735
Canny	0.2851	0.2201	5.4241	0.2030	0.1434	6.8397
Sobel	0.1543	0.0563	5.0602	0.0899	0.0427	6.6001
This paper	0.8815	0.8758	12.2234	0.9572	0.9573	13.7379

To be more accurate and objective evaluation of segmentation results of the algorithm, this paper uses the above-mentioned evaluation method for quantitative evaluation of the segmentation results. The lime heap and cinder heap are selected as evaluation object because the paper mainly monitors industry solid waste. Specific segmentation scale evaluation results of origin RS image 1 are shown in Table 1.

From Table 1, we can see that the lime heap homogeneity index is 2.1045 with the multifeature segmentation method of remote sensing images based on graph theory, which is smaller than the lime heap homogeneity index of spectral-based and shape-based segmentation, and the cinder heap homogeneity index is 0.9877, which is smaller than the cinder heap homogeneity index of texture-based segmentation and is close to the cinder heap homogeneity index of shape-based segmentation; this comparison and contrast herein prove that, by using the multifeature segmentation method based on graph theory, one can get a better measurement of the spectrum and texture information of surface features. The heterogeneity index of the multifeature segmentation based on graph theory is greater than the other three segmentation methods, and it further shows that the multifeature segmentation method based on graph theory can obtain more precise boundaries between different types of surface features. In summary, the results of the multifeature segmentation method based on graph theory make good internal object homogeneity, and at the same time there is an obvious contrast between adjacent objects.

The results obtained with the other segmentation methods and the proposed algorithm over two high-resolution RS images are shown in Table 2. The parameters of PRI, GCA, and VoI of each segmentation methods are computed. Quadtree, watershed, mean shift, multiresolution, and the proposed method are the region-based segmentation methods. The quadtree method starts at the root of the tree

that represents the whole image. If it is found nonuniform (not homogeneous), then it splits into four son squares (the splitting process). If, in contrast, four son squares are homogeneous, they are merged as several connected components (the merging process). This process continues recursively until no further splits or merges are possible. The multiresolution approach partitions the image at different scale, using a pyramid or quadtree structure. The watershed approach considers the gradient magnitude of an image as a topographic surface. Pixels having the highest gradient magnitude intensities correspond to watershed lines, which indicate the region boundaries. Mean shift method is defined as finding modes in a set of data samples, showing an underlying probability density function. Canny operator and sobel operator are boundary-based segmentation method.

From Table 2, taking into account the quality of the results from the evaluation parameters, it will be noticed that the best results are reached by the proposed method. The value of PRI of this method is the highest compared with the other segmentation methods. This is mainly due to the fact that this method combines the spectrum, shape, and texture of image and the segmentation region is close to real region. From the results of Table 2, we can see that the results of region-based segmentation methods are better than the boundary-based segmentation methods.

4. Summary

In this paper, we took into account a number of feature information of the image and used R -cut theory for RS images segmentation. Experimental comparison shows that multifeature segmentation method based on graph theory achieved better segmentation results than the methods based on single feature. Overall, the method can be used in high-resolution RS images. Even though the method also has

shortcomings, such as the effectiveness and the implementing speed of algorithms which are not very satisfactory, in the future, we will keep on seeking an efficient solving process and the weight calculation formula to apply to RS image segmentation.

Conflict of Interests

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

Acknowledgment

This work is supported by National Natural Science Foundation of China (61162013, 61461003).

References

- [1] J. Bai, S. Xiang, and C. Pan, "A graph-based classification method for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 803–817, 2013.
- [2] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [3] S. Wang and J. M. Siskind, "Image segmentation with ratio cut," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 675–690, 2003.
- [4] W. Cui and Y. Zhang, "Graph based multispectral high resolution image segmentation," in *Proceedings of the International Conference on Multimedia Technology (ICMT '10)*, pp. 1–5, October 2010.
- [5] Z. Kato and T.-C. Pong, "A Markov random field image segmentation model for color textured images," *Image and Vision Computing*, vol. 24, no. 10, pp. 1103–1114, 2006.
- [6] D. E. Ilea and P. F. Whelan, "Image segmentation based on the integration of colourtexture descriptors—a review," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2479–2501, 2011.
- [7] H. Wang, N. Ray, and H. Zhang, "Graph-cut optimization of the ratio of functions and its application to image segmentation," in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP '08)*, pp. 749–752, 2008.
- [8] Y. Tian, *Object oriented multi-scale segmentation of high resolution remote sensing image [M.S. thesis]*, Shanghai Jiaotong University, Shanghai, China, 2009.
- [9] J. Zhang, Z. Sun, and X. Yu, "Image segmentation based on graph theory algorithm simulation research," *Computer Simulation*, vol. 28, no. 12, pp. 268–270, 2011.
- [10] L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 879–893, 2012.
- [11] S. Sarkar and G. Healey, "Hyperspectral texture synthesis using histogram and power spectral density matching," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2261–2270, 2010.
- [12] S. Mei, M. He, Z. Wang, and D. Feng, "Spectral-spatial end-member extraction by singular value decomposition for AVIRIS data," in *Proceedings of the 4th IEEE Conference on Industrial Electronics and Applications (ICIEA '09)*, pp. 1472–1476, Xi'an, China, May 2009.
- [13] Q.-L. Tan, Z.-J. Liu, and W. Shen, "An algorithm for object-oriented multi-scale remote sensing image segmentation," *Journal of Beijing Jiaotong University*, vol. 31, no. 4, pp. 111–119, 2007.
- [14] C. Chen and G. Wu, "Evaluation of optimal segmentation scale with object oriented method in remote sensing," *Remote Sensing Technology and Application*, vol. 26, no. 1, pp. 96–102, 2011.
- [15] O. J. Morris, M. D. J. Lee, and A. G. Constantinides, "Graph theory for image analysis: an approach based on the shortest spanning tree," *IEE Proceedings F Communications, Radar and Signal Processing*, vol. 133, no. 2, pp. 146–152, 1986.
- [16] Y. Fisher, *Fractal Image Compression: Theory and Applications*, Springer, 1995.
- [17] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.
- [18] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops '05)*, p. 34, San Diego, Calif, USA, June 2005.
- [19] R. Unnikrishnan and M. Hebert, "Measures of similarity," in *Proceedings of the 7th IEEE Workshop on Applications of Computer Vision (WACV '05)*, pp. 394–400, January 2005.
- [20] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of the 8th International Conference on Computer Vision (ICCV '01)*, vol. 2, pp. 416–423, July 2001.
- [21] M. Meilă, "Comparing clusterings—an axiomatic view," in *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, pp. 577–584, August 2005.

Research Article

Development of a Gastric Cancer Diagnostic Support System with a Pattern Recognition Method Using a Hyperspectral Camera

Hiroyuki Oghara,¹ Yoshihiko Hamamoto,¹ Yusuke Fujita,¹ Atsushi Goto,² Jun Nishikawa,² and Isao Sakaida²

¹Department of Biomolecular Engineering, Yamaguchi University Graduate School of Medicine, Tokiwadai 2-16-1, Ube, Yamaguchi 755-8611, Japan

²Department of Gastroenterology and Hepatology, Yamaguchi University Graduate School of Medicine, Minami-Kogushi 1-1-1, Ube, Yamaguchi 755-8505, Japan

Correspondence should be addressed to Yoshihiko Hamamoto; hamamoto@yamaguchi-u.ac.jp

Received 24 December 2014; Revised 3 March 2015; Accepted 16 March 2015

Academic Editor: Chao-Cheng Wu

Copyright © 2016 Hiroyuki Oghara 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.

Gastric cancer is a completely curable cancer when it can be detected at its early stage. Thus, because early detection of gastric cancer is important, cancer screening by gastroscopy is performed. Recently, the hyperspectral camera (HSC), which can observe gastric cancer at a variety of wavelengths, has received attention as a gastroscope. HSC permits the discerning of the slight color variations of gastric cancer, and we considered its applicability to a gastric cancer diagnostic support system. In this paper, after correcting reflectance to absorb the individual variations in the reflectance of the HSC, a gastric cancer diagnostic support system was designed using the corrected reflectance. In system design, the problems of selecting the optimum wavelength and optimizing the cutoff value of a classifier are solved as a pattern recognition problem by the use of training samples alone. Using the hold-out method with 104 cases of gastric cancer as samples, design and evaluation of the system were independently repeated 30 times. After analyzing the performance in 30 trials, the mean sensitivity was 72.2% and the mean specificity was 98.8%. The results showed that the proposed system was effective in supporting gastric cancer screening.

1. Introduction

Gastric cancer is a completely curable cancer when it can be detected at its early stage. Thus, because early detection of gastric cancer is important, cancer screening by gastroscopy is performed. However, about 20% of gastric cancers are reportedly missed [1], and, in addition, detection greatly depends on the physician's proficiency. Accordingly, regardless of the physician's skill, the development of diagnostic support system that provides constant diagnostic accuracy is urgently needed.

Because there is wide variability of cancer, not just in gastric cancer, even the same carcinoma can differ from person to person. Therefore, a hyperspectral camera (HSC), which can observe gastric cancer at a variety of wavelengths,

has received attention [2]. The HSC has advanced the field of remote sensing. The images captured by the HSC contain 81 reflectance data values per 5 nm in the wavelength range from 400 to 800 nm for each pixel.

The HSC can discern the slight color differences of gastric cancer [3], and it is thought to be applicable to a gastric cancer diagnostic support system. However, because the color of the inner wall of the stomach varies from person to person, the individual differences in color should be absorbed to discriminate gastric cancer.

In this paper, a certain type of reflectance correction is performed to absorb the individual differences, and a system is designed using the corrected reflectance. The design of such a system involves two tasks. One is that of selecting the optimum wavelength. The HSC can observe gastric cancer

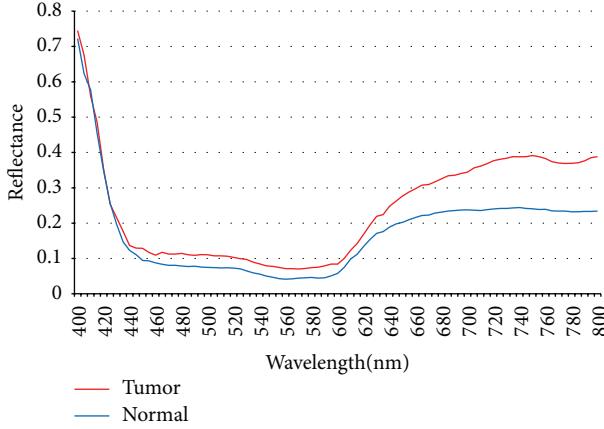


FIGURE 1: Relationship between wavelength and reflectance.

with various wavelengths, but, considering cost and real-time processing, the reflectance is obtained from the optimum wavelength alone. The other task is that of determining cutoff values to distinguish gastric cancer in classifier design. These problems are solved with a pattern recognition method using training samples alone. First, with regard to the selection of the optimum wavelength, using the feature selection method in which the Mahalanobis distance [4] is defined as the criterion function, an optimum wavelength is selected from the candidate wavelengths. Next, with regard to classifier design, a minimum-distance classifier [5] is modified and used. The efficacy of the designed system is evaluated by the hold-out method [5] using test samples independent of training samples.

2. Materials and Methods

Endoscopic resections were performed in 104 cases of gastric cancer at Yamaguchi University Medical School Hospital between April 2010 and August 2012 [6], and the gastric cancers were photographed by HSC immediately after the resections. Reflectance values were measured in pairs of normal and tumor tissues in each of the 104 cases treated by endoscopic submucosal dissection. Using one of the cases as an example, Figure 1 shows the relationship between wavelength and reflectance of the normal and tumor sites, respectively. In general, the reflectance of the tumor site is larger than that of the normal site, as well as when the wavelength is greater than 650 nm, and the difference in reflectance values between the normal site and the tumor site can be clearly seen. In system design, reflectance values at 51 wavelengths between 550 nm and 800 nm were used because there is heavy overlapping between normal and tumor sites in the wavelength range from 400 to 550 nm for almost all samples. The resolution of the HSC is 480×640 pixels.

Because the tissue type, shape, and color of gastric tumors vary, reflectance is not always uniform even in tumor sites. Thus, 10 points were obtained from the tumor regions. The 10 points were chosen so that they were uniformly dispersed throughout the tumor as much as possible. In the same

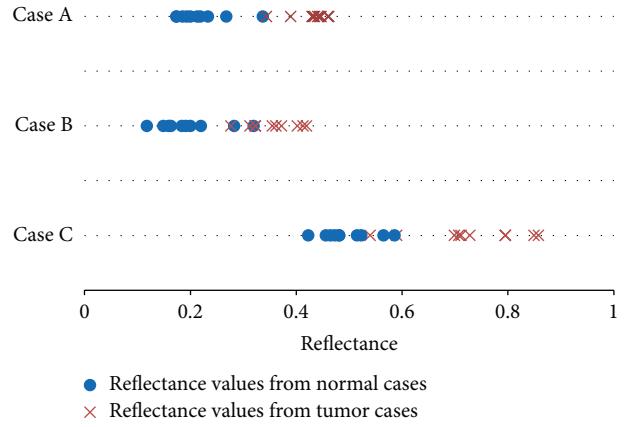


FIGURE 2: Examples of reflectance with individual differences.

manner, 11 points obtained from the normal sites were also chosen. One of the 11 points was used for reflectance correction, and, using the remaining points, the mean of normal sites was estimated. For details of the photographs, please see [2].

2.1. Design of the Proposed System. In pattern recognition, the samples used for system design are called training samples, and those for system evaluation are called test samples. Training samples and test samples must be different [7].

Figure 2 indicates the reflectance values from normal sites with • and those from tumor sites with ×, as examples. These examples show that there are individual differences in reflectance values, in which the values sometimes indicate normal cases and other times tumor cases. These results indicate that a single cutoff value for reflectance cannot distinguish between a normal site and a tumor site. Therefore, a corrected reflectance value is used. The mean reflectance value $x_{(\lambda)}^i$ is estimated from 10 reflectance values of a normal training sample corresponding to a case i at wavelength λ . Using the mean reflectance value $x_{(\lambda)}^i$, the corrected reflectance value $X_{(\lambda)}^i$ for $x_{(\lambda)}^i$ is described as follows:

$$\begin{aligned} X_{(\lambda)}^i &= x_{(\lambda)}^i - x_{(\lambda)}^i \\ &= 0. \end{aligned} \quad (1)$$

Meanwhile, the corrected reflectance value $Y_{(\lambda)}^i$ for the mean reflectance value $y_{(\lambda)}^i$, obtained from 10 reflectance values of a tumor training sample corresponding to the case i , is described as follows:

$$Y_{(\lambda)}^i = y_{(\lambda)}^i - x_{(\lambda)}^i. \quad (2)$$

From now on, we explain the design process using the corrected reflectance values, X and Y .

First, the optimum wavelength is found. For wavelength selection, the criterion of wavelength λ is determined. The Mahalanobis distance $D_{(\lambda)}$, which represents the statistical

distance between the normal and tumor distributions, is used as the criterion:

$$D_{(\lambda)} = \frac{(\mu_{X(\lambda)} - \mu_{Y(\lambda)})^2}{(1/2)(\sigma_{X(\lambda)}^2 + \sigma_{Y(\lambda)}^2)}, \quad (3)$$

where $\mu_{X(\lambda)}$ and $\sigma_{X(\lambda)}^2$ are the respective mean and variance of the corrected reflectance value of a normal site at wavelength λ and $\mu_{Y(\lambda)}$ and $\sigma_{Y(\lambda)}^2$ are the respective mean and variance of the corrected reflectance value of a tumor site. Equation (1), $X_{(\lambda)}^i = 0$, obtained by the correction leads to

$$\begin{aligned} \mu_{X(\lambda)} &= 0, \\ \sigma_{X(\lambda)}^2 &= 0. \end{aligned} \quad (4)$$

Consequently, $D_{(\lambda)}$ is as follows:

$$D_{(\lambda)} = 2 \left(\frac{\mu_{Y(\lambda)}}{\sigma_{Y(\lambda)}} \right)^2. \quad (5)$$

Because there are individual differences in reflectance, using reflectance values from the normal sites in each individual, the reflectance value of the normal site is corrected by (1) and then the reflectance value of the tumor is corrected by (2). Therefore, in all of the normal training samples, the value of corrected reflectance is zero. As a pattern recognition problem, normal and tumor sites are identified as a 2-class problem. The difference in the usual 2-class problem is that the distribution of the normal class is limited to the original point by correction. For this reason, at a glance, rather than being a 2-class problem, this appears to be a 1-class problem for the tumor site.

A wavelength that maximizes the value of $D_{(\lambda)}$ is selected from 51 candidate wavelengths and is defined as the optimum wavelength $\hat{\lambda}$. This is described as follows:

$$D_{(\hat{\lambda})} = \max_{\lambda} D_{(\lambda)}. \quad (6)$$

Second, we explain the design process of a classifier to discriminate between pixels at the normal site and the tumor site within the images. In this paper, the minimum-distance classifier, which is the simplest classifier, is modified and used. The minimum-distance classifier assigns a pattern x to the class associated with the nearest mean value of the two classes and is described as follows:

$$|x - \mu_{X(\hat{\lambda})}| < |x - \mu_{Y(\hat{\lambda})}| \rightarrow X \text{ is normal site}, \quad (7a)$$

$$|x - \mu_{X(\hat{\lambda})}| > |x - \mu_{Y(\hat{\lambda})}| \rightarrow X \text{ is tumor site}. \quad (7b)$$

Here, $|a|$ is the absolute value of a . The cutoff value of this minimum-distance classifier is the midpoint, $\mu_{Y(\hat{\lambda})}/2$, between the mean corrected reflectance value of a normal site, $\mu_{X(\hat{\lambda})}(= 0)$, and the mean corrected reflectance value of a tumor site, $\mu_{Y(\hat{\lambda})}$.

There is generally a trade-off relationship between sensitivity and specificity; that is, the higher the sensitivity,

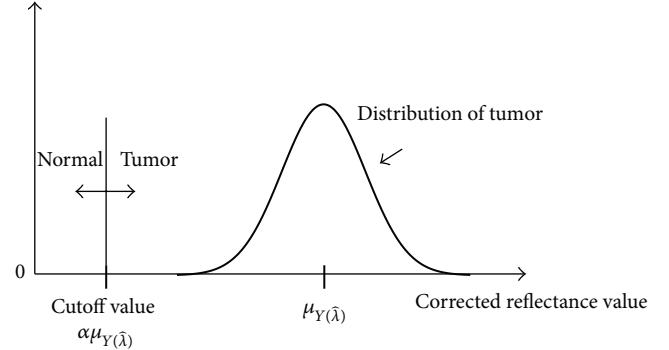


FIGURE 3: Modified minimum-distance classifier.

the lower the specificity and vice versa. Because the system is used for gastric cancer screening, a cutoff value that yields the maximum sensitivity to avoid missing cancer is expected while maintaining high specificity. A cutoff value is defined by multiplying parameters α and $\mu_{Y(\hat{\lambda})}$, that is, $\alpha\mu_{Y(\hat{\lambda})}$. Accordingly, optimization of a cutoff value represents optimization of parameter α . When especially $\alpha = 1/2$, this classifier becomes the minimum-distance classifier (see Figure 3).

In this study, to optimize the cutoff value parameter α , we took the approach of assigning the highest value of discrimination performance to be the optimum value by actually performing discrimination. Because optimization of parameter α was performed in the classifier design stage, parameter α should be optimized using training samples alone. Thus, training samples were further divided into subtraining samples and subtest samples, and, by using subtraining samples, a classifier was designed that discriminates the subtest samples. This process was applied to each predetermined candidate value of parameter α , and, among the candidate values, a value that met the conditions of discrimination ability was selected as the optimum value $\hat{\alpha}$. The idea of resampling the available samples as subsamples comes from the literature [8, 9]. The candidate values of parameter α were determined to be $\{1/2, 1/3, 1/4, 1/5\}$. Optimization of the parameter was conducted according to the following procedures. Procedure 5 is the condition of discrimination ability.

Procedure 1. Training samples are randomly divided into subtraining samples and subtest samples.

Procedure 2. Using subtraining samples, $\mu_{Y(\hat{\lambda})}$ is obtained.

Procedure 3. Cutoff values are obtained using $\alpha\mu_{Y(\hat{\lambda})}$ and $\alpha \in \{1/2, 1/3, 1/4, 1/5\}$.

Procedure 4. Using each cutoff value, subtest samples are discriminated.

Procedure 5. Regarding discrimination ability for the subtest samples, parameter α with a specificity of 99% or more and maximal sensitivity is selected as the optimum value $\hat{\alpha}$.

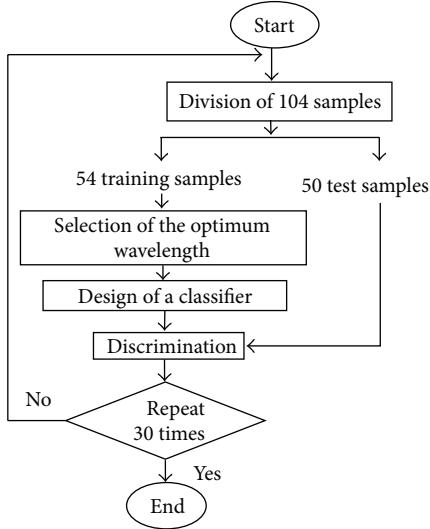


FIGURE 4: Flowchart of system design and assessment.

2.2. Validation of the Proposed System. The best method to evaluate the system is to assess the percentages of the sensitivity and specificity after the test samples are discriminated practically. In this paper, 104 cases were available as samples. The 104 samples were randomly divided into 54 training samples and 50 test samples, and, using the 54 training samples alone, a system was designed through which 50 test samples were discriminated. The trials described above were repeated independently 30 times, and the discrimination ability of the system was evaluated. The flowchart of the evaluation is shown in Figure 4. In addition, in parameter optimization of the cutoff values, the 54 training samples were randomly divided into 27 subtraining samples and 27 subtest samples.

In the test samples, at first, the mean value is calculated from 11 reflectance values of normal sites in the training samples, and the reflectance value that is closest to the mean value of the 11 reflectance values is selected from them as a typically normal reflectance value. The mean value of the remaining 10 reflectance values is calculated again and is defined as the mean reflectance value of normal sites. With respect to tumor sites, similar to the normal sites of the training samples, the mean value of 10 reflectance values of tumor sites is defined as the mean reflectance value of tumor sites. The corrected reflectance values of the normal and tumor sites are obtained by subtracting the typically normal reflectance values from the respective mean reflectance values.

3. Results and Discussion

Figure 5 shows the corrected reflectance values from the values in Figure 2. The reflectance values of the tumor sites are relatively larger in each case as compared with those of the normal sites. The wavelength and the value of parameter α are each acquired from 30 independent trials. With respect to the wavelength, 770 nm was chosen 11 of 30 times, and,

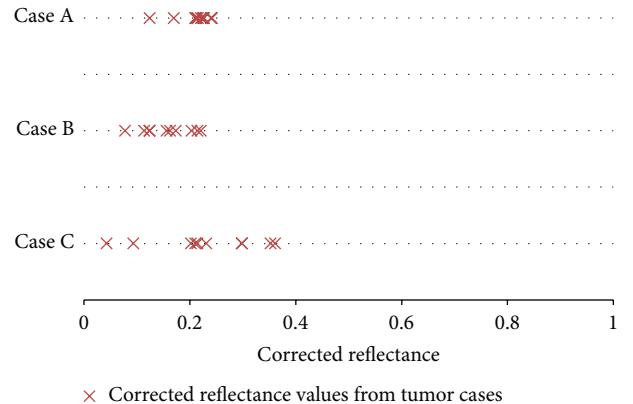


FIGURE 5: Examples of corrected reflectance values for training samples.

TABLE 1: $\hat{\lambda}$ and $\hat{\alpha}$ values with the highest frequencies of selection.

	First rank	Second rank	Third rank
Optimum wavelength $\hat{\lambda}$	770 nm	675 nm	680 nm
Optimum parameter $\hat{\alpha}$	1/4	1/5	1/3

TABLE 2: Discrimination ability of the system using 770 nm and 1/4 for test samples.

Rate of discrimination	Sensitivity (%)	Specificity (%)	Youden index
85.5	72.2	98.8	0.710
[84.6, 86.4]	[70.5, 73.9]	[98.4, 99.2]	[0.692, 0.728]

The numbers in the upper row indicate the mean values, and the numbers in the lower row indicate the 95% confidence interval.

for α , 1/4 was chosen 15 times. In this study, the most frequent wavelength was 770 nm, and 1/4 was the optimum solution. The second-rank wavelength of 675 nm was chosen 5 of 30 times, with 1/5 chosen 7 times. Also, the third-rank wavelength of 680 nm and 1/3 were chosen 3 and 4 times in 30, respectively. These are shown in Table 1.

The results of the discrimination with the use of 770 nm and 1/4 are shown in Table 2. Within the 30 discrimination trials, the mean sensitivity and specificity were 72.2% and 98.8%, respectively. Also, the Youden index [10], which is defined as sensitivity + specificity - 1, is shown for reference. High specificity is needed for gastric cancer screening, and sensitivity must also be high so that no cancer is missed. Although a sensitivity of 72.2% seems to be low, the discrimination in this study is conducted for each pixel, and, thus, despite the low sensitivity of each discrimination, pixels that are discerned as cancer congregate to form one region within the image, leading to avoidance of missing cancer by the physician.

The point of this study can be found in the correction of individual differences of reflectance. Therefore, to clarify the effects of this correction, we conducted an experiment

TABLE 3: Discrimination ability with and without correction.

	Optimum wavelength	Optimum parameter	Rate of discrimination (%)	Sensitivity (%)	Specificity (%)	Youden index
With correction	770 nm	1/4	85.5	72.2	98.8	0.710
Without correction	780 nm	1/5	67.0	73.8	60.2	0.340

TABLE 4: Effect of an increase in the number of features.

Wavelength	Optimum parameter	Rate of discrimination (%)	Sensitivity (%)	Specificity (%)	Youden index
770 nm	1/4	85.5	72.2	98.8	0.710
770 nm, 675 nm	1/4	86.4	74.9	97.9	0.729

comparing techniques that perform correction with techniques that do not perform correction. In both techniques, feature selection was performed using the Mahalanobis distance, and the modified minimum-distance classifier was used. As shown in Table 3, when comparing with and without correction, discrimination ability was improved with correction, and, accordingly, the validity of the correction was revealed.

We used only one wavelength for discrimination in our original study. Therefore, we conducted an experiment to determine whether discrimination performance could be improved by increasing the number of wavelengths. Specifically, the discrimination experiment was conducted on two-dimensional feature space by combining the optimum wavelength of 770 nm and the second-rank wavelength of 675 nm. The results are shown in Table 4. Even when the number of features was increased, the discrimination performance remained virtually unchanged. By using the value of the Youden index, hypothesis tests were conducted of the differences in average value with respect to the performance of 770 nm independently and the combined performance of 770 nm and 675 nm, and there was no significant difference in the results of either test ($P = 0.13$).

Real-time processing was required in this study. Not only does increasing the number of features lengthen the processing time, but also the hardware scale increases. Since no clear improvement in discrimination performance is obtained by increasing the number of features, by emphasizing the real-time processing, we have adopted a one-dimensional system by using one wavelength. For practical use, there may be a problem with the photographic speed of the HSC, which photographs at many wavelengths. However, since the system uses a single optimum wavelength, the photographic speed of the HSC does not matter, and thus real-time processing is achievable.

Finally, when this system is applied in practical use, correction will be a problem. Correction requires normal samples, but such samples are not actually available. Therefore, in this paper, we hypothesize that when the camera photographs the inside of the stomach, almost all of the pixels within the image will be normal pixels. If the image contains many pixels of cancer, a doctor can easily detect cancer without the support of the system. In general, this hypothesis is considered to be formed for the images used for gastric

cancer screening. If this hypothesis is satisfied, one pixel is randomly selected within the images and can be used as a normal pixel for correction.

This research depends on the data that is acquired. This means that the individual wavelength and cutoff value should be optimized depending on the hyperspectral camera that is used. Therefore, the values of 770 nm and 1/4 might not be valid when another hyperspectral camera is used. However, as revealed in this study, to resolve the problem of individual differences in patients, the value of this study is in establishing an approach whereby real-time processing is possible.

4. Conclusion

In this paper, we developed a diagnostic support system for gastric cancer that could discern between a pixel of a normal site and a pixel of a tumor site for each pixel in the images of 104 gastric cancer cases photographed by a HSC. Based on the results of 30 independent trials with the optimal wavelength 770 nm and cutoff value of 1/4, it was shown that this system is effective in screening for gastric cancer, achieving an average sensitivity of 72.2% and average specificity of 98.8%.

From the standpoint of this study, whether a lesion is gastric cancer is ultimately determined by the physician, and the system supports the physician to avoid missing gastric cancer. For this purpose, the system can discriminate on a pixel-by-pixel basis and support a physician's interpretation with a color display of the regions consisting of pixels discriminated as a tumor in the images.

The data used here are from images of tissues photographed by the HSC immediately after gastric cancer resection. In the future, we aim to use the system in the clinical setting, and we are planning to perform experiments using images photographed from within the stomach.

Conflict of Interests

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

References

- [1] K. Aida, H. Yoshikawa, C. Mochizuki et al., "Clinicopathological features of gastric cancer detected by endoscopy as part

- of annual health checkup,” *Journal of Gastroenterology and Hepatology*, vol. 23, no. 4, pp. 632–637, 2008.
- [2] S. Kiyotoki, J. Nishikawa, T. Okamoto et al., “New method for detection of gastric cancer by hyperspectral imaging: a pilot study,” *Journal of Biomedical Optics*, vol. 18, no. 2, Article ID 026010, 2013.
 - [3] S. Satori, Y. Aoyanagi, U. Hara, R. Mitsuhashi, and Y. Takeuchi, “Hyperspectral sensor HSC3000 for nano-satellite TAIKI,” in *2nd Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications*, vol. 7149 of *Proceedings of SPIE*, p. 71490M, Noumea, New Caledonia, November 2008.
 - [4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2nd edition, 2001.
 - [5] K. Fukunaga, *Introduction to statistical pattern recognition*, Computer Science and Scientific Computing, Academic Press, Boston, Mass, USA, 2nd edition, 1990.
 - [6] A. Goto, J. Nishikawa, S. Kiyotoki et al., “Use of hyperspectral imaging technology to develop a diagnostic support system for gastric cancer,” *Journal of Biomedical Optics*, vol. 20, no. 1, Article ID 016017, 2015.
 - [7] S. J. Raudys and A. K. Jain, “Small sample size effects in statistical pattern recognition: Recommendations for practitioners,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 3, pp. 252–264, 1991.
 - [8] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, 1993.
 - [9] Y. Hamamoto, S. Uchimura, and S. Tomita, “A bootstrap technique for nearest neighbor classifier design,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 1, pp. 73–79, 1997.
 - [10] W. J. Youden, “Index for rating diagnostic tests,” *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

Research Article

A New Methodology for Spectral-Spatial Classification of Hyperspectral Images

Zelang Miao and Wenzhong Shi

Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Correspondence should be addressed to Wenzhong Shi; johnshiforscholar@gmail.com

Received 29 October 2014; Revised 27 January 2015; Accepted 6 February 2015

Academic Editor: Chao-Cheng Wu

Copyright © 2016 Z. Miao and W. Shi. 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.

Recent developments in hyperspectral images have heightened the need for advanced classification methods. To reach this goal, this paper proposed an improved spectral-spatial method for hyperspectral image classification. The proposed method mainly consists of three steps. First, four band selection strategies are proposed to utilize the statistical region merging (SRM) method to segment the hyperspectral image. The segmentation map is subsequently integrated with the pixel-wise classification method to classify the hyperspectral image. Finally, the final classification result is obtained using the decision fusion rule. Validation tests are performed to evaluate the performance of the proposed approach, and the results indicate that the new proposed approach outperforms the state-of-the-art methods.

1. Introduction

Hyperspectral images are generally composed of hundreds to thousands of spectral bands. This rich spectral information can effectively distinguish different objects and physical materials and thus cause broad applications in the mineral detection, environment monitoring, and precision agriculture. The classification technology is currently the predominate method for analyzing hyperspectral images and has received much attention. Over the past decades, numerous pixel-wise classification methods, which only use spectral information, have been proposed to classify remote sensing images. In reviewing the literature, pixel-wise classification methods mainly include maximum-likelihood [1], spectral angle classifier [1, 2], neural networks [1], genetic algorithms [3, 4], decision tree [1], and kernel-based methods [5–7]. Particularly, support vector machine (SVM) provides higher classification accuracy in most cases [5, 8, 9]. In this field, Lu and Weng [10] presented a good review and analysis of classification methods and techniques for remote sensing images.

Although pixel-wise classification methods have been researched for years, the spatial information has still not been

sufficiently investigated. Generally, the spatial information is important for classification accuracy that can cause decrease of the classifier performance if neglected, particularly for very high spatial resolution satellite images. Previous studies show that pixel-wise methods will sometimes produce classification maps that look noisy (also known as “salt and pepper” effects) if the image spatial information is not used [10, 11]. An alternative to current pixel-wise methods is to combine spatial information with spectral information in classifier systems to form so-called spectral-spatial classification methods. This approach simultaneously considers the spectral and spatial information and it can generally achieve higher classification accuracy than pixel-wise classifiers [11–16]. Usually, there are two means to implement spectral-spatial classification. One of the spectral-spatial classification methods incorporates spatial information from its neighborhood, using a fixed size window. Mathematical morphology is used to construct morphological profiles (MPs) [11, 12] to define spatial information. Another approach is the integration of contextual/textural information in the kernel methods proposed by Camps-Valls and Bruzzone [17]. However, such spectral-spatial methods have the problem of scale selection, especially when the remote sensing images consist of

complicated structures [14], such as roads, buildings, and other man-made objects.

Another spectral-spatial scheme includes the postclassification spatial information using a segmentation map. An approach for classifying high spatial resolution urban satellite imagery is based on the different segmentation results of various scales [18]. The partitional clustering method [14] and the watershed transformation [15] have also been used to measure spatial structures. The advantage of aforementioned methods is that it is easy to perform with low computational complexity. It should be pointed out that, however, the partitional clustering suffers from the following two limitations: (1) the cluster numbers have to be set by users in advance; (2) the clustering result depends on the initialization values and thus the clustering result is unrobust. These drawbacks raise the problem that it is unclear if the partitional clustering can improve the pixel-wise classification accuracy.

Based on the aforementioned analysis, this study presents a new spectral-spatial classification approach for hyperspectral images. The spatial information is obtained from the statistical region merging (SRM) [19], not the partitional clustering technique. The method is easy to implement and the experimental results are presented in the latter part of this paper showing not only that the method can improve classification accuracy but also that the results are robust.

The main contributions of this paper are two-fold:

- (1) proposing a strategy for band selection from the hyperspectral image;
- (2) proposing a method for spectral-spatial classification using SRM based on the designed band selection strategy.

The remainder of this paper is organized as follows. The spectral-spatial classification using grouping clustering is introduced in Section 2, while the proposed method is validated on two experimental images in Section 3. Section 4 includes the conclusions and future work.

2. Methodology

The proposed spectral-spatial classification combines advances in SVM classification and SRM segmentation methods. The proposed method has three main steps, as summarized in Figure 1. Details of each step are introduced as follows.

2.1. Hyperspectral Image Segmentation. To segment the hyperspectral image, statistical region merging (SRM) [19] was selected to achieve the purpose. The advantage of this algorithm is that it can segment an image into regions in a fast and robust manner. In the SRM, let the observed image be denoted as I which contains $|I|$ pixels. The image I contains $\{R, G, B\}$ color channel values belonging to the set $\{0, 1, 2, \dots, g\}$ (where $g = 255$). Let the perfect scene of an observed image I be denoted as I^* . The observed color channel is sampled from a family of Q , taken on values from $[0, g/Q]$ and distributions at each pixel of I^* . Q controls the number of regions: the higher Q is, the greater

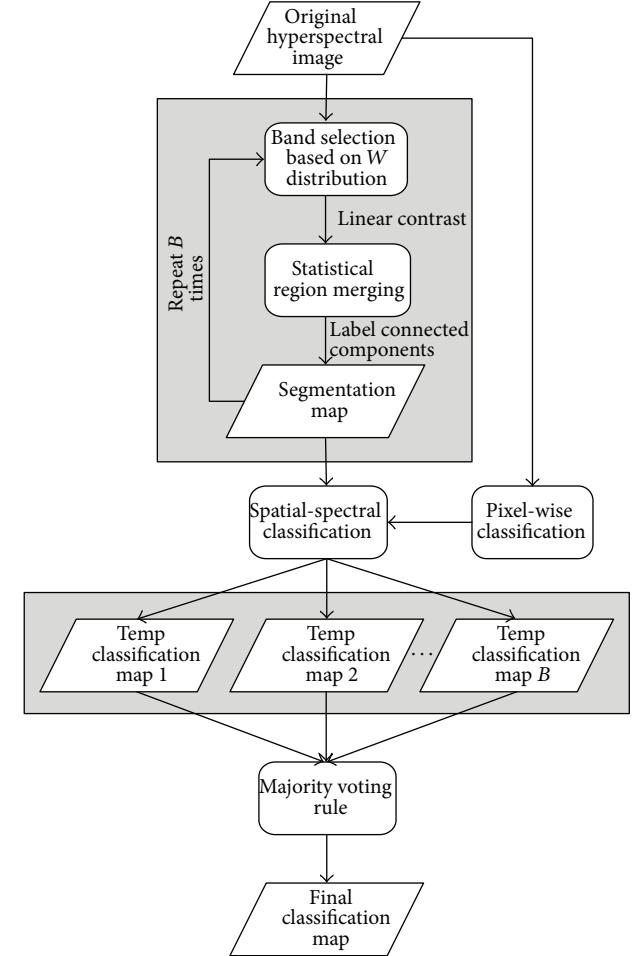


FIGURE 1: Flowchart of the proposed method.

the number of regions generated is. Two components are essential in defining the SRM algorithm: the merging predicate and the merging order. The merging predicate is defined as

$$P(R, R') = \begin{cases} \text{true}, & \text{if } \forall a \in \{R, G, B\}, |\bar{R} - \bar{R}'| \leq b(R, R') \\ \text{false}, & \text{otherwise,} \end{cases} \quad (1)$$

where $b(R, R') = g\sqrt{(1/2Q)(1/|R| + 1/|R'|)} \ln(2/\delta)$ and R and R' represent a fixed couple of regions of I and $\delta = 1/|I|^2$; merge R and R' if $P(R, R') = \text{true}$. The merging order is to choose a function f to sort pixel pairs in I . One choice of f is

$$f_a(t, t') = |t_a - t'_a|, \quad (2)$$

where t_a and t'_a are the pixel channel values.

The original SRM algorithm is used for segmenting color images that contain only three spectral channels and thus cannot directly segment hyperspectral images. Although

- (1) Perform linear contrast stretch algorithm [20] on the hyperspectral image. This step can ensure the grey value of each hyperspectral band is in [0–255] and enhance the image quality simultaneously.
- (2) Generate a random number v that satisfies the uniform distribution.
- (3) Select the k th band if $F_W(k-1) < v < F_W(k)$, ($1 \leq k \leq p$), where F_W denotes the cumulative density function of the W distribution.
- (4) Set $f_W(k) = 0$ and renormalize the W distribution.
- (5) Repeat Step 2 to Step 4 until three spectral bands have been selected.
- (6) Apply SRM to segment the image composed by the selected three spectral bands.

ALGORITHM 1: The segmentation of hyperspectral image.

a minimum heterogeneity rule based SRM method [21] is proposed for multispectral satellite image segmentation, this study designed a strategy from different viewpoint that fully utilizes the rich spectral information of hyperspectral images. To this end, an improved SRM algorithm is presented in this study. The central idea of the proposed approach is to select three spectral bands from hyperspectral images, which is relying on four different band selection strategies. The advantage of band selection is that it can be repeated and thus will generate multiple information sources and redundant information, which complement each other and improve the robustness. The band selection strategies include a static method which selects three predetermined bands and three dynamic methods which select three bands based on the W distribution. Details of the band selection process are described as follows.

(1) *The First Three PCs.* Firstly, the principle component analysis (PCA) transformation is performed on the original hyperspectral image to select the first three principle components (PCs) as the input of SRM. Although this study utilizes PCA to perform band selection, there are also numerous band selection technologies that can be used, such as linear discrimination analysis (LDA) [22] and nonparametric weighted feature extraction (NWFE) [23]. The interested reader is referred to [17]. It is obvious that this band selection method is static. By contrast, the following three methods are dynamic methods that select three bands based on the W distributions, including (1) uniform distribution, (2) W_{LDA} distribution, and (3) W_{Entropy} distribution.

(2) *Uniform Distribution.* The weights of hyperspectral bands are supposed to be equivalent and three spectral bands are randomly selected as the input of SRM.

(3) *W_{LDA} Distribution.* The W_{LDA} distribution is based on the assumption that the hyperspectral bands weights are unequivocal [24] and can be measured by LDA. The weight of each band is computed by the following equation:

$$f_{W_{\text{LDA}}}(j) = \frac{J_j}{\sum_{k=1}^p J_k}, \quad J_j = \text{tr}(S_w^{-1} S_b), \quad j = 1, 2, \dots, p, \quad (3)$$

where J_j denotes the discrimination power of the j th band, and S_w and S_b represent within-class scatter matrix and

between-class scatter matrix, respectively. Here, S_w and S_b are defined as

$$\begin{aligned} S_w &= \sum_{i=1}^C \sum_{t=1}^N (x_t^i - \mu_i)(x_t^i - \mu_i)^T, \\ S_b &= \sum_{i=1}^C N(\mu_i - \mu)(\mu_i - \mu)^T, \end{aligned} \quad (4)$$

where C is number of classes, N is number of samples, x represents spectral value, μ_i is i th class mean, and μ is the overall mean of spectral values.

(4) *W_{Entropy} Distribution.* Unlike the W_{LDA} distribution, the W_{Entropy} distribution uses entropy to compute the band weight. Entropy [25] is a statistical measure of randomness that can be used to characterize the texture of the input image. The weight of each band is determined by

$$\begin{aligned} f_{W_{\text{Entropy}}}(j) &= \frac{H_j}{\sum_{k=1}^p H_k}, \\ H_j &= -\sum_{i=1}^n p(x_{ji}) \log_b p(x_{ji}), \quad j = 1, 2, \dots, p, \end{aligned} \quad (5)$$

where H_j is the entropy of j th spectral band.

Once the W distributions are obtained, band selection can be performed using the pseudorandom number generation theory [24, 26]. After the band selection process, the hyperspectral image is segmented using Algorithm 1.

2.2. Spectral-Spatial Classification. The spectral-spatial classification is performed to postprocess pixel-wise SVM classification result after segmentation results obtained by SRM. In this study, the scheme [14, 15] (see Figure 2), which combines advances of segmentation and classification results, is selected to implement spectral-spatial classification. Particularly, the majority voting algorithm is selected as the decision fusion rule, as it is easy to implement. The SRM segmentation and spectral-spatial classification will be repeated 200 times, resulting in 200 classification results. Finally, these 200 classification results will be fused using the majority voting algorithm to produce the final classification result.

TABLE 1: Class-specific accuracies in percentage for the ROSIS image by different classifiers.

Class	Samples		Method				
	Train	Test	Pixel-wise SVM (%)	Three PCs (%)	W_{LDA} (%)	$W_{Entropy}$ (%)	$W_{Uniform}$ (%)
C1	252	567	90.30	93.83	98.57	99.47	96.17
C2	135	355	98.56	88.02	90.08	89.82	40.70
C3	720	1697	92.29	98.56	98.51	98.59	96.44
C4	1260	2961	92.82	98.41	98.59	98.59	89.54
C5	91	214	99.41	99.63	99.41	99.48	98.74
C6	198	463	70.00	78.70	79.33	80.13	96.23
C7	173	323	68.65	73.42	99.67	100	86.47
C8	644	1619	81.41	92.32	95.19	94.89	87.42
C9	513	1125	96.52	98.31	99.58	99.68	55.12

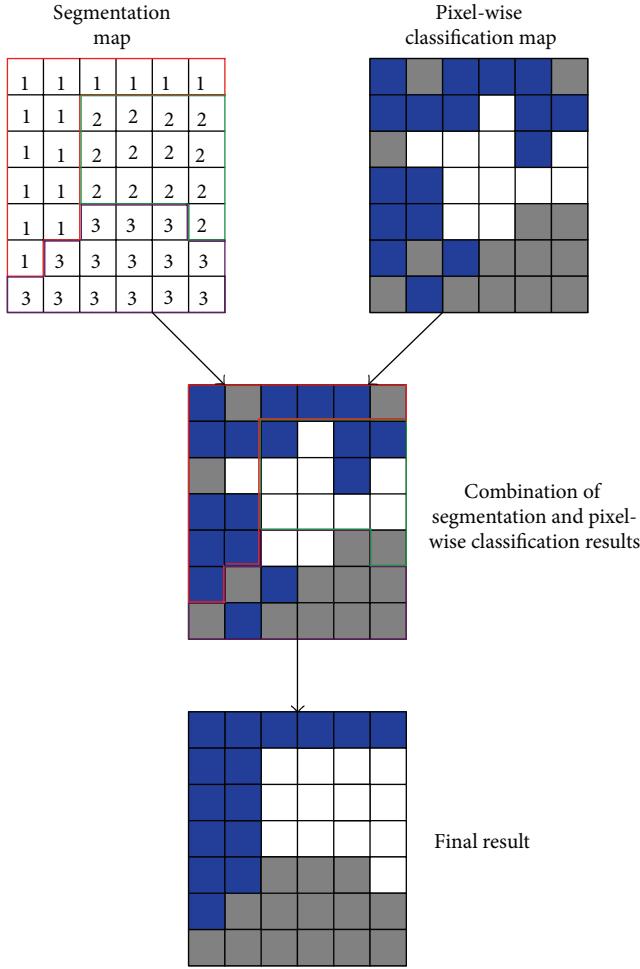


FIGURE 2: The logic flow of the spectral-spatial classification [15].

3. Experiments

In order to evaluate the performance of the proposed spectral-spatial classification approach, experiments on two hyperspectral images were carried out. The first experiment used a ROSIS image whereas an AVIRIS image was used in

the second experiment. In this study, MATLAB with R2010b version was used as the coding environment on a PC that has Intel Core2Quad processor with 2.83-GHz clock speed.

3.1. First Experiment. The University of Pavia image is of an urban area recorded by the ROSIS-03 optical sensor, with an image size of 610×340 pixels. The image has a spatial resolution of 1.3 m per pixel and the number of spectral bands is 115, which ranges from 0.43 to 0.86 μm . It should be noted that the 12 noisiest channels have been removed in the preprocessing step. There are 9 classes in the first experiment, denoted as C1, C2, C3, C4, C5, C6, C7, C8, and C9. The training and testing data sets used in this experiment were provided by Professor Paolo Gamba from University of Pavia. A false color image (bands 50, 27, and 17) and the ground truth data are shown in Figures 3(a) and 3(b), respectively.

The supervised classification was firstly created by the multiclass SVM and without feature selection. Table 1 shows the training and testing sample numbers for each class. This study implemented the SVM classification by LIBSVM library [27]. The Gaussian radial basis function (RBF) kernel was used in this experiment and the optimal parameters C and γ were determined by fivefold cross validation, resulting in $C = 128$, $\gamma = 0.125$. The pixel-wise SVM classification result is shown in Figure 5(a). After pixel-wise SVM classification, SRM algorithm was then applied to segment the hyperspectral image. Three bands were selected based on W distribution as the input of SRM. A Matlab toolbox [28] was used to perform the SRM algorithm. The Q value of SRM was defined as 1024 in this study. Figure 4 shows the SRM results using different band selection strategies.

The spectral-spatial classification was then performed after the segmentation maps were obtained. Figures 5(b)–5(e) show the spectral-spatial classification results using different band selection strategies. From the visual analysis, it can be seen that the pixel-wise SVM result looks like more “noisy” than that of spectral-spatial classification methods. By contrast, the latter provide more homogeneous regions than pixel-wise SVM. In order to quantitatively evaluate the performance of the proposed approach, two measures, including (1) overall accuracy (OA, the number of well-classified samples divided by the number of test samples)

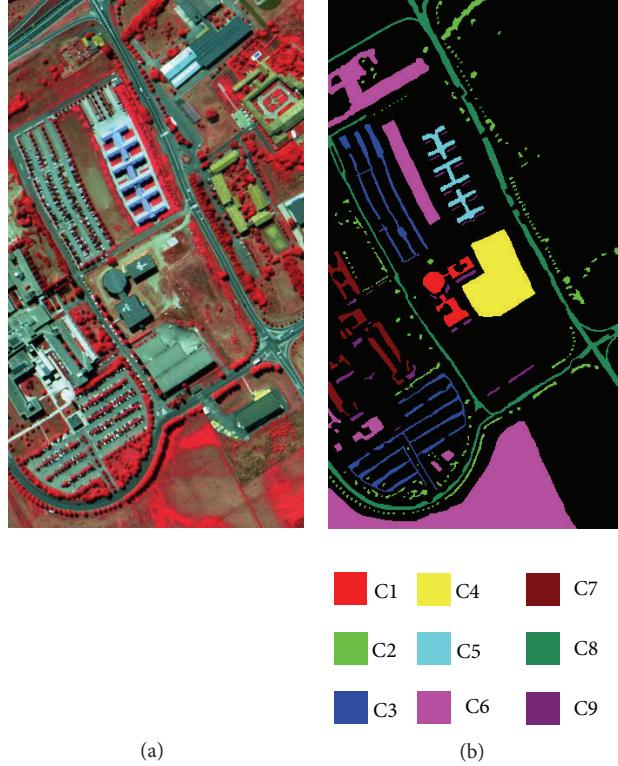


FIGURE 3: ROSIS image of University of Pavia. (a) False color image. (b) Corresponding reference map.

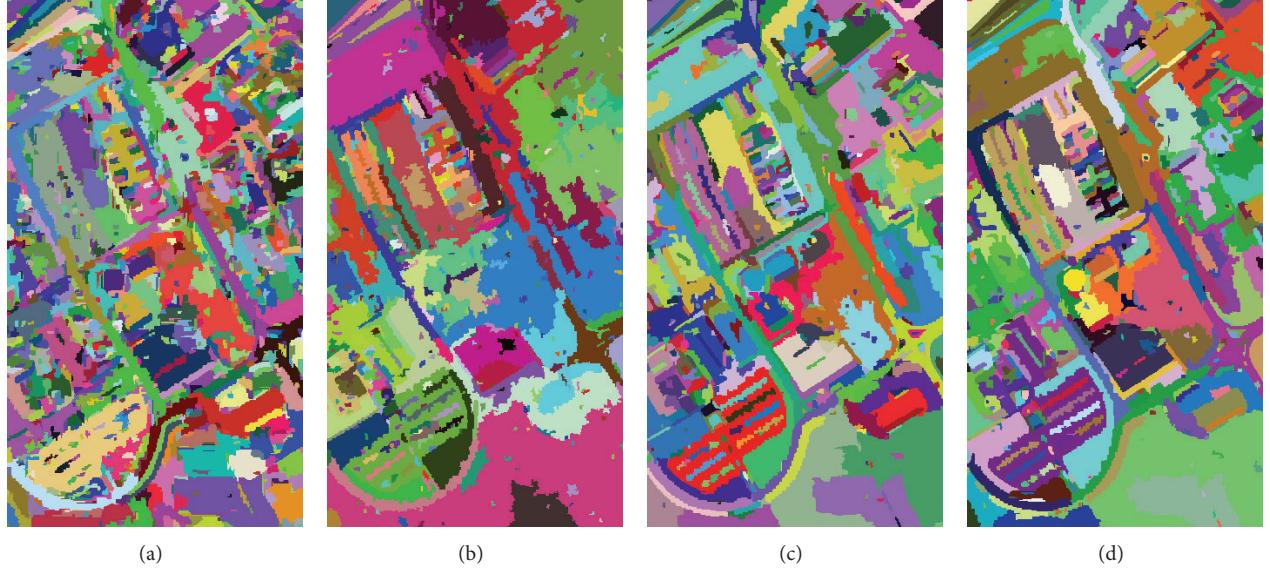


FIGURE 4: SRM results on the ROSIS image using different band selection strategies: (a) the first three PCs; (b) uniform distribution; (c) W_{LDA} distribution; (d) W_{Entropy} distribution.

and (2) kappa coefficient (κ , the percentage of agreement corrected by the amount of agreement that would be expected by chance alone), are used in this study.

Table 1 reports the classification accuracy for each class. From Table 1, it can be seen that spectral-spatial classification methods can improve the classification accuracy of

the pixel-wise SVM except C2 class. The reason for this phenomenon is that C2 class is the feature with small area, whose spatial structure is likely damaged in the process of image segmentation. Despite this drawback, the spectral-spatial classification methods still show eminently satisfactory results. The comparison of the SVM classification and

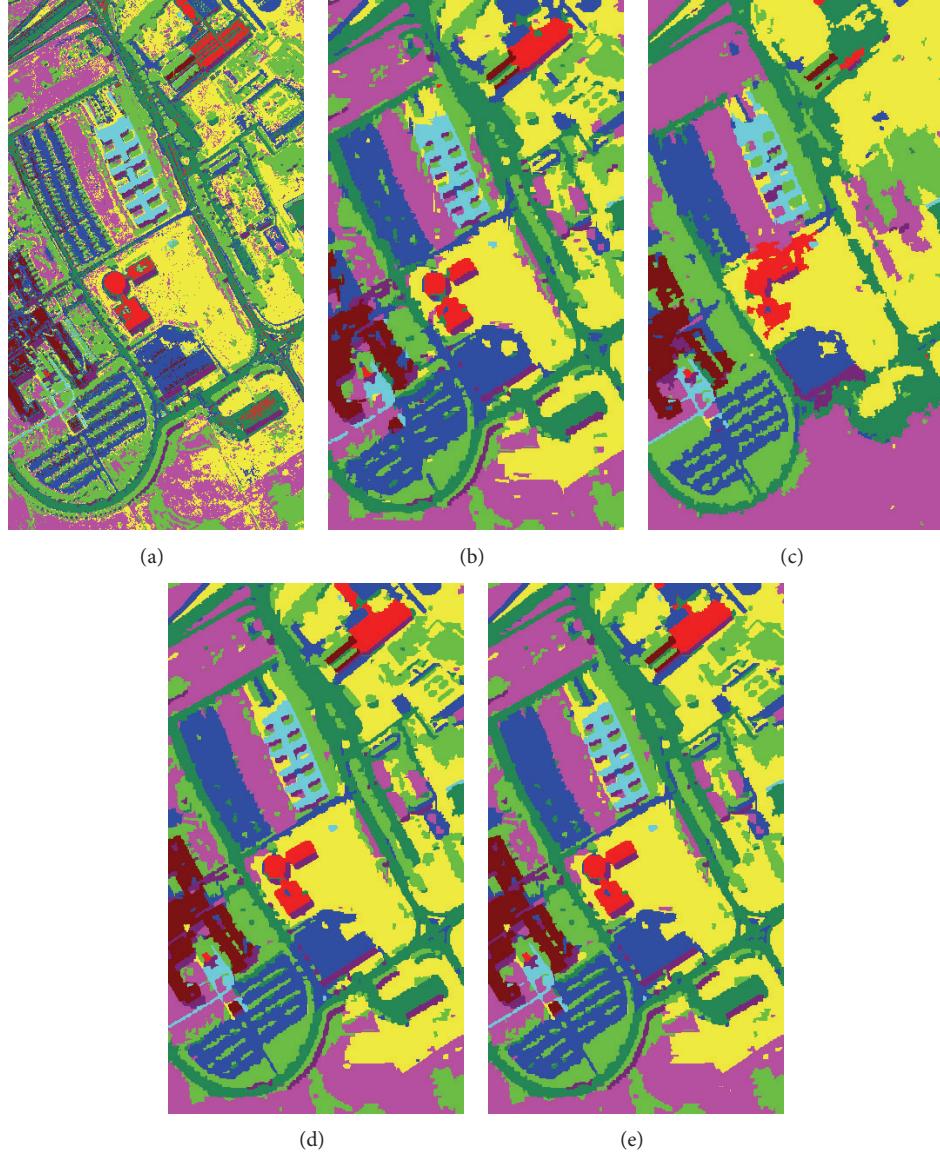


FIGURE 5: University of Pavia image classification result: (a) SVM classification result; (b) the first three PCs; (c) majority vote result of uniform; (d) majority vote result of LDA; (e) majority vote result of Entropy.

TABLE 2: Comparison of the SVM and the developed spectral-spatial classification method for University of Pavia image.

	Method	OA (%)	κ (%)
Static	Pixel-wise SVM	80.49	75.59
	Three PCs	86.81	83.20
Dynamic	Uniform	88.81	85.18
	LDA	89.15	86.18
	Entropy	89.49	86.59

the spectral-spatial classification is given in Table 2. As is seen from Table 2, compared to the pixel-wise SVM, OA and κ improve about 6%~9% and 7%~11%, respectively. This table clearly shows that the proposed approach has higher classification accuracy with respect to the pixel-wise SVM (i.e.,

the approach proposed outperforms the pixel-wise SVM). The improvement in the volume of classification accuracy coincides with the assumption that integrating multiple information sources (i.e., spectral and spatial information) can reduce classification errors. Meanwhile, Table 2 also shows that both three dynamic band selection methods achieve higher classification accuracy than static methods. This is due to the fact that dynamic band selection methods can provide more rich spatial information than static methods. In particular, the band selection based on W_{Entropy} achieves the highest classification accuracy in this experiment.

To assess the impact of the presented algorithm on the results of hyperspectral image classification, a comparison was carried out among SVM plus majority vote method (SVMMV) [15], dynamic subspace method with random band selection method (DSM) [24], dynamic subspace

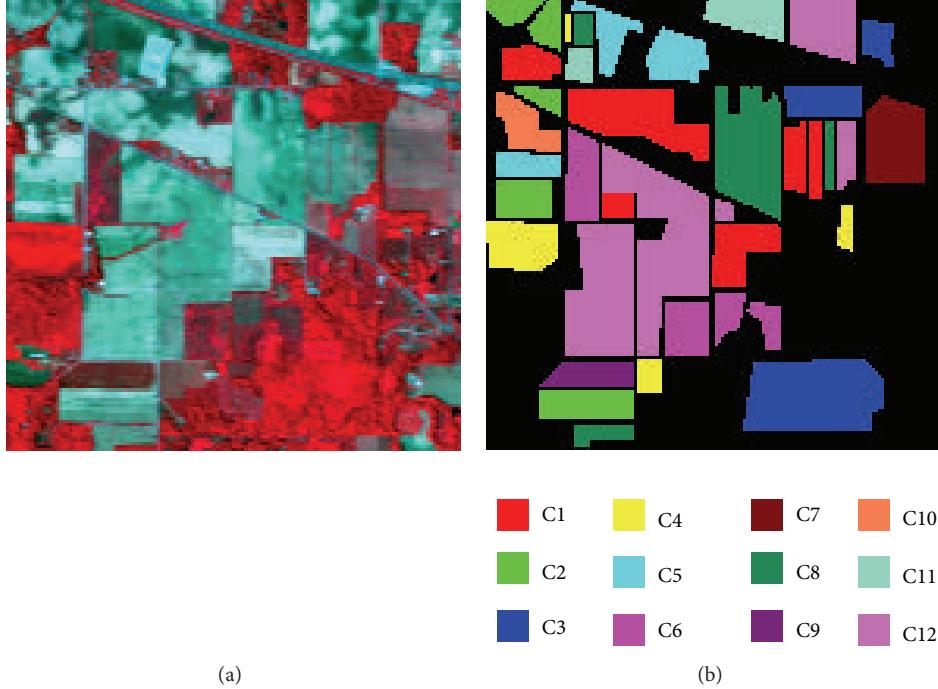


FIGURE 6: (a) AVIRIS image of Indian Pines (50, 27, and 17). (b) Corresponding reference map.

TABLE 3: Quantitative evaluation of different spectral-spatial classification methods on the ROSIS dataset.

Method	OA (%)	κ (%)
SVMMV [15]	85.42	81.30
DSM [24]	87.51	85.20
DSMw2 [24]	88.76	86.22
The proposed approach	89.49	86.59

method with LDA distribution method (DSMw2) [24], and the proposed approach. As shown in Table 3, it depicts the behavior of OA and κ among these four methods. The OA yielded by SVMMV, DSM, DSMw2, and the proposed approach was equal to 85.42%, 87.51%, 88.76%, and 89.49%, respectively, while κ values were equal to 81.30%, 85.03%, 86.22%, and 86.59%, respectively. The proposed approach outperforms SVMMV, DSM, and DSMw2 obviously, which indicates that the proposed method is more suitable for hyperspectral image classification than the other three methods.

3.2. Second Experiment. The Indiana Indian Pines hyperspectral image captured by the AVIRIS sensor on June 12, 1992, was used in the second experiment. The data and corresponding true ground data, as shown in Figure 6, are provided by Professor David A. Landgrebe from Purdue University. The AVIRIS data is composed of 220 spectral bands with a spatial resolution of 20 m per pixel [29]. In this experiment, a subsection of the original Indian Pines with a size of 145×145 pixels was used. Twenty bands with water absorption were discarded, resulting in 200 bands. There are

16 different classes in the original Indiana image. Four of these 16 classes were discarded due to their sample size, resulting in 12 classes for this experiment labelled as C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, and C12. Detailed information about the 12 classes is given in Table 4, with a number of samples for each class in the available reference data.

First, SRM based on four band selection strategies were applied to segment the hyperspectral image, as shown in Figure 7. Here, the parameter Q of SRM was given by 8192. Next, 30% samples for each class were randomly chosen from the ground truth data as training samples. Based on these training datasets, the optimal parameters C and γ were determined as 512 and 0.0078, respectively. Figure 8(a) shows the classification result of the pixel-wise SVM. After that, the SRM segmentation results, based on four band selection schemes, were used to refine the pixel-wise SVM classification result, producing results in Figures 8(b)–8(e). As can be seen from Figure 8, SVM classifier suffers from “salt and pepper” effects that lead to the decrease of classification performance. By contrast, spectral-spatial classifiers produce more homogeneous areas and hence tackle this limitation to a certain extent, which in turn improves the classification accuracy.

For the purpose of quantitative comprise, Table 4 summarizes the classification accuracy for each class of different methods. As hinted by Table 4, spectral-spatial classification methods can improve the classification accuracies of most classes produced by pixel-wise SVM. Table 5 comprises the pixel-wise SVM and spectral-spatial classification method based on different band selection strategies. As can be seen from Table 4, compared to pixel-wise SVM, the proposed approach based on four band selection strategies improves

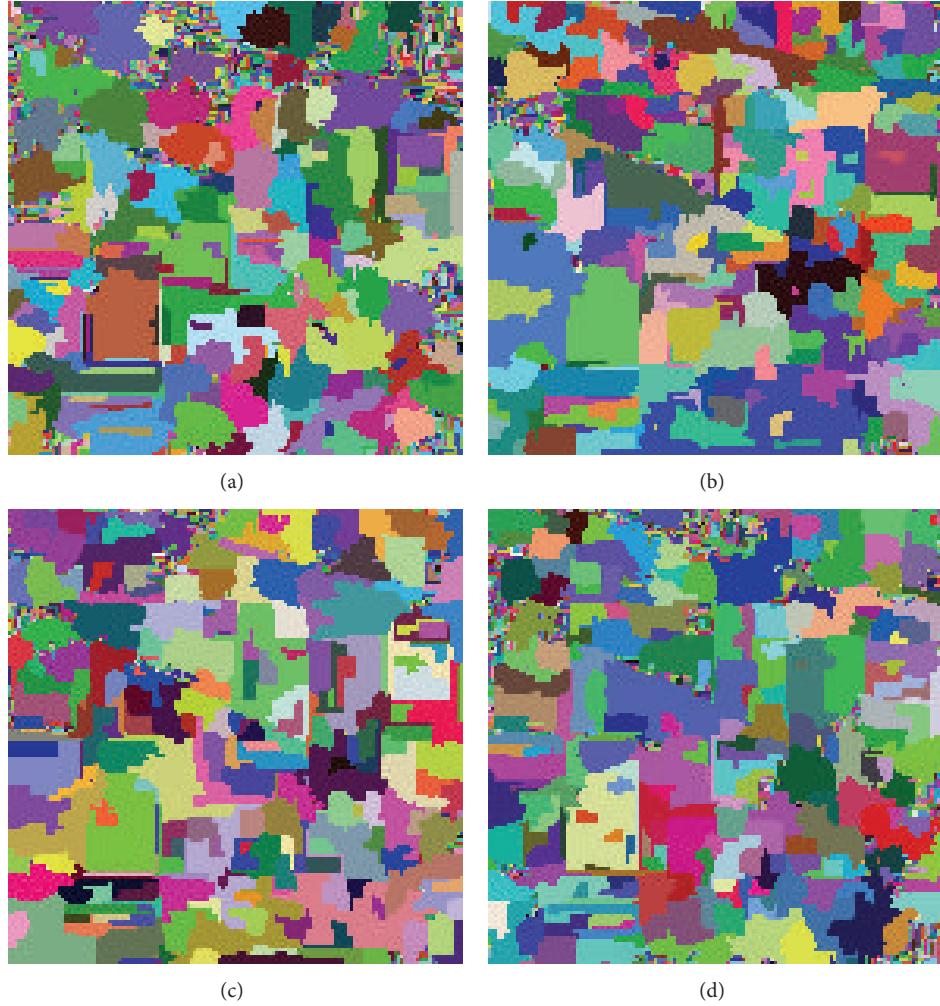


FIGURE 7: Examples of SRM segmentation results using different band selection strategies: (a) the first three PCs; (b) uniform distribution; (c) W_{LDA} distribution; (d) $W_{Entropy}$ distribution.

TABLE 4: Comparison of class-specific accuracies in percentage for the Indiana image by different methods.

Class	Samples		Pixel-wise SVM (%)	Method			
	Train	Test		Three PCs (%)	W_{LDA} (%)	$W_{Entropy}$ (%)	$W_{Uniform}$ (%)
C1	422	1012	84.67	88.91	92.12	93.10	93.51
C2	252	582	73.91	72.54	93.65	94.96	95.20
C3	392	902	95.79	97.76	99.92	98.38	98.92
C4	150	347	92.70	93.36	95.98	96.38	96.38
C5	198	416	83.91	97.88	98.86	98.53	99.19
C6	232	515	96.39	89.96	98.53	99.33	99.06
C7	150	339	99.40	99.18	99.18	99.39	99.39
C8	277	691	71.53	75.83	78.41	78.10	78.31
C9	52	160	99.30	99.06	99.53	100	100
C10	64	170	64.44	87.61	75.64	79.06	78.21
C11	112	268	73.99	78.16	99.47	96.05	96.05
C12	760	1708	85.00	96.56	98.91	98.78	98.99

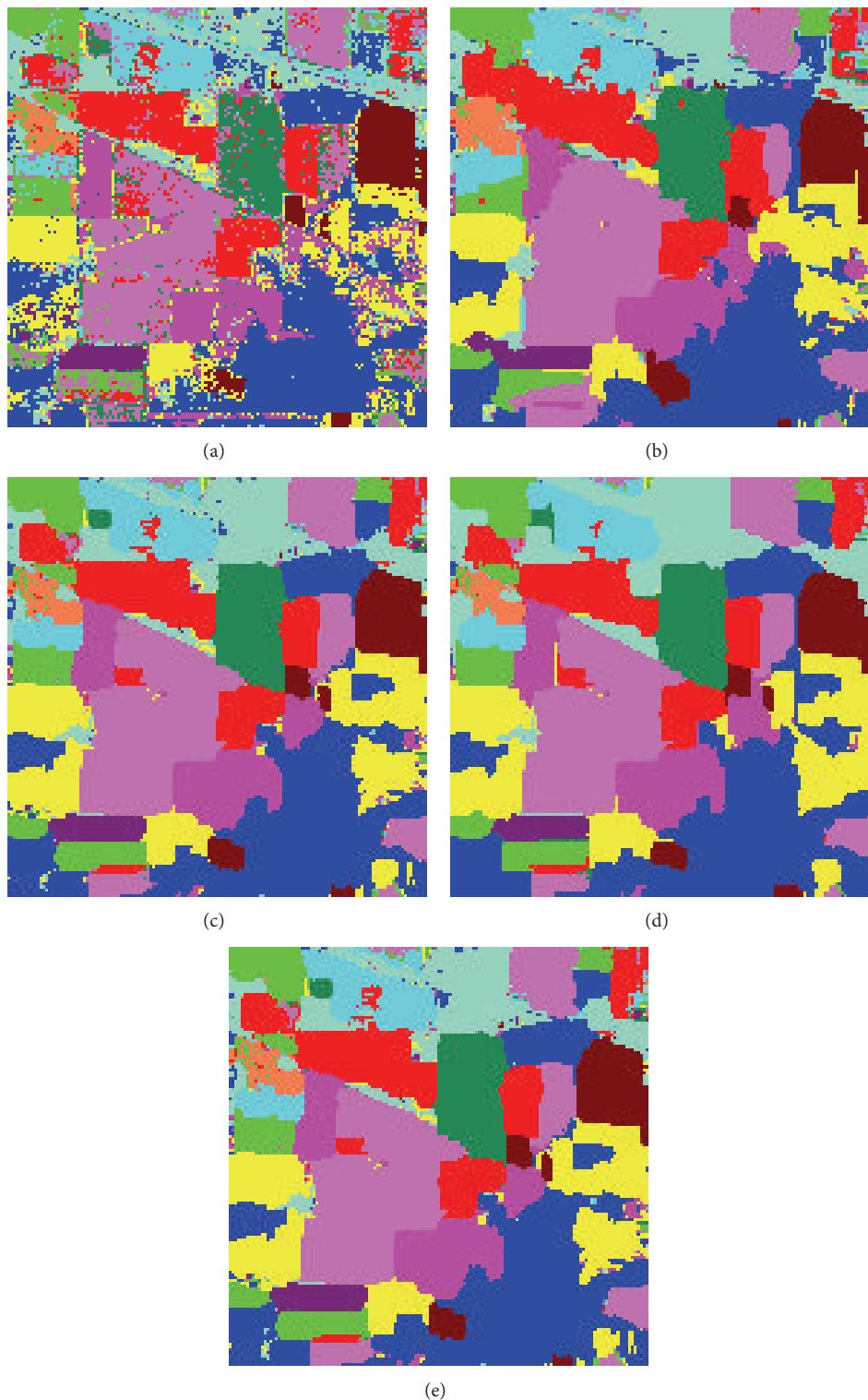


FIGURE 8: Indiana image classification result: (a) SVM classification result; (b) the first three PCA bands; (c) majority vote result of uniform; (d) majority vote result of LDA; (e) majority vote result of Entropy.

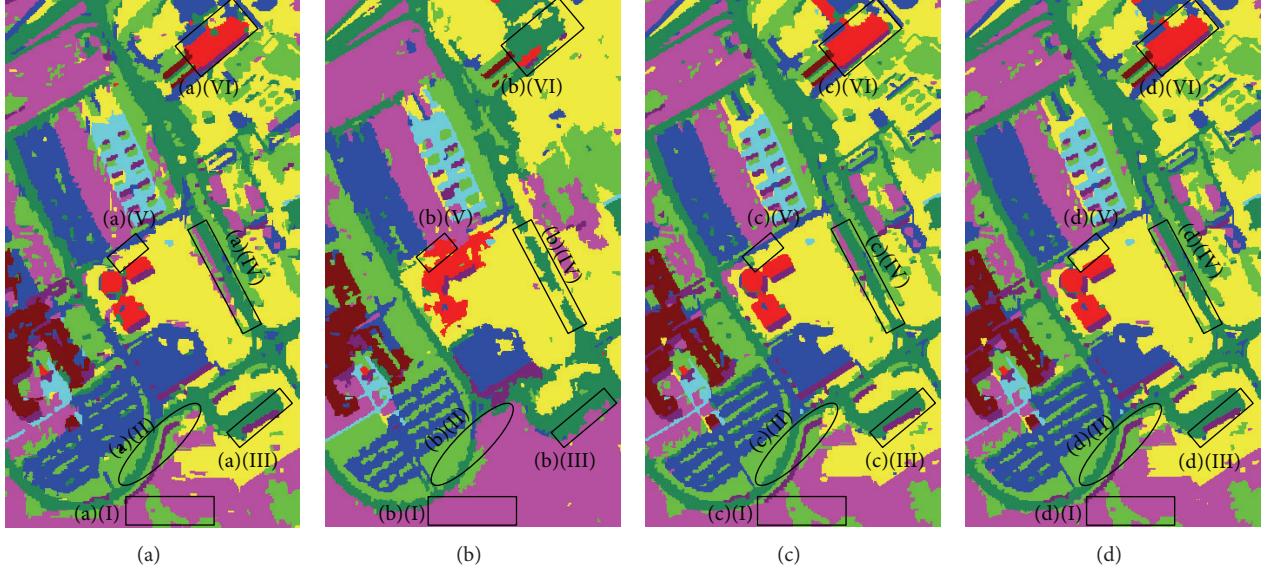


FIGURE 9: This figure shows the visual comparison of spectral-spatial classification of University of Pavia image using different band selection strategies: (a) the first three PCs; (b) uniform distribution; (c) W_{LDA} distribution; (d) $W_{Entropy}$ distribution.

TABLE 5: Comparison of the SVM and the developed spectral-spatial classification method for Indiana image.

	Method	OA (%)	κ (%)
Static	Pixel-wise SVM	85.32	83.14
	Three PCs	90.41	88.97
Dynamic	Uniform	95.27	94.56
	LDA	95.03	94.29
	Entropy	95.05	94.32

OA values by 5.09%, 9.95%, 9.71%, and 9.73%, respectively, while κ values are improved by 5.83%, 11.42%, 11.15%, and 11.18%, respectively. Similar to the first experimental results, both three dynamic band selection methods achieve higher accuracy than static band selection method, which again verifies the superiority of the proposed approach. Meanwhile, as opposed to the first case, band selection method based on uniform distribution produces the best performance in the second case. This indicates that uniform distribution is more suitable for urban areas where the spatial structural is complicated, while entropy distribution is a more proper way for plain areas without many image details.

In the second experiment, to assess the suitability of the proposed approach for the classification of hyperspectral images, a comparison analysis was carried out on three other methods (i.e., SVMMV, DSM, and DSMw2). As reported in Table 6, the proposed method resulted in the highest OA and κ . The quantitative analysis confirms the suitability of the proposed approach on the classification of hyperspectral images.

3.3. Discussion. In the first experiment, band selection using $W_{Entropy}$ gives the highest accuracy (see Table 2). From Table 1, it can be seen that the accuracy of classes C2, C4,

TABLE 6: Quantitative evaluation of different spectral-spatial classification methods on the AVIRIS dataset.

Method	OA (%)	κ (%)
SVMMV [15]	93.78	92.88
DSM [24]	90.20	88.30
DSMw2 [24]	89.50	87.50
The proposed approach	95.27	94.56

and C9 of the spectral-spatial classification based on $W_{Uniform}$ distribution decreases compared to the pixel-wise SVM classification. From Figure 9, it can be seen that the $W_{Uniform}$ distribution method shows a good classification only in area (I) compared to the W_{LDA} and $W_{Entropy}$ distributions. In (II)–(VI) areas, the $W_{Uniform}$ distribution method misclassified pixels more than the W_{LDA} and $W_{Entropy}$ distribution methods did.

In the second experiment, the classification method based on $W_{Uniform}$ achieves the highest accuracy (see Table 4). Figure 10 shows the visual comparison of the spectral-spatial classification results based on different band selection strategies. As can be seen from Figure 10, both dynamic methods performed well on area (I) compared to the band selection method based on the first three PCs. Figure 10 also indicates that dynamic methods achieve similar classification results which has also been confirmed by Table 4. Hence, for images with large spatial structures, the band selection strategy has a low influence on the spectral-spatial classification method when the image has a large spatial structure.

4. Conclusion

An advanced spectral-spatial classification method for classification of hyperspectral images, which combines advances

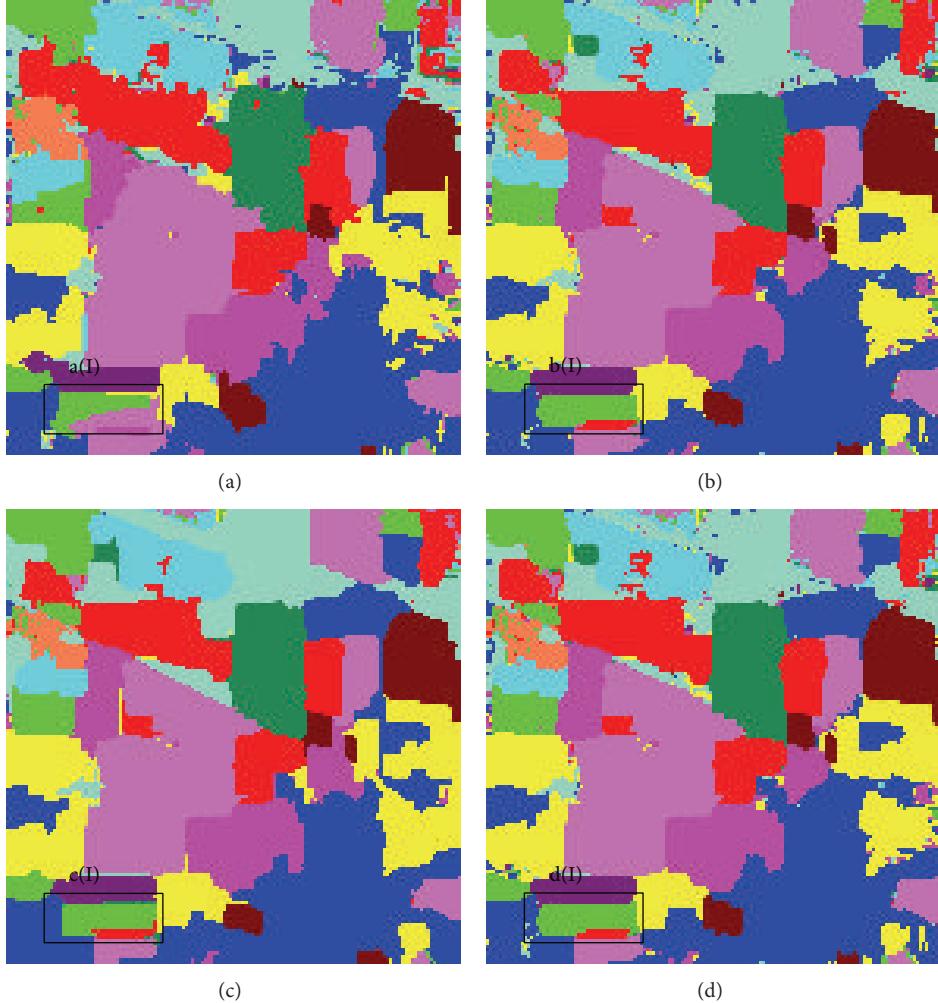


FIGURE 10: This figure gives a visual comparison of spectral-spatial classification for the Indiana image, using different band selection strategies: (a) the first three PCs; (b) uniform distribution; (c) W_{LDA} distribution; (d) $W_{Entropy}$ distribution.

of region-based segmentation and image fusion, has been proposed in this study. The proposed approach has been achieved by (a) integrating pixel-wise support vector machine (SVM) classification and statistical region merging (SRM) segmentation results; (b) multiclassification results fusion using majority voting. Four different band selection strategies have been studied to implement the SRM algorithm to segment the hyperspectral image. The proposed approach has two advantages: (1) it does not need to set cluster numbers in advance; (2) the segmentation does not depend on initial values. These advantages result in higher robustness than the partitional clustering technique and thus make the proposed approach ideal for advanced spectral-spatial classification of hyperspectral images. Furthermore, the proposed approach is easy and efficient to implement. Although the developed method was used to classify hyperspectral images in this study, this method can also be used to classify multispectral images.

In this study, the spatial information is derived from the region-based segmentation results, which suffers from two

main drawbacks: (1) it is difficult to control the scale and (2) it destroys class boundaries. To tackle these limitations, further research will therefore be conducted to improve the segmentation results. Particularly, edge information will be studied to produce precise segmentation result of remote sensing images.

Conflict of Interests

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

Acknowledgments

The work presented in this paper is partly supported by Ministry of Science and Technology of China (Project nos. 2012BAJ15B04 and 2012AA12A305), the National Natural Science Foundation of China (41331175), and Ling Jun Ren Cai Project of National Administration of Surveying,

Mapping and Geo-Information, China. The authors would like to thank Professor David A. Landgrebe from Purdue University for providing AVIRIS datasets, Professor Paolo Gamba from University of Pavia for providing the ROSIS hyperspectral images, and Dr. Hua Zhang and Dr. Yi Liu from China University of Mining and Technology for their kind discussion and support on computer programming.

References

- [1] H. Z. M. Shafri, A. Suhaili, and S. Mansor, "The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis," *Journal of Computer Science*, vol. 3, no. 6, pp. 419–423, 2007.
- [2] F. A. Kruse, A. B. Lefkoff, J. W. Boardman et al., "The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment*, vol. 44, no. 2-3, pp. 145–163, 1993.
- [3] B. C. K. Tso and P. M. Mather, "Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3 I, pp. 1255–1260, 1999.
- [4] U. Maulik and S. Bandyopadhyay, "Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 5, pp. 1075–1081, 2003.
- [5] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.
- [6] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Vilá-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, 2006.
- [7] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. L. Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 6, pp. 1822–1835, 2008.
- [8] M. Pal and P. M. Mather, "Support vector machines for classification in remote sensing," *International Journal of Remote Sensing*, vol. 26, no. 5, pp. 1007–1011, 2005.
- [9] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [10] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823–870, 2007.
- [11] M. Fauvel, *Spectral and spatial methods for the classification of urban remote sensing data [Ph.D. thesis]*, University of Iceland, 2007.
- [12] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, 2008.
- [13] Y. Tarabalka, J. Chanussot, J. A. Benediktsson, J. Angulo, and M. Fauvel, "Segmentation and classification of hyperspectral data using watershed," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pp. III652–III655, Boston, Mass, USA, July 2008.
- [14] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2973–2987, 2009.
- [15] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognition*, vol. 43, no. 7, pp. 2367–2379, 2010.
- [16] Y. Tarabalka and J. C. Tilton, "Spectral-spatial classification of hyperspectral images using hierarchical optimization," in *Proceedings of the 3rd Workshop on Hyperspectral Image and Signal Processing (WHISPERS '11)*, pp. 1–4, June 2011.
- [17] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*, John Wiley & Sons, Hoboken, NJ, USA, 1st edition, 2009.
- [18] E. Frauman and E. Wolff, *Segmentation of Very High Spatial Resolution Satellite Images in Urban Areas for Segments-Based Classification*, 2005, http://www.ecognition.com/sites/default/files/226_frauman_urs2005.pdf.
- [19] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1452–1458, 2004.
- [20] F. A. Kruse, A. B. Lefkoff, J. W. Boardman et al., "The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment*, vol. 44, no. 2-3, pp. 145–163, 1993.
- [21] H. Li, H. Gu, Y. Han, and J. Yang, "An efficient multiscale SRMMHR (Statistical Region Merging and Minimum Heterogeneity Rule) segmentation method for high-resolution remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 2, pp. 67–73, 2009.
- [22] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [23] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096–1105, 2004.
- [24] J.-M. Yang, B.-C. Kuo, P.-T. Yu, and C.-H. Chuang, "A dynamic subspace method for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2840–2853, 2010.
- [25] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 2003.
- [26] L. Devroye, *Non-Uniform Random Variate Generation*, Springer, New York, NY, USA, 1st edition, 1986.
- [27] C.-C. Chang and C.-J. Lin, *LIBSVM: A Library for Support Vector Machines*, 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [28] S. Boltz, "Image Segmentation Using Statistical Region Merging," 2014, <http://www.mathworks.com/matlabcentral/file-exchange/25619-image-segmentation-using-statistical-region-merging>.
- [29] R. O. Green, M. L. Eastwood, C. M. Sarture et al., "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.

Research Article

Airborne Polarimetric Remote Sensing for Atmospheric Correction

Tianquan Liang, Xiaobing Sun, Han Wang, Rufang Ti, and Cunming Shu

Key Laboratory of Optical Calibration and Characterization, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Science, Hefei 230031, China

Correspondence should be addressed to Tianquan Liang; liangpolaris@126.com

Received 2 October 2014; Revised 9 January 2015; Accepted 17 January 2015

Academic Editor: Seong G. Kong

Copyright © 2016 Tianquan Liang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The problem, whose targets can not be effectively identified for airborne remote sensing images, is mainly due to the atmospheric scattering effect. This problem is necessary to be overcome. According to the statistical evaluations method and the different characteristics of polarization between the objects radiance and atmospheric path radiation, a new atmospheric correction method for airborne remote sensing images was proposed. Using this new method on the airborne remote sensing images which acquired on the north coast areas of China during the haze weather, we achieved a high quality corrected atmosphere-free image. The results demonstrate the power of the method on the harbor area. The results show that the algorithm, improving image contrast and image information entropy, can effectively identify the targets after atmospheric correction. The image information entropy was enhanced from 5.59 to 6.62. The research provides a new and effective atmospheric correction technical approach for the airborne remote sensing images.

1. Introduction

Airborne remote sensing plays an important role in searching materials of interest [1], as well as an important platform of national science missions for remote sensing monitoring. This technique exhibits many potentialities in a large amount of applications such as geology mapping, land resources planning, terrain cover information classification, monitoring growth situation of crop, ocean targets detecting, and military surveillance. It is especially important that the aerial remote sensing can provide real-time, actionable data, automatic detection, and classification. The data acquired by aerial remote sensing platform consists of objects radiance and atmospheric path radiance information. The scene is strongly affected by the presence of the atmosphere along the path from the sun to the target to the sensor. These potentialities have motivated the development of atmospheric correction [2].

One of the simplest methods of atmospheric correction is the dark object subtraction method [3, 4]. This method approximates the path radiance from the darkest object in

the scene. Currently, different methods are used to solve the direct transfer radiative problems; for example, they provide atmospheric parameters for atmospheric correction, such as 6S (Second Simulation of the Satellite Signal in the Solar Spectrum) model [5, 6]. It provides a computer code which can accurately simulate atmospheric radiative. There are some other standard procedures such as LOWTRAN and MODTRAN [7, 8]. In the atmospheric transfer equation, it can provide powerful atmospheric parameters to recover the scenes [9]. The atmospheric correction has been experienced for long history. There have been several types of approaches: scene-based empirical approaches, radiative transfer modeling approaches, and hybrid approaches [10, 11].

Recently some researchers have used as few as two images taken through a polarizer at different orientations in the haze weather to recover the scene based on effective atmospheric degradation physical model [12, 13]. The algorithm gives birth to better results, but the work concentrated on the horizontal direction and the scene should include the sky region to estimate the unknown parameters. Christoph Borel accomplished the atmospheric correction by using the recently

released vectorized version of 6Sv of airborne POLDER polarimetric imagery. However, the approach has limitations and the typical bilinear, thin plates spline, and cubic 2-D interpolation methods produced unsatisfactory results [14]. We take advantage of polarization information and statistical evaluations information for the airborne polarization image atmospheric correction. The basic statistical shows that it is very often that some pixels have very low intensity. The pixels were completely shadowed in some case and their radiances were mainly contributed by atmospheric scattering (called “path radiance”). The main advantages of statistical evaluations algorithm are that it is effective and easy to estimate the parameters. These parameters are necessary for the airborne image correction. Compared to other atmospheric correction methods, it does not require any real time ground measurement parameters so that it is simple and relatively straightforward to apply. We construct the airborne polarization remote sensing atmospheric correction model which combined polarization information and statistical evaluations method. The image contrast and entropy are increased dramatically when this algorithm is applied in the airborne polarization remote sensing image acquired in the haze weather.

2. Principles and Method

To make the paper self-contained, this section will review the known formation model of haze images, as shown in Figure 1.

The airborne polarization remote sensing acquired images composed of two main components. The first originated from the object radiance. Let us denote by $J(x, y)$ the object radiance as if it was taken in clear atmospheric circumstances, without scattering and absorbing in the line of sight. Due to attenuation in the haze, the airborne polarization camera senses a fraction of this radiance. This attenuated signal is called direct transmission. The second component is known as path radiance, or airlight, or ambient light. It originates from the atmosphere illumination (e.g., sunlight), a particle radiates the light it absorbs, behaving as a light source suspending in the atmosphere, the light scattered a great many times by a huge number of particles. All these particles generate the atmospheric light, a portion of which is scattered into the line of sight by the atmospheric. The model can be described as follows:

$$I(x, y) = J(x, y)t(x, y) + A(1 - t(x, y)), \quad (1)$$

where $I(x, y)$ is the total intensity which acquired by airborne polarization sensor and $t(x, y)$ is the medium transmission which describing the portion of the light that is not scattered and reaches the airborne polarization sensor. The goal of atmospheric correction is to recover $J(x, y)$ from $I(x, y)$. $t(x, y) = \exp(-\beta(\lambda) * z(x, y))$ called atmospheric transmission impact factor, where $\beta(\lambda)$ represents the atmospheric extinction coefficient which is determined by the particle material, size, shape, and concentration. $z(x, y)$ indicates the airborne polarization sensor flying height information. “ $-$ ” indicates the light is weakened.

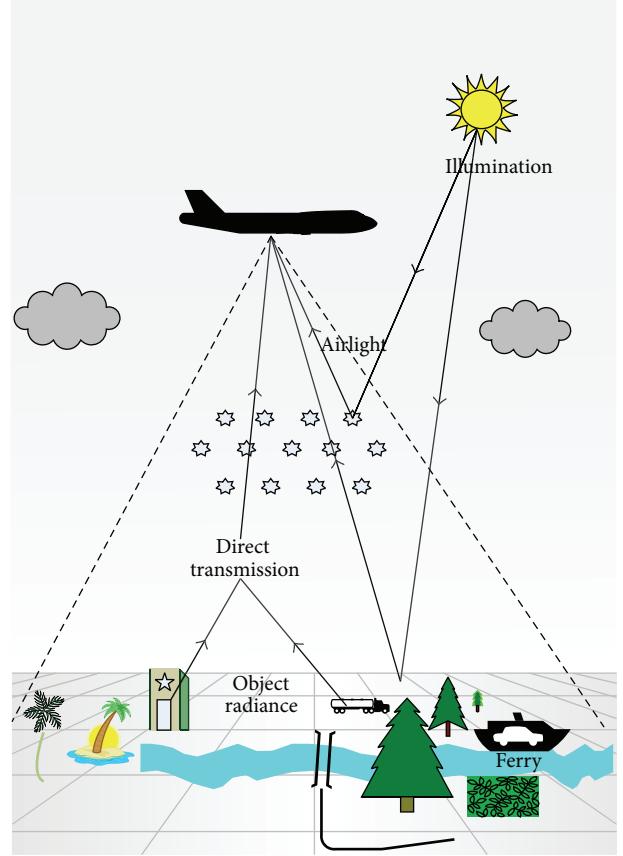


FIGURE 1: The image model for airborne polarization remote sensing.

We will observe strong polarization effect of atmospheric path radiance due to scattering by the atmosphere particles. The assumption which we make in this paper is that light emanating from objects has insignificant polarization. Because scattering does not change the direct transmission light polarization state, it follows that the polarization of the direct transmission is also insignificant. Thus the polarization of the atmospheric path radiance dominates the airborne polarization sensor measured light. As a result atmospheric path radiance information can be expressed as

$$\begin{aligned} A(1 - t(x, y)) &= A^{\text{atm}}(x, y)_{\max} + A^{\text{atm}}(x, y)_{\min} \\ &= \frac{A^{\text{atm}}(x, y)_{\max} - A^{\text{atm}}(x, y)_{\min}}{p_A} \\ &\approx \frac{I(x, y)_{\max} - I(x, y)_{\min}}{p_A}, \end{aligned} \quad (2)$$

where p is atmospheric path radiance degree of polarization. It can be defined as

$$p_A = \frac{A^{\text{atm}}(x, y)_{\max} - A^{\text{atm}}(x, y)_{\min}}{A^{\text{atm}}(x, y)_{\max} + A^{\text{atm}}(x, y)_{\min}}. \quad (3)$$

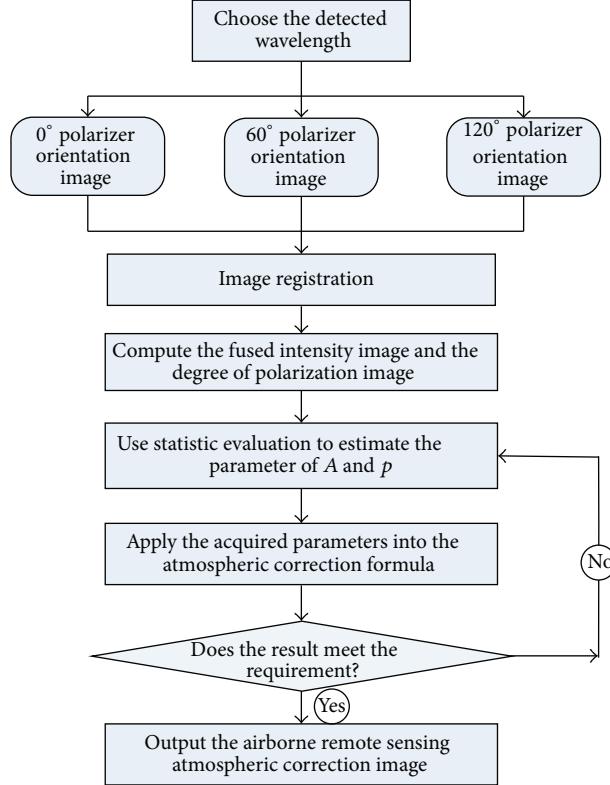


FIGURE 2: The flowchart of the airborne polarized atmospheric correction method.

According to above analysis, the path radiance can be derived from (2) and (3). The atmospheric transmission impact factor $t(x, y)$ can be expressed as (5). Consider

$$A^{\text{atm}} = \frac{I(x, y) \times P(x, y)}{p_A(x, y)} \quad (4)$$

$$t(x, y) = 1 - \frac{I(x, y) \times P(x, y)}{A(x, y) \times p_A(x, y)}. \quad (5)$$

Notice that A represents the atmospheric path radiance without influence of object radiance. The parameters A and p_A were unknown. How do we estimate these parameters? We will bring forth new ideas to give the value of these parameters. It is a innovative method distinguish from other methods. It is statistical evaluations method, and we use statistic for all the pixels and then set threshold value to estimate these parameters, because we find that the images acquired by the airborne polarization sensor will have dark pixels which have low intensities. It is mainly due to some factors, like shadows effect and dark objects or surfaces, for example. We will explicitly introduce the computation procedure at the end of this section.

Polarization property of electromagnetic radiation can be described by the Stokes vector $(I, Q, U, V)^T$ [15]. Because the natural objects have low value of circular polarization, in practical measurement we do not consider circular polarization component, namely, the V component of the Stokes vector. Utilizing a single rotatable linear polarizer, we set the

reference direction, and the Stokes vector can be calculated by rotating three arbitrary polarization orientation angles of polarizer:

$$I(\alpha) = \frac{1}{2} (I + Q \cos 2\alpha + U \sin 2\alpha), \quad (6)$$

where α is the angle between the reference direction axis and the orientated direction. Stokes vector can deduce the degree of polarization:

$$P = \frac{\sqrt{Q^2 + U^2}}{I}. \quad (7)$$

Based on these discussions, the final airborne atmospheric correction result, namely, the objects radiance $J(x, y)$, is recovered by

$$J(x, y) = \frac{I(x, y) - (I(x, y) \times P(x, y) / p_A(x, y))}{1 - (I(x, y) \times P(x, y) / A(x, y) \times p_A(x, y))}. \quad (8)$$

In order to give an intuitive act of the airborne polarized atmospheric correction method, it is necessary to depict the flowchart of correcting method as shown in Figure 2.

3. Instrument and Experiment Description

We use the instrument named the Directional Polarimetric Camera (DPC) which was designed by Anhui Institute of Optics and Fine Mechanics, Chinese Academy of

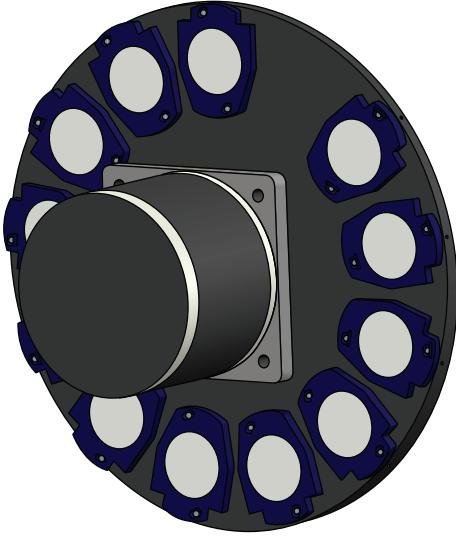


FIGURE 3: Polarizer/filter wheel.

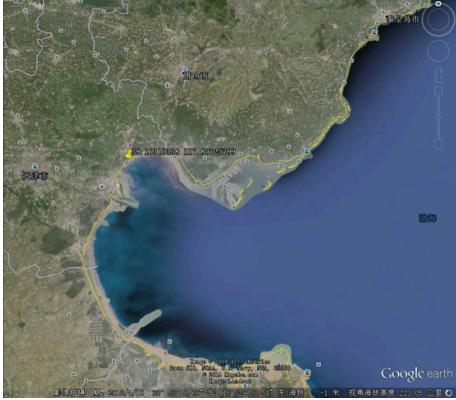


FIGURE 4: Tianjin-Bohai-Tangshan areas in China's north coast.

Science [16]. It has three polarized wavebands (495 nm/670 nm/870 nm) and three unpolarized wavebands (555 nm/780 nm/815 nm). It recorded images through a rotated filter wheel with spectral filter and in three polarized spectral bands with linear polarized filters orientated at 0° , 60° , and 120° . The images needed to be registered before spectra or the Stokes vectors can be computed. The wheel was shown in Figure 3. It is divided into thirteen sections. The FOV is from $-30^\circ \sim +30^\circ$ and on the diagonal it is from $-40^\circ \sim +40^\circ$.

The DPC have taken three flights experiments at Tianjin-Bohai-Tangshan areas of China north coast (as shown on Figure 4) during March 21/26/27, 2012. The experiment data were huge; one flight acquired almost twenty thousand images. The scenes included farmland, city circle, countryside, harbor, and ocean. The weather was haze. As seen from the airplane the scenes were obscured by the atmosphere.

The plane has fixed a POS instrument with DPC at the same time. It will record the plane's position information including latitude, longitude, and altitude. It can create highly accurate flight mission planning and accurate sensor positioning and image acquisition. According to the POS

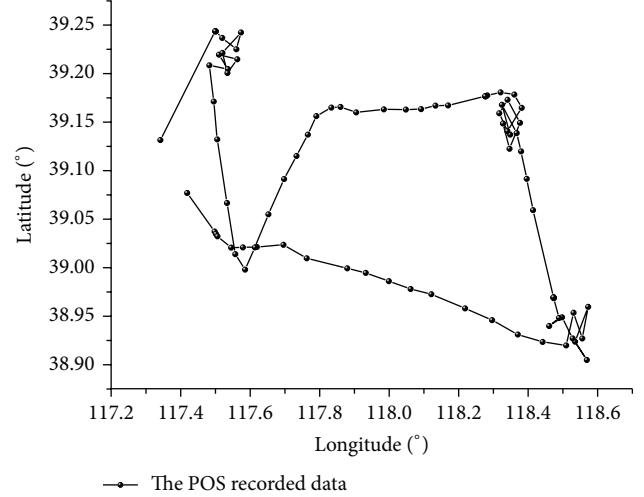


FIGURE 5: The accurate flying trace in the Tianjin-Bohai-Tangshan area.

recorded information, we depicted the accurate flying trace of Tianjin-Bohai-Tangshan area as shown in Figure 5. The flight mission sets three important observer areas. The flying routes were intensive in the interesting areas as shown in Figure 5.

4. Experiment Results

The experiment data which we use were taken on March 21, 2012. We choose one group 670 nm data to prove our method. The data which we choose were acquired at coast ($39^{\circ}09'19.78''$ N, $117^{\circ}48'29.83''$ E); the flying altitude was 3608 m. Because the DPC structure adopts polarizer/spectral filter wheel, one group data will have some difference scenes. Generating images that need to be registered before the Stokes vectors can be computed. This procedure was depicted in the flowchart as shown in Figure 2.

In this section we first display the original images which DPC acquired in the haze weather. It includes three different polarizer orientation angles (namely, 0° , 60° , and 120°), as shown in Figure 6.

We then use the airborne polarization atmospheric correction method to recover the scene. The scene which we chose was a port at Tianjin, China. This port changed into a park, because of "Kyiv" aircraft carrier which China government bought from Russia. Our approach requires image registration for three different polarizer orientation angle images, because of the plane is flying when record the scene it will have some difference in sequence images. Using the registered images we can compute the fused intensity image and the degree of the polarization image. Then we use statistical evaluations to estimate the unknown parameters A and p_A . Finally, we apply the acquired parameter to recover the scene, namely, the airborne polarization image atmospheric correction. The atmospheric correction results are as shown in Figure 7. The contrast of features in the atmospheric corrected image is greatly improved relative to the fused intensity image and three original polarizer orientation

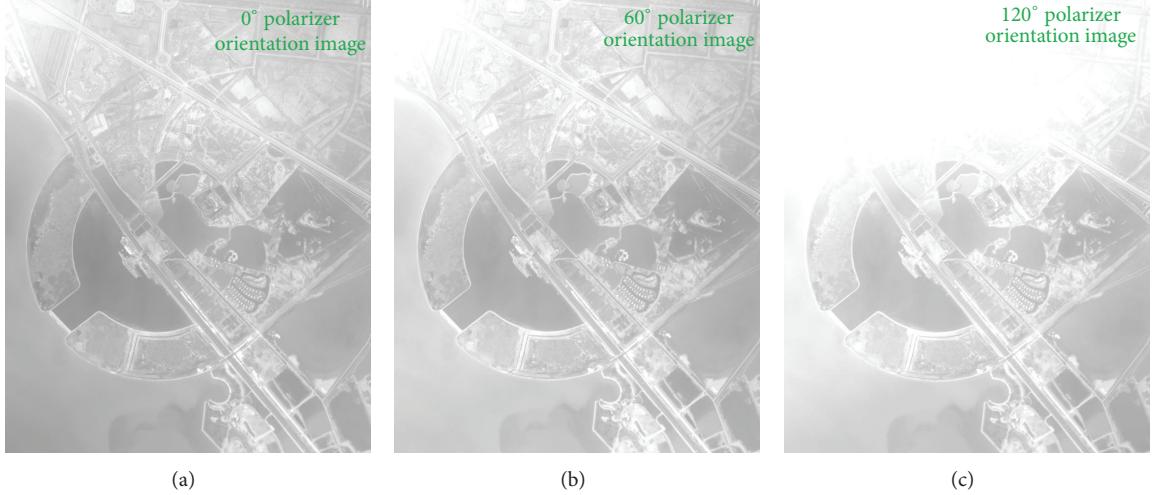


FIGURE 6: The raw images of three different polarizer orientation angles ((a)/(b)/(c) represented 0°/60°/120° polarizer orientation angle images, resp.).

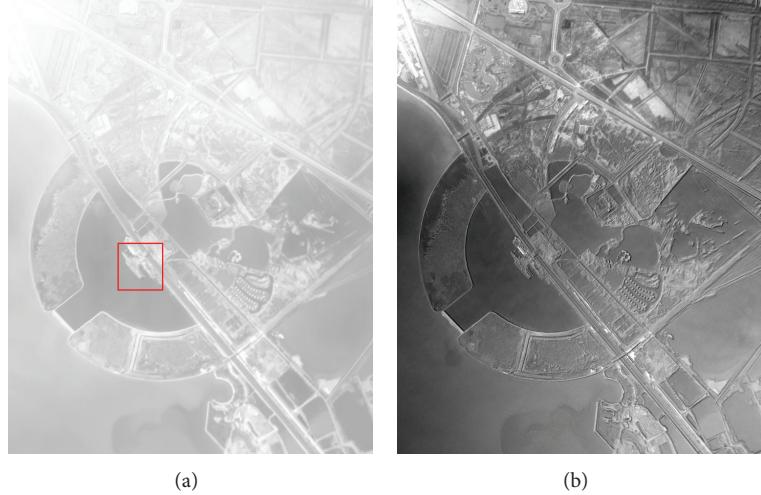


FIGURE 7: The fused intensity image (a) and the atmospheric corrected image (b).

angles raw images. Moreover, the algorithm removed the glint light, which existed in the raw images. Thus the “Kyiv” aircraft carrier is clearly visible in the scene, while in the raw images that area looks like white and blur. The recovered image show details unseen in the raw images of Figure 7.

In order to give restricted and objective results, atmospheric corrected results should make quantitative evaluations. We chose aircraft carrier area as shown in Figure 8, the red rectangle area marked in Figure 7. Compare the histogram of atmospheric correction results as shown in Figure 9. After atmospheric correction the image nonzero value distributed broadly and evenly. It proves that the image contrast is higher than the intensity image. We chose the aircraft carrier as the target and computed the average value of the target area. We chose the ocean around the target as the background and computed the average value of the background area. The object and background area were chosen as the same area of intensity image and atmospheric

corrected image, as shown in Figure 8. The image contrast was enhanced from 0.058 to 0.536. The contrast principle was defined as

$$R = \frac{|f_{\text{object}} - f_{\text{background}}|}{f_{\text{background}}}. \quad (9)$$

In 1948, Shannon introduced a general uncertainty measure on random variables which takes different probabilities among states into account [17]. Today this measure is known as “Shannon’s entropy.” Given events occurring with probability P , the Shannon entropy is defined as

$$H = \sum_{i=1}^m P_i \log_2 \frac{1}{P_i} = -\sum_{i=1}^m P_i \log_2 P_i. \quad (10)$$

Shannon’s entropy can compute an image, where probabilities of the gray level distributions are considered in the

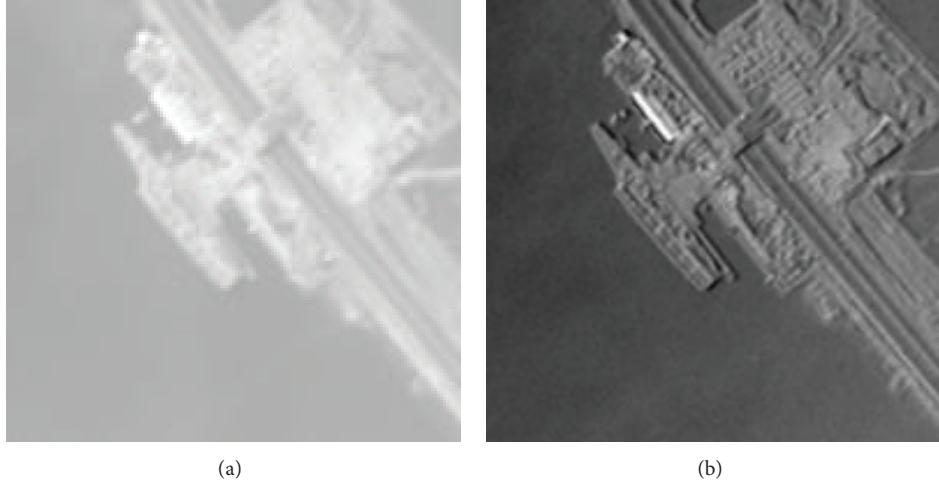


FIGURE 8: The red rectangle area marked in Figure 7 before (a) and after (b) atmospheric correction.

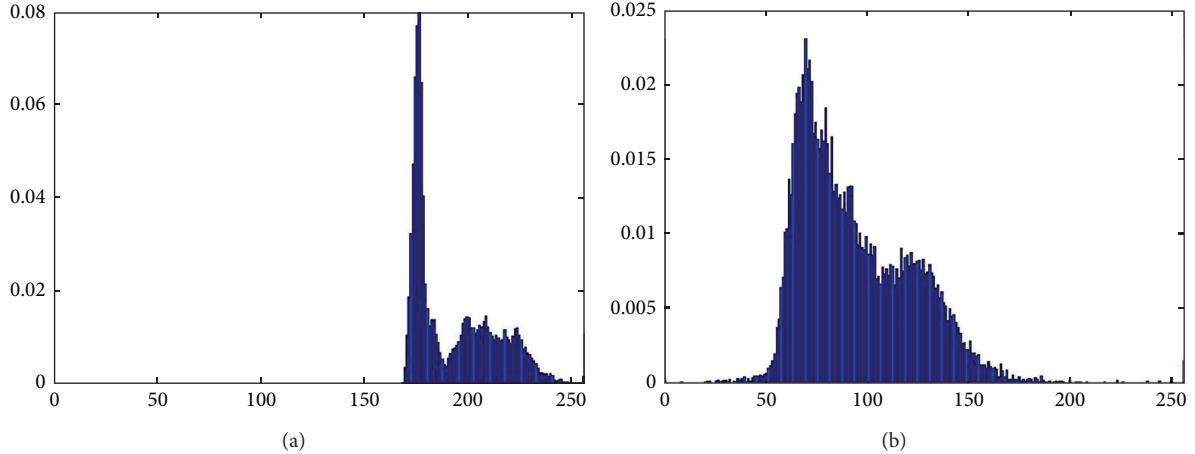


FIGURE 9: The normalized histogram of before (a) and after (b) atmospheric correction.

Shannon entropy formula. A probability distribution of gray values can be estimated by counting the number of times each gray value occurs in the image and dividing those numbers by the total number of occurrences. The image scene is more disorderly (namely, the scene objects more clearly), and the Shannon entropy is higher. After atmospheric correction, the entropy of intensity image has been enhanced from 5.59 to 6.62. As a result of the effect of atmosphere on the intensity image, which has led the whole system relatively orderly, the scenes texture is blurring. The whole system of atmospheric corrected image is more complex and the scenes details texture become clear.

5. Conclusions

In this paper, an approach for airborne polarization atmospheric correction was presented. It is based on analysis of images acquired through Directional Polarimetric Camera (DPC). The method is physics-based which enables a very

effective recovery of the scene. It is a potentially useful tool for airborne atmospheric correction and target recognition applications. Additional work certainly needs to be done to further reach the goal of making the method full proof. In addition, further research and development of hardware are needed to obtain fast acquisition that does not need image registration and can handle uneven atmosphere.

Conflict of Interests

The authors declare no conflict of interests.

Acknowledgments

The authors would like to thank Professor Seong Kong and anonymous reviewer(s) for their valuable comments and suggestions to improve this paper. This work was supported by the National Basic Research Program (973) of China (no. 2010CB950803, no. 61322502).

References

- [1] C. Archer and J. Morgenstern, "Improved target recognition with live atmospheric correction," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*, vol. 8743 of *Proceedings of the SPIE*, Baltimore, Md, USA, May 2013.
- [2] S. Liang, *Quantitative Remote Sensing of Land Surfaces*, Science Press, 2009.
- [3] J. Chavea, "Image-based atmospheric corrections—revisited and improved," *Photogrammetric Engineering & Remote Sensing*, vol. 62, no. 9, pp. 1025–1036, 1996.
- [4] M. S. Moran, R. D. Jackson, P. N. Slater, and P. M. Teillet, "Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output," *Remote Sensing of Environment*, vol. 41, no. 2-3, pp. 169–184, 1992.
- [5] E. F. Vermote, D. Tanré, J. L. Deuzé, M. Herman, and J.-J. Morcrette, "Second simulation of the satellite signal in the solar spectrum, 6s: an overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 3, pp. 675–686, 1997.
- [6] E. F. Vermote, D. Tanré, J. L. Deuze et al., *Second Simulation of the Satellite Signal in the Solar Spectrum (6S)*, 6S User Guide Version 6.0, NASA-GSFC, Greenbelt, Md, USA, 1994.
- [7] J. M. van den Bosch and R. E. Alley, "Application of LOWTRAN 7 as an atmospheric correction to airborne visible/infrared imaging spectrometer (AVIRIS) data," in *Proceedings of the 10th Annual International Geoscience and Remote Sensing Symposium (IGARSS '90)*, pp. 175–177, May 1990.
- [8] M. W. Matthew, S. M. Adler-Golden, A. Berk et al., "Status of atmospheric correction using a MODTRAN4-based algorithm," in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, vol. 4049 of *Proceedings of the SPIE*, pp. 199–207, April 2000.
- [9] X. Xiru, *The Physics of Remote Sensing*, Peking University Press, 2006.
- [10] S. Liang and H. Fang, "An improved atmospheric correction algorithm for hyperspectral remotely sensed imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 2, pp. 112–117, 2004.
- [11] Y. J. Kaufman and C. Sendra, "Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery," *International Journal of Remote Sensing*, vol. 9, no. 8, pp. 1357–1381, 1988.
- [12] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," *Applied Optics*, vol. 42, no. 3, pp. 511–525, 2003.
- [13] 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 (CVPR '01)*, vol. 1, pp. 325–332, 2001.
- [14] C. Borel and C. Spencer, "Atmospheric correction of airborne POLDER polarimetric imagery using vectorized 6S," in *Imaging Spectrometry XI*, vol. 6302 of *Proceedings of SPIE*, pp. 63020T-1–63020T-12, August 2006.
- [15] J. R. Schott, *Fundamentals of Polarimetric Remote Sensing*, SPIE Press, 2009.
- [16] C. Ligang, *Study of laboratory calibration of the airborne polarization CCD camera with wide field of view [Ph.D. thesis]*, Graduate University of Chinese Academy of Sciences, 2008.
- [17] J.-B. Jeon, J.-H. Kim, J.-H. Yoon, and K.-S. Hong, "Performance evaluation of teeth image recognition system based on difference image entropy," in *Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology (ICCIIT '08)*, pp. 967–972, November 2008.

Research Article

Land Use Transformation Rule Analysis in Beijing-Tianjin-Tangshan Region Using Remote Sensing and GIS Technology

Shang-min Zhao,¹ Wei-ming Cheng,² Hai-jiang Liu,³ Yao Xia,² Hui-xia Chai,⁴ Yang Song,² Wen-jie Zhang,² and Tian You²

¹Department of Surveying and Mapping, College of Mining Engineering, Taiyuan University of Technology, Taiyuan 030024, China

²State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China

³China National Environmental Monitoring Centre, Beijing 100012, China

⁴Chinese Academy for Environmental Planning, Beijing 100012, China

Correspondence should be addressed to Shang-min Zhao; zhaoshangmin@tyut.edu.cn

Received 8 July 2014; Revised 12 January 2015; Accepted 14 January 2015

Academic Editor: Yongqiang Zhao

Copyright © 2016 Shang-min Zhao 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.

Based on land use classification system, this paper acquires the land use distribution status at 2000, 2005, and 2010 in Beijing-Tianjin-Tangshan Region using remote sensing images, field survey data, images in Google Earth, and visual interpretation methods. Then, the land use transformation rules from 2000 to 2010 are achieved using GIS (geographic information system) technology. The research results shows the following: (1) as to the distribution area of the land use types, dry field has the largest area, followed by forest land, building land, paddy field, water area, grassland, and unused land; (2) from 2000 to 2010, the area of building land has the largest increase, which is mainly transformed from cropland and sea reclamation area; the largest decreased land use type is paddy field, which mainly transforms to dry field and building land; (3) the high increase of building land and decrease of cropland suggest the land use transformation in the quick development process of economy; meanwhile, the total area of forestland and grassland changes little, so the ecological environment does not have apparent deterioration in the 1st decade of the new century.

1. Introduction

As an important environmental element, land is a carrier for people's social and economic activities [1]. Land use means the result after people employing the land and its resources [2, 3], so its distribution status and dynamic change have important impact not only on land management practices, economic, and social developing processes, but also on government policy at regional, national, and even global level [4, 5]. Land use change is an important element of environmental change processes, which has much significance in sustainable development processes [6] so as to become one of the hot research fields.

With the development of science and technology, land use change monitoring is widely conducted by using remote

sensing and geographic information system (GIS) technologies [7]. Remote sensing images are used to acquire the land use status at different phases [8, 9]; based on multitemporal land use data, GIS technology has been proved to be an efficient method to conduct land use dynamic monitoring, transformation rule acquisition, spatial-temporal pattern analysis, and so on [10–12].

As China's capital, Beijing, is a political, cultural, and economic center in China, so it has unique land use characteristics after long-term comprehensive action of natural and humanity factors [13]. Tianjin is the biggest coastal open city in northern China, which is adjoining to Bohai Sea; with fast development of the economy, land use in Tianjin faces severe transformation, especially in coastal area, where large sea area changes to sea reclamation area, such as building

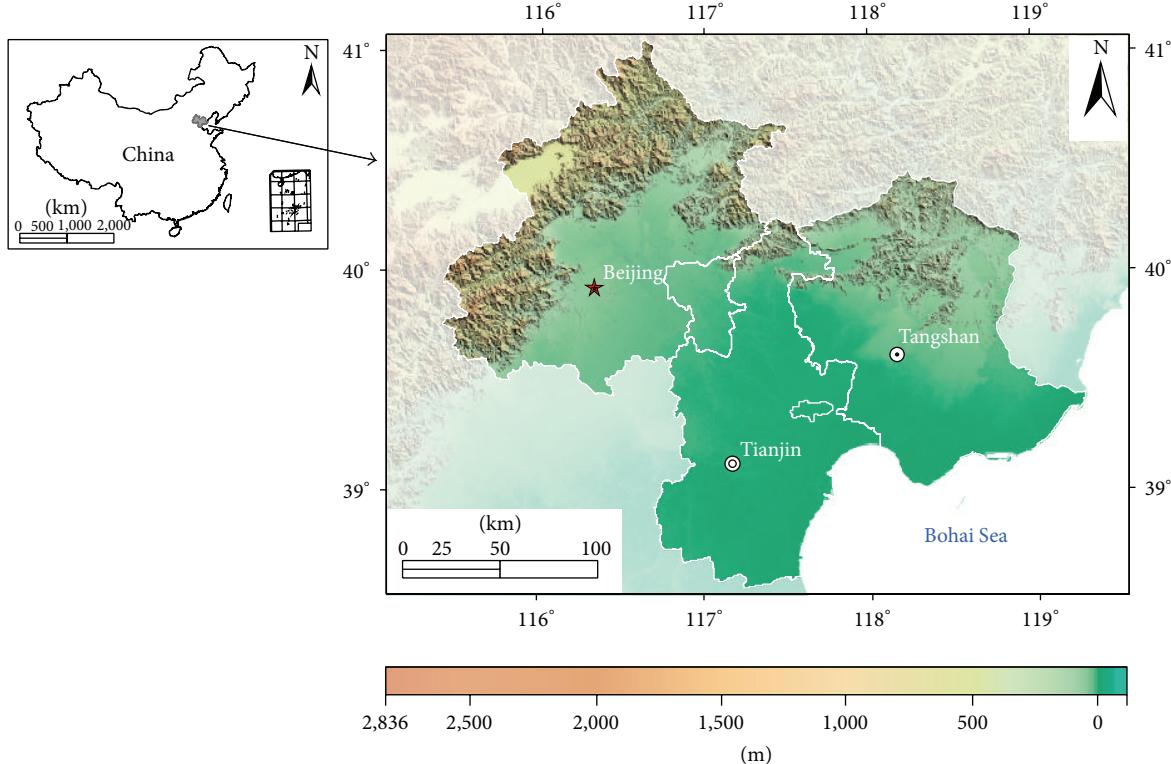


FIGURE 1: Location and relief map of the study area.

land and water area [14]. Tangshan is a central city in Hebei Province, which surrounds the central Bohai gulf region, so it is the strategic place and extremely important corridor for adjoining North China Region and Northeast China Region [15]. Tangshan combined with Beijing and Tianjin forms an important metropolitan circle in China, Beijing-Tianjin-Tangshan metropolitan circle, which is defined as Beijing-Tianjin-Tangshan Region in this research. Compared to other metropolitan circles, Beijing-Tianjin-Tangshan metropolitan circle has relatively lower research degree [16]. The fast development of economy, deep opening-up policy, infrastructure modernization at highest level, and especially the site for 2008 Olympic Games in Beijing make it essential to monitor the transformation rule of land use in this region.

Hence, this research firstly collected and processed remote sensing images in Beijing-Tianjin-Tangshan Region at different phases; then land use distribution status in 2000, 2005, and 2010 were visually interpreted in the study region by using remote sensing images; finally, the land use transformation rules from 2000 to 2010 are acquired and analyzed by using spatial analysis function in ArcGIS software. It has scientific, economic, and environmental significance at some extent for acquiring the land use distribution status and transformation rules from 2000 to 2010 in Beijing-Tianjin-Tangshan Region.

2. Study Area

This research chooses Beijing-Tianjin-Tangshan metropolitan economic circle as the study area, which includes China's capital, Beijing, municipality directly under the Central

Government, Tianjin, and northern part of Hebei Province, Tangshan City. Additionally, in order to avoid hollow area, the study area also includes some counties of Langfang City in Hebei Province. As the geographic boundary of the study area extends from 115.4°E to 119.3°E and from 38.6°N to 41.1°N, the Albers Equal Area projection with two standard parallels of 25°N and 47°N and central meridian of 117°E is chosen as the uniform projection in this research, which considers minimum deformation, area statistics, and consistency to national projection (Figure 1).

The study area locates in the northern part of North China Plain, which is the transition area between North China Region and Northeast China Region. With Yanshan Mountains to the north, Taihangshan Mountains to the west, and Bohai Sea to the east, the geographical location of the study area has much significance (Figure 1).

The main distribution of coastal plain and semiwet monsoon climate make the study area not only an important grain production base, but also a fast developing metropolitan economic circle. The fast development of the economy, establishment of the national central cities, and the Beijing Olympic Games make the land use distribution here transform severely. So it is important to conduct land use transformation rule research in the study area, which is significant to understand the ecological and environmental change in this period.

3. Material and Methodology

The main data sources used in this research are remote sensing images in 2000, 2005, and 2010. Additionally, field

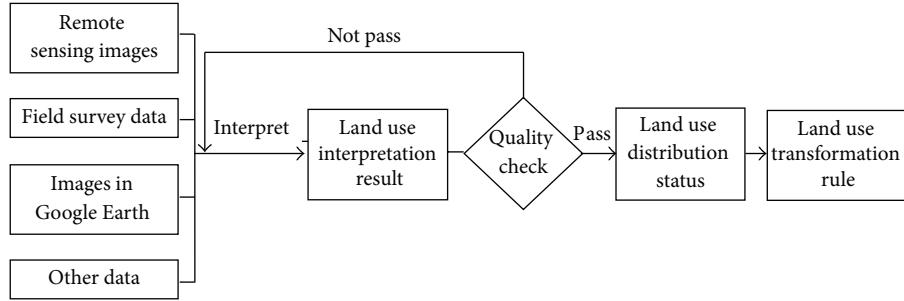


FIGURE 2: Work flow of this research.

survey data, topographic maps, images in the Google Earth, vegetation data, and geomorphologic data, and so on are also referenced.

The main data sources and the methodology in this research are specifically illustrated in the following.

3.1. Remote Sensing Images. The main remote sensing images are Landsat TM images in 2000, 2005, and 2010. The images are mostly collected in summer season (especially August) to decrease the weather influences. The path-rows of the images in the study area are 121-031, 121-033, 122-032, 122-033, 123-032, 123-033, and 124-032, respectively.

The remote sensing images in 2000 are processed through data selection, downloading, band selection and stack, atmospherically correction, geometric correction based on topographic maps at 1:50,000 scales, mosaic, projection, clip, and so on [17]. The processes to the remote sensing images in 2005 and 2010 are similar to these in 2000, except for that the geometric corrections are based on the processed images in 2000.

3.2. Methodology. Many researchers have tried to acquire land use change status by using remote sensing images [5, 7, 17–19], so several techniques for achieving dynamic monitoring of land use have been proposed, applied, and evaluated, such as principle component analysis [20, 21], comparison among two or more images at different phases and so on [22–25]. These techniques mainly aim at the automatic extraction of the land use information, as to the extraction quality, the traditional visually interpretation method combined with field survey data has the highest quality and accuracy, despite of large labour, time, and money input. Hence, the visual interpretation method is adopted in this research.

The remote sensing images are the fundamental data for acquiring the land use distribution status at different phases. The processed images are mainly Landsat TM images, which have 7 bands, and the usual band combination in the land use interpretation is 4-3-2 presented by red, green, and blue colour; sometimes, other band combinations are also used, such as 7-4-2 and 5-4-3. Meanwhile, other auxiliary data are also used in land use visual interpretation, such as the images in the Google Earth with higher spatial resolution, the field survey data with field pictures, land use types and locations, the Chinese geomorphologic data, and the vegetation data at 1:1,000,000 scales.

Under the ArcGIS software, the land use distribution status is visually interpreted by using multidata sources, especially the remote sensing images. According to the interpretation results of the land use distribution status, the land use transformation rules from 2000 to 2010 are acquired and analyzed. Thus, the work flow of this research is shown in the Figure 2.

The accuracy of the interpretation results is fundamental in this research, which also affects the quality of the land use transformation rules. In 2008, 2009, and 2010, about 400 sample points are checked. Through field survey, the locations, pictures, and land use types are acquired for these sample points. Based on these points, the accuracy of the interpretation results is estimated for higher than 85%. Finally, the interpreted results are revised according to the filed survey results of these sample points.

4. Analysis to the Distribution Statuses of Land Use in 2000, 2005, and 2010

Based on the visual interpretation results for the remote sensing images, the land use distribution statuses are acquired in 2000, 2005, and 2010 in the study area, which are analyzed, respectively, as in the following.

4.1. Analysis to the Distribution Status of Land Use in 2000. Based on the data sources and the methodology, land use distribution status in 2000 is interpreted in the study area, which is shown in Figure 3(a).

In Figure 3, the legend gives the land use types represented by different numbers. The representations keep consistency in this research. Through area statistics to every land use type, the results are shown in Table 1.

Figure 3(a) and Table 1 show the land use distribution status in 2000 in the study area: dry field is the most widely distributed type, which mainly distributes in the central and southern flat part; then forestland is distributing mainly in the western and northern mountainous and hill region; the next is building land, including mainly residential areas at all levels and coastal areas; paddy field mainly distributes in the eastern part, especially the northeastern part; other land use types, such as water area, grassland, and unused land, are the least distributed types; water area mainly distributes in lakes, ponds, and channels; as to grassland, it mainly distributes in the northern hill areas.

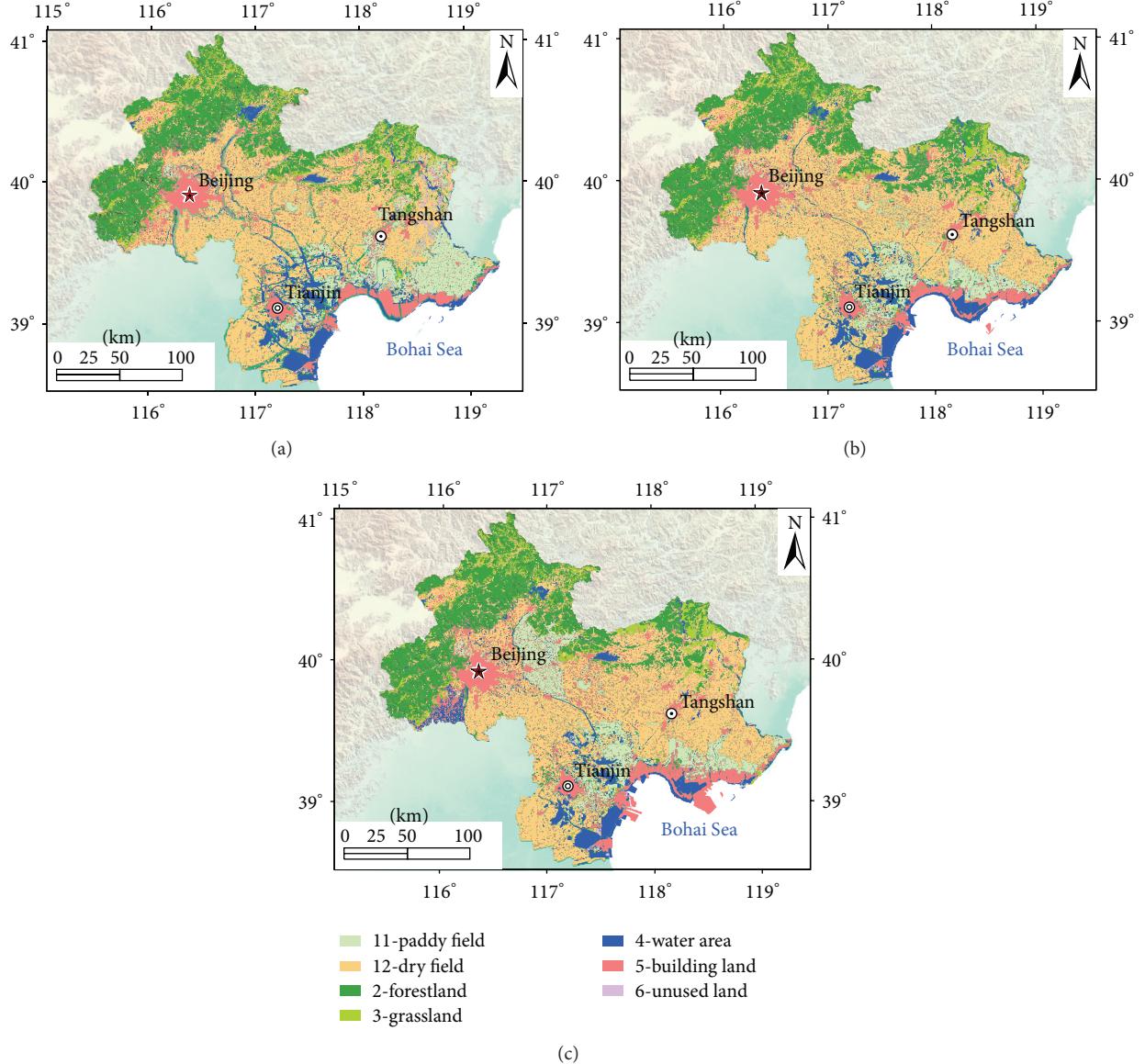


FIGURE 3: Land use distribution status in different phases ((a) 2000; (b) 2005; (c) 2010).

TABLE 1: Area statistics of land use types in 2000, 2005, and 2010 ($\times 10^4 \text{ km}^2$).

Time	Land cover type	Paddy field	Dry field	Forestland	Grassland	Water area	Building land	Unused land	Total
2000	Area	4403.7	16280.7	9251.6	2627.8	3169.9	6667.8	500.0	42901.5
	Ratio	10.3%	37.9%	21.6%	6.1%	7.4%	15.5%	1.2%	100%
2005	Area	2744.7	17796.4	10001.3	2287.3	3484.4	6765.2	160.5	43239.8
	Ratio	6.3%	41.2%	23.1%	5.3%	8.1%	15.6%	0.4%	100%
2010	Area	3754.8	16492.6	9499.8	2384.4	3626.4	7634.5	129.1	43521.6
	Ratio	8.6%	37.9%	21.8%	5.5%	8.3%	17.5%	0.3%	100%

4.2. Analysis to the Distribution Status of Land Use in 2005. Through visual interpretation and revision, land use distribution status in 2005 can be acquired by using remote sensing images in 2005 and referencing the land use distribution

results in 2000. After passing quality check, the land use distribution results in 2005 in the study area are shown in Figure 3(b). The area and ratio of the land use types in 2005 are computed, which are shown in Table 1.

TABLE 2: Land use transformation area from 2000 to 2005 ($\times 10^4 \text{ km}^2$).

2000	2005								
	Void	Paddy field	Dry field	Forestland	Grassland	Water area	Building land	Unused land	Sum
Void	0.0	0.0	0.4	0.0	0.0	300.0	39.6	0.0	340.1
Paddy field	0.0	2136.3	1856.5	57.6	39.5	76.3	200.1	38.7	4405.0
Dry field	0.0	312.3	14026.7	682.0	261.3	235.6	755.7	7.4	16280.9
Forestland	4.0	6.6	401.8	8263.2	463.6	34.4	76.1	1.9	9251.6
Grassland	0.0	40.5	235.9	828.3	1445.3	30.4	43.9	3.6	2627.8
Water area	0.0	125.3	271.7	70.1	28.5	2474.6	185.9	13.7	3170.0
Building land	0.3	81.6	692.6	80.5	42.1	324.9	5441.7	4.9	6668.7
Unused land	0.0	42.1	310.9	19.5	6.9	8.1	22.1	90.2	500.0
Sum	4.3	2744.7	17796.4	10001.3	2287.3	3484.4	6765.2	160.5	43244.1

Figure 3(b) and Table 1 show the distribution status of land use in 2005: the whole land use distribution status in 2005 is similar to it in 2000. As to the specifics, the total area has increased 338.3 km^2 , which is mainly due to the added coastal land adjoining to the Bohai Sea; the areas of dry field and forestland increase remarkably; inversely, the areas of paddy field, unused land, and grassland decrease evidently; as to other land use types, the distribution of building land and water area does not have apparent change.

4.3. Analysis to the Distribution Status of Land Use in 2010. Referencing the visually interpreted land use distribution results in 2000 and 2005, the land use distribution status in 2010 in the study area is visually interpreted based on remote sensing images and auxiliary data, which is shown in Figure 3(c). Through computing the area of every land use types in 2010, the statistics results are shown in Table 1.

From Figure 3 and Table 1 we can see that the whole land use distribution status in 2010 is similar to them in 2000 and 2005. Compared to the land use distribution statuses in 2000 and 2005, the land use distribution status in 2010 has the following changes: the total area continuously increases from 2000 to 2010, which is mainly due to the coastal construction; the area of paddy field increases from 2005 to 2010 but decreases from 2000 to 2010; the area of dry field decreases from 2005 to 2010 but has a little increase from 2000 to 2010; the area of forestland decreases from 2005 to 2010 but increases a little from 2000 to 2010; the area of grassland has evident decrease from 2000 to 2010; the area of water area consistently increases from 2000 to 2010; the area of building land also consistently increases from 2000 to 2010, especially from 2005 to 2010; the area of unused land consistently decreases from 2000 to 2010, which shows the rapid change of land use distribution status.

5. Analysis to the Transformation Rules of Land Use from 2000 to 2010

Land use transformation is due to that one land use type in this phase may transform to other land use types in the later phase. So land use transformation status is mainly acquired by calculating the land use transformation areas of every land use type during two phases. Usually, the transformations

between every two land use types are mutual-transformative, such that the paddy field may transform to the dry field and vice versa during two phases. In order to eliminate the mutual-transformative situations, the absolute transformation status is presented and defined, which is acquired by computing the subtraction of the roll-out and roll-in area for every land use type during two phases.

Hence, based on the land use distribution statuses in 2000, 2005, and 2010 in the study area, the land use transformation statuses from 2000 to 2005, from 2005 to 2010, and from 2000 to 2010 are computed and analyzed based on GIS spatial analysis function; moreover, the absolute transformation status is acquired and analyzed from 2000 to 2010 for every land use type.

5.1. Analysis to the Transformation Status of Land Use from 2000 to 2005. The land use transformation status from 2000 to 2005 can be computed by overlapping the land use distribution results of 2000 and 2005. As the distribution areas for 2000 and 2005 do not completely coincide, union function is used instead of clipping function. The processed land use transformation distribution status from 2000 to 2005 in the study area is shown as in Figure 4(a).

In Figure 4(a), the number before the short line represents the land use type in 2000, the early phase; the number after the short line represents the land use type in 2005, the latter phase. The land use types represented by these numbers can be retrieved in Figure 3. For the number “0,” it represents that the land use type does not exist in the corresponding phase but exists in other phase; this region mainly exists in Bohai Sea and is defined as sea reclamation area.

Making an area statistic to the land use transformation types, the computed results are shown in Table 2.

In Table 2, the right column gives the areas of land use types in 2000; the lowest line shows the areas of land use types in 2005; middle part shows the land use transformation status; the void type corresponds to the “0” number in Figure 4.

Figure 4(a) and Table 2 show that, from 2000 to 2005, paddy field mainly transforms to dry field, so the area of paddy field decreases and that of dry field increases; the forestland mainly originates from grassland and dry land, so the area of the forestland increases and that of the grassland

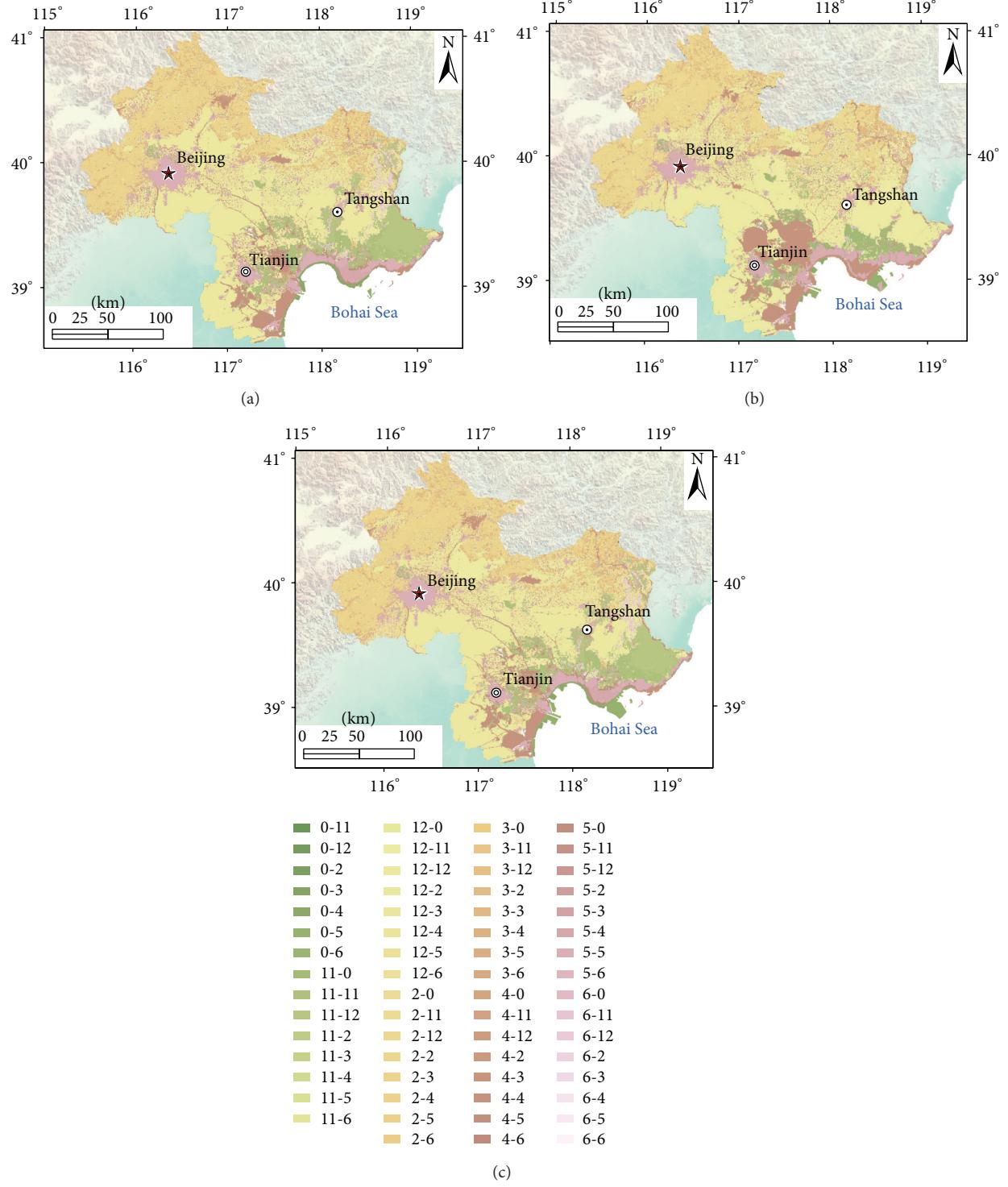


FIGURE 4: Land use transformation status from 2000 to 2010 ((a) from 2000 to 2005; (b) from 2005 to 2010; (c) from 2000 to 2010).

decreases. The unused land mainly transforms to the dry field. Undoubtedly, the transformed dry field from unused land has low quality.

5.2. Analysis to the Transformation Status of Land Use from 2005 to 2010.

Processing the land use distribution statuses

in both 2005 and 2010 by union function, the land use transformation status from 2005 to 2010 can be achieved as shown in Figure 4(b). The explanation to Figure 4(b) can reference them for Figure 4(a). Making an area statistics to the land use transformation status in Figure 4(b), the results are shown as in Table 3.

TABLE 3: Land use transformation area from 2005 to 2010 ($\times 10^4 \text{ km}^2$).

2005	2010								
	Void	Paddy field	Dry field	Forestland	Grassland	Water area	Building land	Unused land	Sum
Void	0.0	0.0	0.0	4.0	2.6	63.7	247.0	0.0	317.3
Paddy field	0.0	2340.9	176.7	5.5	3.5	13.7	203.4	0.9	2744.7
Dry field	0.0	1249.4	14952.0	275.1	130.6	500.8	678.1	10.3	17796.4
Forestland	0.0	10.5	600.3	8975.0	297.2	38.4	72.8	7.1	10001.3
Grassland	0.0	4.1	198.1	189.1	1856.8	10.0	25.2	4.1	2287.3
Water area	32.9	29.8	134.3	17.2	46.5	2958.0	250.8	14.9	3484.4
Building land	0.1	96.4	418.6	28.3	39.6	36.9	6139.0	6.3	6765.2
Unused land	0.0	24.1	12.7	5.7	7.5	5.1	20.0	85.5	160.5
Sum	33.0	3755.1	16492.8	9499.8	2384.4	3626.4	7636.4	129.1	43557.1

TABLE 4: Land use transformation area from 2000 to 2010 ($\times 10^4 \text{ km}^2$).

2000	2010								
	Void	Paddy field	Dry field	Forestland	Grassland	Water area	Building land	Unused land	Sum
Void	0.0	0.0	0.1	0.0	2.6	230.0	391.6	0.0	624.3
Paddy field	0.0	2057.7	1790.9	32.7	28.8	66.9	423.8	3.0	4403.8
Dry field	0.0	1337.4	12762.1	340.9	220.7	611.4	994.0	14.3	16280.8
Forestland	0.0	18.9	308.3	8292.0	510.7	37.7	76.6	7.3	9251.5
Grassland	0.0	34.0	289.6	699.9	1506.2	33.2	59.4	5.5	2627.8
Water area	3.5	120.2	285.4	57.4	48.8	2332.3	304.6	17.8	3170.0
Building land	0.6	162.8	734.4	65.2	60.7	304.9	5330.7	8.5	6667.8
Unused land	0.0	23.9	321.9	11.9	5.8	9.9	53.9	72.6	499.9
Sum	4.1	3754.9	16492.7	9500.0	2384.3	3626.3	7634.6	129.0	43526.0

The explanation to Table 3 can reference them in Table 2. Figure 4(b) and Table 3 show that, from 2005 to 2010, paddy field and dry field have inverse change tendencies compared to them from 2000 to 2005; large area of dry field transforms to paddy field which leads to increase of paddy field and decrease of dry field; forestland mainly transforms to grassland and dry land, so the area of the forestland decreases, but the grassland increases. The area of building land has evident increase, which is mainly transformed from cropland and sea reclamation area.

5.3. Analysis to the Transformation Status of Land Use from 2000 to 2010. Overlapping land use distribution statuses in both 2000 and 2010 by union function, land use transformation status in whole period from 2000 to 2010 is computed as shown in Figure 4(c). The explanation to Figure 4(c) can reference them for Figure 4(a). Making an area statistics to the land use transformation types in Figure 4(c), the results are shown as in Table 4.

The explanation to Table 4 can reference them in Table 2. Figure 4(c) and Table 5 show that, from 2000 to 2010, paddy field mainly transforms to dry field and building land; dry field mainly transforms to paddy field, building land, and grassland; forestland mainly transforms to grassland and dry field; grassland mainly transforms to forestland and dry field; water area mainly transforms to building land and dry field; building land mainly transforms to dry field and water area;

the area of unused land has evident decrease, which mainly transforms to dry field and building land.

5.4. Analysis to the Absolute Transformation of Land Use from 2000 to 2010. The above three subsections analyze transformation statuses in three periods which are from 2000 to 2005, from 2005 to 2010, and from 2000 to 2010. In the transformation statuses analysis, land use type transformation is mutual, so it is difficult to find the absolute transformation area between every two land use types. In view of this, the absolute transform area is computed to represent the subtraction area between the roll-out area and roll-in area for every two land use types, which is shown in Table 5.

In Table 5, the right column represents the areas of the land use types which transform to other land use types from 2000 to 2010; the lower line represents the areas of the land use types which are transformed from other land use types from 2000 to 2010; the middle part represents the absolute transformation areas between every two land use types.

Table 5 shows the absolute transformation status from 2000 to 2010 in the study area: the area of paddy field has the largest decrease, about $648.9 \times 10^4 \text{ km}^2$, which mainly transforms to dry field and the building land; on the contrary, building land has the largest increase, about $996.8 \times 10^4 \text{ km}^2$, which is mainly transformed from cropland (both paddy field and dry field); dry field has a little increase, which is because

TABLE 5: Land use absolute transformation area from 2000 to 2010 ($\times 10^4 \text{ km}^2$).

2000	2010								
	Void	Paddy field	Dry field	Forestland	Grassland	Water area	Building land	Unused land	Sum
Void	0	0	0.1	0	2.6	226.5	391.0	0	620.2
Paddy field	0	0	453.5	13.8	-5.2	-53.3	261.0	-20.9	648.9
Dry field	-0.1	-453.5	0	32.6	-68.9	326	259.6	-307.6	-211.9
Forestland	0	-13.8	-32.6	0	-189.2	-19.7	11.4	-4.6	-248.5
Grassland	-2.6	5.2	68.9	189.2	0	-15.6	-1.3	-0.3	243.5
Water area	-226.5	53.3	-326	19.7	15.6	0	-0.3	7.9	-456.3
Building land	-391.0	-261.0	-259.6	-11.4	1.3	0.3	0	-45.4	-966.8
Unused land	0	20.9	307.6	4.6	0.3	-7.9	45.4	0	370.9
Sum	-620.2	-648.9	-211.9	248.5	-243.5	456.3	966.8	370.9	0

lots of paddy field and unused land transform to dry field, but part of dry field transforms to building land; the increase of forestland and decrease of grassland are mainly due to that part of grassland transforms to forestland, about $189.2 \times 10^4 \text{ km}^2$; water area has an evident increase which mainly originates from the sea reclamation area; for the unused land, it has large decrease which mainly transforms to dry field; large area of the sea reclamation area is utilized, which mainly transforms to the building land and water area. As part of water area also which belongs to the building land, such as ponds in the parks, the area of the building land increases even larger; the large increase of building land and decrease of cropland show the rapid change of land use in the economy development process; meanwhile, large area of paddy field transforms to dry field and building land, which represents the deterioration situation of the cropland in both quality and quantity aspects. The change of forestland and grassland shows that the ecological environment does not change worse which may result from large campaigns in the study area, such as the Olympic Games in 2008.

6. Discussion

This research has some breakthroughs in the land use transformation rule analysis from 2000 to 2010 in the study area, but some improvements are also needed. So the innovations and prospects in this research are discussed in the following.

6.1. Innovations

(1) *Significant Study Area.* Compared to other researches, this research selects the Beijing-Tianjin-Tangshan metropolitan circle in China as the study area, which has much larger area than the study regions in other researches [10, 26, 27]. Moreover, the deep opening-up and reform policy, fast development of the economy, and hosting site for Olympic Games in 2008 make the study area in this research have much significance and necessary to make a land use dynamic research.

(2) *Land Use Distribution Acquisition Method.* This research selects traditional remote sensing method, visually interpretation combined with field survey data to acquire the land use distribution data. Although this method consumes large

labour, time, and money, the acquired data have the highest accuracy at present situation [28]. Based on the sample points, the accuracy is estimated for higher than 85%, which guarantees the quality and availability of the analyzed results.

(3) *Achievement of the Land Use Transformation Rules.* In most land use dynamic monitoring researches, land use distribution change matrix is usually adopted [7, 29], but the change matrix only gives the mutual transition status between every two land use types, and the absolute transition between every two land use types is seldom presented and analyzed. In this research, the achievement of the absolute transformation is helpful in understanding the real transformation status of the land use types.

6.2. Research Prospects

(1) *Improvement of the Land Use Data.* Although the authors had done their best to make the quality and accuracy of the land use data as high as possible, the data can also be improved in some aspects. For example, the land covered by water is classified as water area; sometimes, the classified water area should be changed to building land, such as pond, small lakes located in the building land. The improvement of the land use data will bring more reasonable results. Deep understanding of the land use situation and class system is important development of this research, which can be used to acquire better data.

(2) *Further Analysis to the Land Use Data.* The basic land use distribution data at three phases are acquired in this research, and the land use transformation rules are computed and analyzed in the study region from 2000 to 2010. However, the data can be studied much deeper in the future, such as acquiring the driving factors of the land use transformation [30], analyzing the spatial-temporal patterns of the land use distribution [26, 31], and so on.

7. Conclusion

From this research, the following conclusions can be drawn.

(1) In the study area, the distribution status of the land use types shows that dry field is the most widely

distributed type and mainly distributes in central and southern part; then forestland is distributing mainly in western and northern mountainous and hill regions; the next is building land, including mainly residential areas at all levels and coastal areas; paddy field mainly distributes in the eastern part, especially the northeastern part; other land use types, including water area, grassland, and unused land, distribute sparsely.

(2) From 2000 to 2005 and then to 2010, the area changes of land use types are as follows: dry field increases firstly and then decreases, whose area has a little increase in the whole period; forestland increases apparently firstly and then decreases, whose area has a small increase in the whole period; building land keeps increasing, especially from 2005 to 2010; paddy field decreases evidently firstly and then increases, whose area shows a decrease tendency in the whole period; grassland and unused land decrease remarkably; as to water area, it shows an increasing tendency, especially from 2005 to 2010.

(3) The land use transformation rule in the whole period is as follows: building land has the largest increase, which is mainly transformed from sea reclamation area, paddy filed, and dry field; conversely, paddy field has maximum decrease, which mainly transforms to dry filed and building land. The transformation rule of the land use types shows the rapid development of the economy and the deterioration situation of the cropland, in both quality and quantity aspects.

Conflict of Interests

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

Acknowledgments

This research is under the auspices of the National Natural Science Foundation of China (41301469, 41171332), the Open Foundation of the LREIS, the National Science Technology Support Plan Project (2012BAH28B01-03), and the Qualified Personnel Foundation of Taiyuan University of Technology (QPFT) (tyut-rc201221a). The authors would like to express their sincere gratitude to the anonymous editors and referees.

References

- [1] C. Lu, Q. Yang, Y. Tian, F. Wen, Y. Hu, and P. Wan, "Study on the driving force of land use based on the primary factor analysis," *Journal of Anhui Agricultural Sciences*, vol. 34, no. 21, pp. 5627–5628, 2006.
- [2] W. B. Meyer, "Past and present land-use and land-cover in the USA," *Consequences*, vol. 1, pp. 25–33, 1995.
- [3] G. S. Kiran and U. B. Joshi, "Estimation of variables explaining urbanization concomitant with land-use change: a spatial approach," *International Journal of Remote Sensing*, vol. 34, no. 3, pp. 824–847, 2013.
- [4] D. S. Ojima, K. A. Galvin, and B. L. I. I. Turner, "The global impact of land-use change," *BioScience*, vol. 44, no. 5, pp. 300–304, 1994.
- [5] R. S. Dwivedi, K. Sreenivas, and K. V. Ramana, "Land-use/land-cover change analysis in part of Ethiopia using Landsat Thematic Mapper data," *International Journal of Remote Sensing*, vol. 26, no. 7, pp. 1285–1287, 2005.
- [6] X. Li, "A review of the international researches on land use/land cover change," *Acta Geographica Sinica*, vol. 51, no. 6, pp. 553–557, 1996.
- [7] X. Zhang, T. Kang, H. Wang, and Y. Sun, "Analysis on spatial structure of landuse change based on remote sensing and geographical information system," *International Journal of Applied Earth Observation and Geoinformation*, vol. 12, supplement 2, pp. S145–S150, 2010.
- [8] J. Heo and T. W. FitzHugh, "A standardized radiometric normalization method for change detection using remotely sensed imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 66, no. 2, pp. 173–181, 2000.
- [9] J. Rogan, J. Miller, D. Stow, J. Franklin, L. Levien, and C. Fischer, "Land-cover change monitoring with classification trees using Landsat TM and ancillary data," *Photogrammetric Engineering and Remote Sensing*, vol. 69, no. 7, pp. 793–804, 2003.
- [10] J. G. Masek, F. E. Lindsay, and S. N. Goward, "Dynamics of urban growth in the Washington DC metropolitan area, 1973–1996, from Landsat observations," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3473–3486, 2000.
- [11] M. Herold, N. C. Goldstein, and K. C. Clarke, "The spatiotemporal form of urban growth: measurement, analysis and modeling," *Remote Sensing of Environment*, vol. 86, no. 3, pp. 286–302, 2003.
- [12] D. Maktav and F. S. Erbek, "Analysis of urban growth using multi-temporal satellite data in Istanbul, Turkey," *International Journal of Remote Sensing*, vol. 26, no. 4, pp. 797–810, 2005.
- [13] Y. He, Y. Chen, Z. Li, Y. Yao, and P. Tang, "Analysis on spatial structural characteristics of land use of Beijing City," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 26, no. 2, pp. 313–318, 2010.
- [14] N. Xu, W. Meng, F. Zhai, and H. Li, "Effects of land use change on ecosystem service value in Tianjin Binhai new area," *Urban Environment and Urban Ecology*, vol. 26, no. 1, pp. 5–8, 2013.
- [15] X. Zhang, C. Chen, and K. Dong, "Analysis of temporal and spatial characteristics of land use pattern in Tangshan Coastal Zone from 1956 to 2005 based on RS and GIS," *Acta Agriculturae Boreali-Occidentalis Sinica*, vol. 22, no. 2, pp. 204–208, 2013.
- [16] D. Meng, X. Li, H. Xu, and H. Gong, "The spatial expansion of construction land-use in Beijing-Tianjin-Hebei metropolis circle," *Journal of Geo-Information Science*, vol. 15, no. 2, pp. 289–296, 2013.
- [17] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 296–308, 2015.
- [18] T. Fung, "An assessment of TM imagery for land-cover change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 4, pp. 681–684, 1990.
- [19] E. F. Lambin and A. H. Strahler, "Change-vector analysis in multitemporal space: a tool to detect and categorize land-cover change processes using high temporal-resolution satellite data," *Remote Sensing of Environment*, vol. 48, no. 2, pp. 231–244, 1994.
- [20] T. Fung and E. Ledrew, "Application of principal components analysis to change detection," *Photogrammetric Engineering and Remote Sensing*, vol. 53, no. 12, pp. 1649–1658, 1987.

- [21] J. R. Eastman and M. Fulk, "Long sequence time series evaluation using standardized principal components," *Photogrammetric Engineering & Remote Sensing*, vol. 59, no. 6, pp. 991–996, 1993.
- [22] K. Green, D. Kempka, and L. Lackey, "Using remote sensing to detect and monitor land-cover and land-use change," *Photogrammetric Engineering and Remote Sensing*, vol. 60, no. 3, pp. 331–337, 1994.
- [23] O. E. Frihy, K. M. Dewidar, S. M. Nasr, and M. M. El Raey, "Change detection of the northeastern Nile Delta of Egypt: shoreline changes, spit evolution, margin changes of Manzala lagoon and its islands," *International Journal of Remote Sensing*, vol. 19, no. 10, pp. 1901–1912, 1998.
- [24] K. M. Dewidar, "Landfill detection in Hurghada, North Red Sea, Egypt, using thematic mapper images," *International Journal of Remote Sensing*, vol. 23, no. 5, pp. 939–948, 2002.
- [25] K. M. Dewidar, "Detection of land use/land cover changes for the northern part of the Nile delta (Burullus region), Egypt," *International Journal of Remote Sensing*, vol. 25, no. 20, pp. 4079–4089, 2004.
- [26] D. Wang, J. Gong, L. Chen, L. Zhang, Y. Song, and Y. Yue, "Spatio-temporal pattern analysis of land use/cover change trajectories in Xihe watershed," *International Journal of Applied Earth Observation and Geoinformation*, vol. 14, no. 1, pp. 12–21, 2012.
- [27] P. K. Mallupattu and J. R. Sreenivasula Reddy, "Analysis of land use/land cover changes using remote sensing data and GIS at an Urban Area, Tirupati, India," *The Scientific World Journal*, vol. 2013, Article ID 268623, 6 pages, 2013.
- [28] S. Gao, Y. Cheng, and Y. Zhao, "Unsupervised change detection of satellite images using low rank matrix completion," *Optics Letters*, vol. 38, no. 23, pp. 5146–5149, 2013.
- [29] Z. Li, X. Li, Y. Wang, A. Ma, and J. Wang, "Land-use change analysis in Yulin prefecture, northwestern China using remote sensing and GIS," *International Journal of Remote Sensing*, vol. 25, no. 24, pp. 5691–5703, 2004.
- [30] D. Wang, J. Gong, L. Chenc, L. Zhang, Y. Song, and Y. Yue, "Comparative analysis of land use/cover change trajectories and their driving: forces in two small watersheds in the western Loess Plateau of China," *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, no. 1, pp. 241–252, 2012.
- [31] D. Ruelland, F. Levavasseur, and A. Tribotté, "Patterns and dynamics of land-cover changes since the 1960s over three experimental areas in Mali," *International Journal of Applied Earth Observation and Geoinformation*, vol. 12, supplement 1, pp. S11–S17, 2010.

Research Article

Undersampled Hyperspectral Image Reconstruction Based on Surfacelet Transform

Lei Liu,¹ Jingwen Yan,² Di Guo,³ Yunsong Liu,⁴ and Xiaobo Qu⁴

¹Department of Mathematics, Shantou University, Shantou 515063, China

²Guangdong Provincial Key Laboratory of Digital Signal and Image Processing, Shantou University, Shantou 515063, China

³School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361005, China

⁴Department of Electronic Science, Xiamen University, Xiamen 361005, China

Correspondence should be addressed to Xiaobo Qu; quxiaobo@xmu.edu.cn

Received 7 July 2014; Revised 23 September 2014; Accepted 24 September 2014

Academic Editor: Yongqiang Zhao

Copyright © 2015 Lei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hyperspectral imaging is a crucial technique for military and environmental monitoring. However, limited equipment hardware resources severely affect the transmission and storage of a huge amount of data for hyperspectral images. This limitation has the potentials to be solved by compressive sensing (CS), which allows reconstructing images from undersampled measurements with low error. Sparsity and incoherence are two essential requirements for CS. In this paper, we introduce surfacelet, a directional multiresolution transform for 3D data, to sparsify the hyperspectral images. Besides, a Gram-Schmidt orthogonalization is used in CS random encoding matrix, two-dimensional and three-dimensional orthogonal CS random encoding matrixes and a patch-based CS encoding scheme are designed. The proposed surfacelet-based hyperspectral images reconstruction problem is solved by a fast iterative shrinkage-thresholding algorithm. Experiments demonstrate that reconstruction of spectral lines and spatial images is significantly improved using the proposed method than using conventional three-dimensional wavelets, and growing randomness of encoding matrix can further improve the quality of hyperspectral data. Patch-based CS encoding strategy can be used to deal with large data because data in different patches can be independently sampled.

1. Introduction

Typical hyperspectral imaging (HSI) is acquired on satellites and aerospace probes and then transmitted to grounds. The imaging spectrometer can provide tens to hundreds of narrow-band spectral information for each spatial location. The huge amounts of data but scarce equipment hardware resources on satellites and aerospace severely limit the transmission and storage of hyperspectral images [1, 2]. However, the ground receiving side holds very strong processing capability. If we transfer the system complexity from the satellites and aerospace probes to the ground, it is expected to potentially solve the limitation in traditional high spectral sampling.

Traditionally, super-resolution reconstruction is used to improve the spatial resolution of HSI [3, 4] while compression can improve the transmission efficiency [5–8]. We acquire “all” data and then “throw away” most of it in transmission

process. Can we just directly measure the part that we need to save the cost of storage? Compressed sensing (CS) has been proposed to solve these contradictions in remote sensing. The theory of CS shows that a sparse signal can be recovered from a relatively small number of linear measurements [9, 10]. The difference between the conventional method and CS based transmission method is shown in Figure 1.

CS has been widely used in medical imaging [11, 12] and wireless sensor network [13, 14] to reduce the sampled data in recent years which has been applied in remote sensing. Duarte et al. [15] proposed a single-pixel imaging method. Ma [2] applied CS in single frame imaging. Aravind et al. [16] used ten spectral bands to compare orthogonal matching pursuit with simultaneous orthogonal matching pursuit in reconstruction. Fowler [17] presented compressive-projection PCA to effectively shift the computational burden of PCA from the encoder to the decoder.

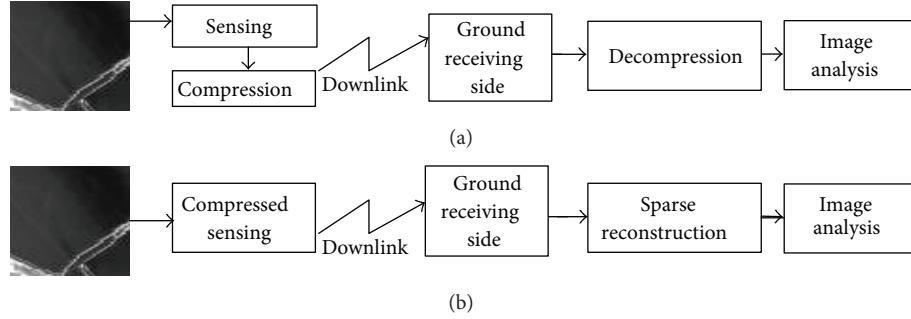


FIGURE 1: Conventional HSI transmission and CS HSI transmission. (a) Conventional sampling, compression, and decompression. (b) Compressed sensing and sparse reconstruction.

According to the CS theory, the reconstruction error is bounded by the sparse approximation error [18]. For a signal $\mathbf{x} \in \mathbb{R}^N$, let $\tilde{\mathbf{x}}$ be the solution to obey CS conditions, Ψ^T a forward transform, and $(\Psi^T \mathbf{x})_s$ the sparse approximation of \mathbf{x} in transform domain by retaining only the s -largest entries of $\Psi^T \mathbf{x}$; then, the reconstruction error is

$$\|\tilde{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 \varepsilon + \frac{C_1}{\sqrt{S}} \|\Psi^T \mathbf{x} - (\Psi^T \mathbf{x})_s\|_1, \quad (1)$$

where C_0 , C_1 are small positive constants and ε is noise power [19]. Equation (1) implies that a sparser representation will reduce the reconstruction error. An optimal sparsifying transform is always important for sparse image reconstruction to reduce the reconstruction error.

In this study, we introduce the surfacelet transform (ST) to sparsely represent hyperspectral images by making use of the spatial and spectral information. The advantage of ST over wavelet on sparse HSI data reconstruction is demonstrated, where the data are reconstructed by using a fast iterative shrinkage-thresholding algorithm [20]. To further improve the reconstruction performance, a 3-dimensional (3D) random encoding is designed and the Gram-Schmidt orthogonalization is adopted. Finally, a patch-based CS encoding scheme is designed to deal with large size data.

The remainder of this paper is organized as follows. First, ST-based compressive sensing HSI reconstruction method is introduced in Section 2. The experimental results are presented in Section 3. Finally, conclusions are given in Section 4.

2. ST-Based Compressive Sensing HSI Reconstruction

The CS theory comes up with two principles: sparsity, which asserts that the reconstructed signal is sparse with a transform Ψ , and incoherence, which requires that the encoding matrix Φ is incoherent with Ψ [10].

2.1. Sparse Representation of HSI Using Surfacelet Transform. Wavelet is commonly used as a typical sparse transform for HSI [21–24]. However, wavelet sometimes fails in sparsely representing HSI because it has only 7 directions [25–27].

Surfacelet transform (ST), proposed by Lu and Do [25], can efficiently capture the surface intrinsic geometrical structure within N -dimensional signals [25]. It offers 3×2^L directional subbands with decomposition level L by combining the multiscale pyramid with the 3-dimensional directional filter banks (3D-DFB). Thus, surfacelet with more directions may help reducing the blocky artifacts caused by orthogonal wavelets [24, 25].

ST is a multiscale version of the 3D-DFB. The input signal $x[n]$ first goes through the 3D hourglass filter $P_{0,0}^{0,0}(\omega)$, which is a three-channel undecimated filter bank. One branch of the three-channel structure of 3D-DFB is given by Figure 2. The output $y[n]$ is then fed into a 2D filter bank, denoted by $IRC_{12}^{(l_2)}$, which operates on the (n_1, n_2) planes. The tree-structured filter bank $IRC_{12}^{(l_2)}$ produces 2^{l_2} output subbands, denoted by $z_i[n]$ for $0 \leq i \leq 2^{l_2}$. Each output is then fed into another 2D filter bank $IRC_{13}^{(l_3)}$ operating on the (n_1, n_3) planes. In the end, we get $2^{l_2+l_3}$ outputs, represented by $z_{i,j}[n]$ for $0 \leq i \leq 2^{l_2}$ and $0 \leq j \leq 2^{l_3}$ [25]. 3 \times 2^L spatial domain basis images of ST with 2 levels of decomposition are shown in Figure 3.

ST is optimal for the C^2 singularities, thus providing sparser representation of smooth curves and surface singularities than wavelet. Theoretically, for the C^2 singularities of an image \mathbf{x} , the best S -term approximation error $\|\tilde{\mathbf{x}}_s - \mathbf{x}\|_2$ using ST has error decay rate of $O(S^{-2})$, while this error decay rate is $O(S^{-1})$ for typical wavelets [28, 29]. Figure 4 shows that a sparser representation is achieved using ST. According to the CS theory [19], the reconstruction error bound is proportional to $\|\Psi^T \mathbf{x} - (\Psi^T \mathbf{x})_s\|_1$, where Ψ^T is the forward transform and S means the number of preserved coefficients in transform domain. Therefore, ST is expected to reduce the image reconstruction error of HSI.

2.2. Incoherence and Gaussian Random Encoding Matrix. Incoherence means a column of the CS encoding matrix must be trying the proliferation in the corresponding sparsity basis Ψ . Since this paper focuses on investigating the spatial-spectral sparsity of HSI, a Gaussian random matrix is chosen as Φ because it is incoherent with the entire existing basis

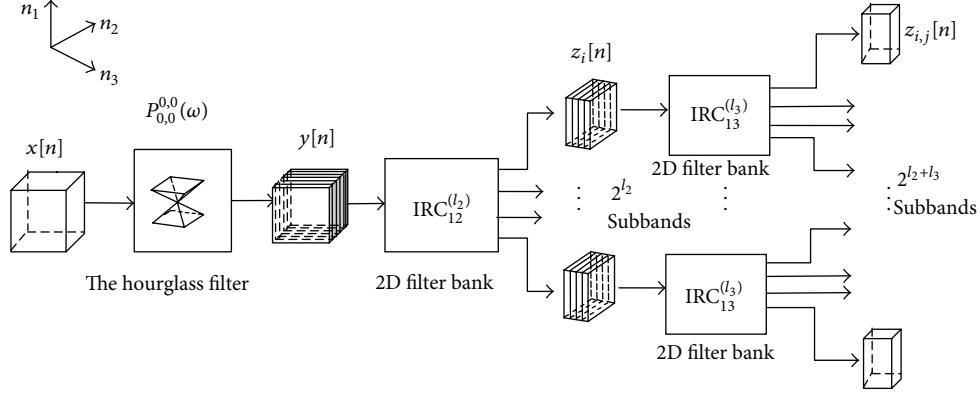
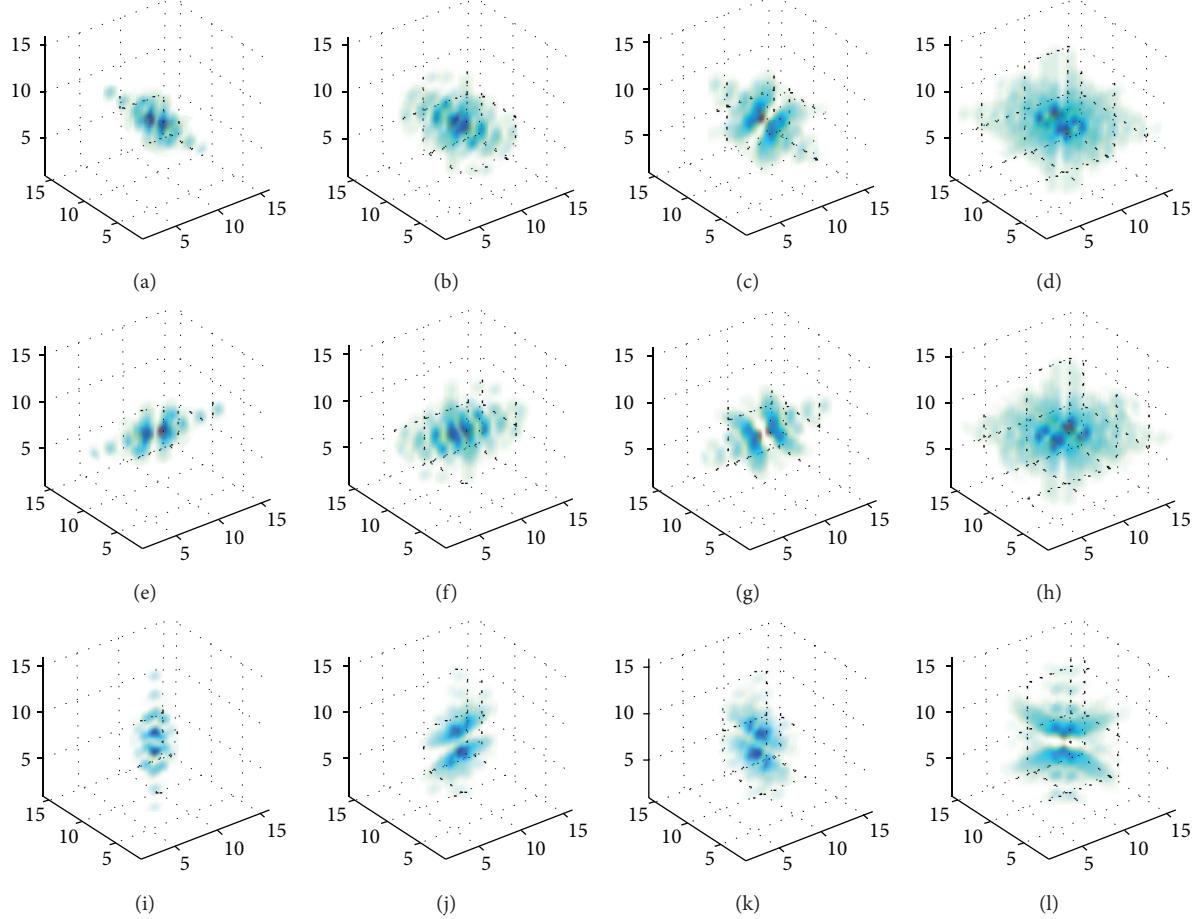


FIGURE 2: One branch of the three-channel structure of 3D-DFB.

FIGURE 3: 3×2^L basis images of ST in 3D spatial domains with 2-level decomposition.

[9, 10]. If each spatial image is encoded separately, Φ can be represented by block matrix:

$$\Phi = \begin{bmatrix} \Phi_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Phi_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Phi_L \end{bmatrix}, \quad (2)$$

where Φ_i ($1 \leq i \leq L$) is an encoding matrix for the band i of HSI. When $\Phi_1 = \Phi_2 = \cdots = \Phi_L$, that is, each 2D image is encoded with the same encoding matrix, then Φ is called 2D random encoding matrix; when $\Phi_1 \neq \Phi_2 \neq \cdots \neq \Phi_L$, that is, each 2D image is encoded differently, then Φ is called 3D random encoding matrix. Physically, 3D random encoding means that digital micromirror device (DMD)

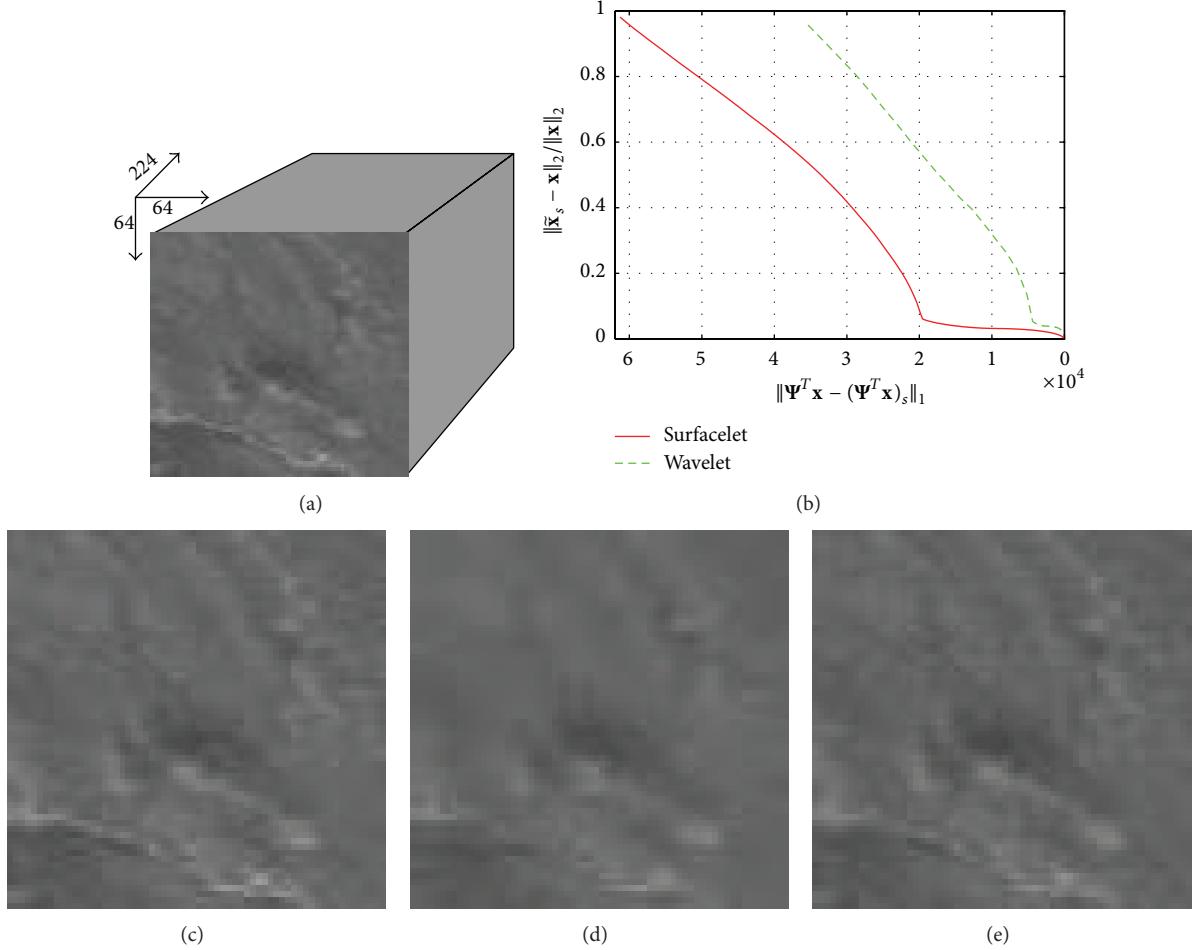


FIGURE 4: The S-term approximation errors and reconstructed images when $\|\Psi^T x - (\Psi^T x)_s\|_1 = 4 \times 10^3$, (a) Lunar Lake HSI data, (b) the S-term approximation error, (c) the original 5th band, (d) sparsely represented image using wavelet (PSNR = 33.12 dB), and (e) sparsely represented image using surfacelet (PSNR = 38.46 dB).

arrays are different for different bands. The performance of both encoding schemes will be discussed in Section 3.

In order to improve the performance of the recovered signal, a Gram-Schmidt orthogonalization (GSO) is used in CS encoding matrix. Given one band random encoding matrix Φ_i ($1 \leq i \leq L$) and the n th column denote as \mathbf{v}_n , the GSO is expressed by

$$\begin{aligned} \bar{\mathbf{v}}_n &= \mathbf{v}_n - \sum_{m=1}^{n-1} \langle \mathbf{v}_n, \bar{\mathbf{v}}_m \rangle \mathbf{v}_m, \quad 2 \leq n \leq J, \\ \varphi_n &= \frac{\bar{\mathbf{v}}_n}{\|\bar{\mathbf{v}}_n\|_2}, \quad 1 \leq n \leq J. \end{aligned} \quad (3)$$

The new matrix $\Phi_i = (\varphi_1, \varphi_2, \dots, \varphi_J)$ is adopted as the CS encoding matrix of one spectral band. The randomness of the matrix is not affected by GSO. In the following, columns of each Φ_i are processed with GSO.

2.3. Reconstruction Algorithm. For HSI $\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(L)}) \in \mathbb{R}^{M \times N \times L}$, where $\mathbf{x}^{(i)}$ represents the i th spectral

band image, $M \times N$ represent the spatial dimensions, and L represents the spectral depth of HSI. Let $\mathbf{x}_i = \text{vec}(\mathbf{x}^{(i)})$ ($1 \leq i \leq L$), where \mathbf{x}_i is a column vector of the i th band HSI with size $MN \times 1$. The data acquisition model for CS is given by

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_L \end{bmatrix} = \begin{bmatrix} \Phi_1 & \mathbf{0}_{J \times MN} & \cdots & \mathbf{0}_{J \times MN} \\ \mathbf{0}_{J \times MN} & \Phi_2 & \cdots & \mathbf{0}_{J \times MN} \\ \vdots & & \ddots & \vdots \\ \mathbf{0}_{J \times MN} & \mathbf{0}_{J \times MN} & \cdots & \Phi_L \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_L \end{bmatrix}, \quad (4)$$

where $\Phi_i \in \mathbb{R}^{J \times MN}$ ($J < MN$) is a random encoding matrix for the i th spectral band image and $\mathbf{y}_i \in \mathbb{R}^{J \times 1}$ is the acquired undersampled data. The sampling ratio is defined as

$$\text{CSR} = \frac{J}{MN}, \quad (5)$$

where $\text{CSR} < 1$, which means the HSI are undersampled.

In this paper, ST is adopted to provide sparser representation of HSI data and is expected to reduce the reconstruction

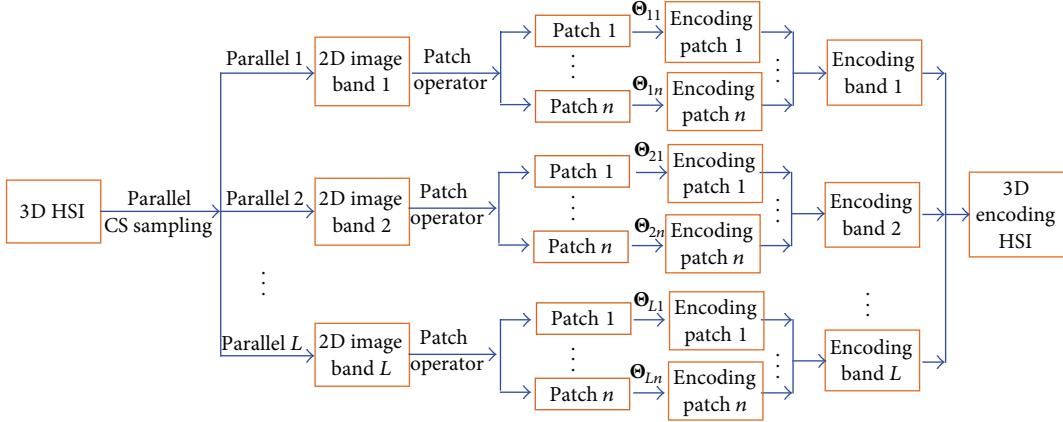


FIGURE 5: Parallel processing of compressive sampling on HIS.

error. Let Ψ represent inverse ST and let Ψ^T denote forward ST; CS recovers \mathbf{x} by solving [18, 19]

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \left\{ \lambda \|\boldsymbol{\alpha}\|_1 + \|\mathbf{y} - \Phi \Psi \boldsymbol{\alpha}\|_2^2 \right\}, \quad (6)$$

where $\boldsymbol{\alpha} = \Psi^T \mathbf{x}$, $\|\cdot\|_p$ ($p = 1, 2$) stands for l_p -norm, and λ is the regularization parameter which decides the tradeoff between the sparsity and the data fidelity. Many researchers seek for a simple and fast algorithm to solve (6), such as the conjugate gradient [30], Bregman iteration [31, 32], low rank reconstruction [33], and other methods. In this paper, we choose FISTA to solve (6) because of its simplicity and fast convergence, whose convergence rate is $O(k^{-2})$, where k is an iteration counter [20].

To solve (6), the smallest Lipschitz constant [20] of the gradient $\|\mathbf{y} - \Phi \Psi \boldsymbol{\alpha}\|_2^2$ is $G = 2\lambda_{\max}((\Phi \Psi)^T (\Phi \Psi))$. The Lipschitz constant ensures the convergence of algorithm [34]. Taking $\boldsymbol{\alpha}_0 = \mathbf{0}$ and $t_1 = 1$ as the initial values, the threshold is λ/G . For $k \geq 1$, solutions are found by iterating from (7) as follows:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_k &= \boldsymbol{\alpha}_k - \frac{1}{G} (\Psi^T \Phi^T (\Phi \Psi \boldsymbol{\alpha}_k - \mathbf{y})), \\ \boldsymbol{\alpha}_k &= \max \{ |\hat{\boldsymbol{\alpha}}_k| - \text{threshold}, 0 \}_{+} \operatorname{sgn}(\hat{\boldsymbol{\alpha}}_k), \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ \boldsymbol{\alpha}_{k+1} &= \boldsymbol{\alpha}_k + \left(\frac{t_k - 1}{t_{k+1}} \right) (\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k-1}). \end{aligned} \quad (7)$$

The maximum number of iterations of FISTA is set as 200 to achieve stable solutions. The final output is $\hat{\mathbf{x}} = \Psi \boldsymbol{\alpha}_k$.

2.4. Patch-Based Compressed Sensing with Surfacelet Reconstruction (PCSST). For a larger HSI data \mathbf{x} , bigger encoding matrix may exceed the memory of computer or leads long computation time. In this case, a patch-based sampling operation can be used. Let \mathbf{R} be a patch operator and let

\mathbf{R}_l ($1 \leq l \leq L$) divide the l th band of \mathbf{x} into n patches, and then the data acquisition model for CS is given by

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_L \end{bmatrix} = \begin{bmatrix} \Theta_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Theta_2 & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \cdots & \Theta_L \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \mathbf{x}_1 \\ \mathbf{R}_2 \mathbf{x}_2 \\ \vdots \\ \mathbf{R}_L \mathbf{x}_L \end{bmatrix}, \quad (8)$$

where \mathbf{y}_i ($1 \leq i \leq L$) denotes measurements of patches in i th band of \mathbf{x} and Θ_i satisfies $\Theta_i = (\Theta_{i1}, \Theta_{i2}, \dots, \Theta_{in})^T$, where Θ_{ij} means random encoding on each block. In our scheme, the encoding on different patches of each band is different and these patches are nonoverlapped.

PCSST allows implementing the proposed method on a larger dataset. It has the potential to allow for parallel computing as shown in Figure 5 and also provides the possibility to reduce the complexity of sensor arrays since the sensing detectors independently sample the data in different patches.

3. Experimental Results

Experiments are conducted in three aspects. First, better reconstructed signal using ST than using wavelet is demonstrated. Second, improving the randomness of encoding matrix for each spectral band is shown to improve the recovery quality and the reconstruction performance of two encoding schemes is compared. Third, a patch-based CS encoding scheme is designed to deal with large data.

The HSI data is obtained from U.S. AVRIS website, including Moffet Field and Lunar Lake [35]. These data contain 224 spectral bands with spatial size 64×64 , and every pixel is encoded with 16 bits. Linear interpolation is employed to fix the junk bands [36] and maintain the consistency of spectral lines, and all data are normalized.

To evaluate the performance, the mean-square-error (MSE), peak signal-to-noise ratio (PSNR), and structural

similarity (SSIM) [37] are adopted as criteria to measure the reconstruction error. Their definitions are

$$\begin{aligned} \text{MSE} &= \frac{1}{M \cdot N} \sum_{j=1}^M \sum_{i=1}^N (\hat{x}(i, j) - x(i, j))^2, \\ \text{PSNR} &= 10 \log_{10} \left(\frac{1}{\text{MSE}} \right), \\ \text{SSIM}(\hat{x}, x) &= \frac{(2\mu_{\hat{x}}\mu_x + C_1)(2\sigma_{\hat{x}x} + C_2)}{(\mu_{\hat{x}}^2 + \mu_x^2 + C_1)(\sigma_{\hat{x}}^2 + \sigma_x^2 + C_2)}, \end{aligned} \quad (9)$$

where x denotes the original image, μ_x and σ_x are mean and standard deviation of x , \hat{x} stands for the recovered image, $\mu_{\hat{x}}$ and $\sigma_{\hat{x}}$ are mean and standard deviation of \hat{x} ; C_1, C_2 are small constant, and $C_1 = C_2 = 0.08$ is adopted in our work, which is used to avoid instability when either $(\mu_{\hat{x}}^2 + \mu_x^2)$ or $(\sigma_{\hat{x}}^2 + \sigma_x^2)$ is very close to zero [37]. Mean SSIM index is adopted in this paper. Simulations run on a dual core 2.5 GHz CPU laptop with 4 GB RAM. A fast ST implementation in C language is adopted.

3.1. Surfacelet and Wavelet with 2D Encoding. To simulate the CS data acquisition, the HSI are undersampled by random encoding matrix Φ . 50% sampled data means $\Phi_i \in \mathbb{R}^{(64^2/2) \times 64^2}$ for $x_i \in \mathbb{R}^{64^2 \times 1}$ ($1 \leq i \leq 224$). In order to improve the quality of the recovered signal, Gram-Schmidt orthogonalization is used to obtain the CS encoding matrix. With the 2D encoding matrix (i.e., each band has the same encoding matrix), the ST-based reconstruction is compared with the 3D wavelets-based reconstruction. Daubechies filters “db4” with 2 levels of decomposition is used, and there are 7 directional subbands for 3D wavelet. The multiscale pyramid with 2 decomposition levels is chosen, and there are 12 directional subbands for ST. The complexity of ST is $O(n \log_{10} n)$ and wavelet is $O(n)$.

In order to well discuss the relationship between λ and PSNR, we compared the PSNR performance from $\lambda = 10$ to $\lambda = 10^{-4}$ as shown in Figure 6. $\lambda = 10^{-2}$ is empirically chosen to give optimal PSNR in our work. As shown in Figure 7, ST significantly improves the PSNR of each spectral band more than 3D wavelet. Edges and curves are better reconstructed using ST than 3D wavelets as shown in Figure 8. At a given spatial location, the spectral line reconstructed using ST is much more consistent with the ground truth than wavelets as shown in Figure 9. Under different sampling ratios, ST achieves much better PSNRs than wavelet as shown in Figure 10, and more advantage of ST is seen at low ratios. Besides, ST shows better performance for the Lunar Lake dataset which has richer geometric textures than Moffet Field dataset. Running time (unit: second) based on ST and WT with different sampled rate is shown in Table 1.

3.2. 2D and 3D Encoding Matrix. The performance of 2D encoding and 3D encoding schemes is compared in this section. The recovered spatial images with two encoding schemes are shown in Figure 11. Edges and curves are better reconstructed using 3D encoding than 2D encoding. The

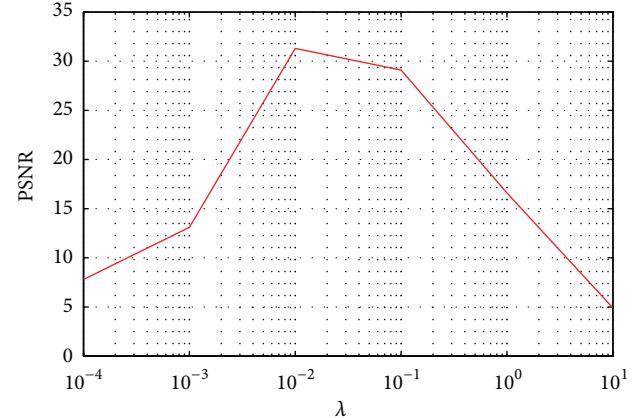


FIGURE 6: PSNR performance versus different λ . Note: test is performed on Lunar Lake dataset at sampling ratio $R = 0.4$.

recovered spectral lines using 3D encoding are more consistent with ground truth than using 2D encoding as shown in Figure 12. The improvement is more obvious for Moffet Field dataset with textures in low intensities as shown in Figures 11(a) and 12(a). These results imply that increasing the encoding randomness among different bands will achieve better reconstruction.

3.3. Reconstruction with PCSST. We tested the methods on Lunar Lake data with size $256 \times 256 \times 224$. A reconstructed spectral band is shown in Figure 13. The image structures recovered by surfacelet are much sharper than wavelet.

4. Conclusions

Compressive sensing is a new sampling theorem. In this paper, the surfacelet transform is introduced into hyperspectral image reconstruction from compressive sampled data. The surfacelet transform is a directional multiresolution transform for 3D data, which is applied to sparsify the hyperspectral images for the first time. Simulations are conducted in three aspects. First, better reconstructed signal using surfacelet than using wavelet is demonstrated. Second, improving the randomness of encoding matrix for each spectral band is shown to improve the recovery quality. Third, a patch-based CS encoding scheme is designed to deal with large data. It provides the possibility to reduce the complexity of sensor arrays because sensing detectors in different patches independently sample data. Experiments demonstrate that reconstruction of spectral lines and spatial images is significantly improved using the proposed method than using conventional three-dimensional wavelets.

In the future, our work includes the following two aspects.

The first aspect is optimizing sparse representation of hyperspectral imaging, for example, combining adaptive sparse representation [38–40] and surfacelet transform. An adaptive dictionary may provide a sparser representation leading to a lower reconstruction error. Therefore, it is meaningful to try a low-complexity training method in the future.

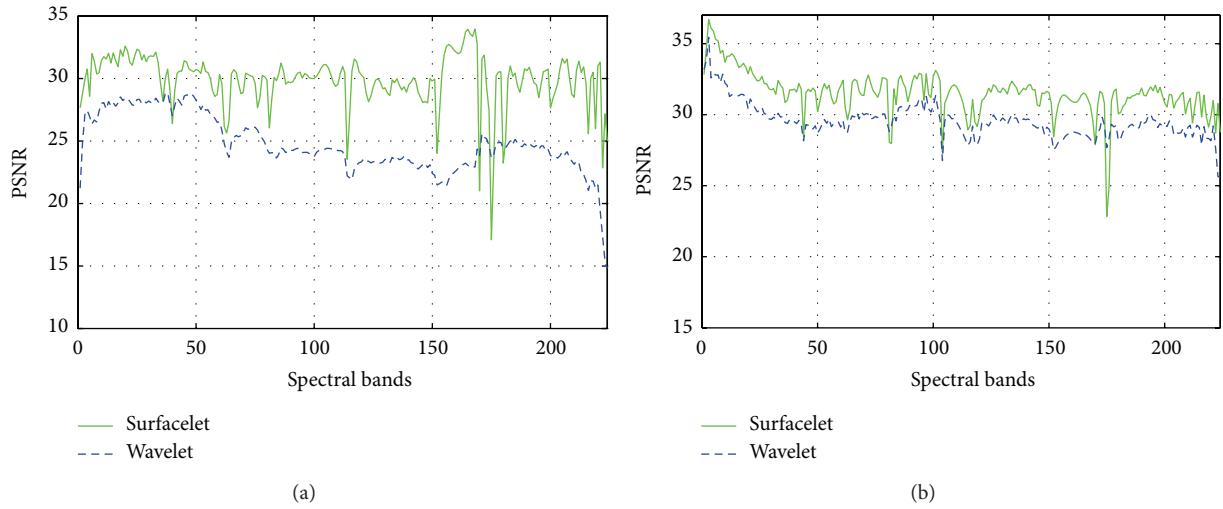


FIGURE 7: PSNR performance using wavelet and surfacelet when sampling ratio is 0.40. (a) Moffet Field dataset and (b) Lunar Lake dataset.

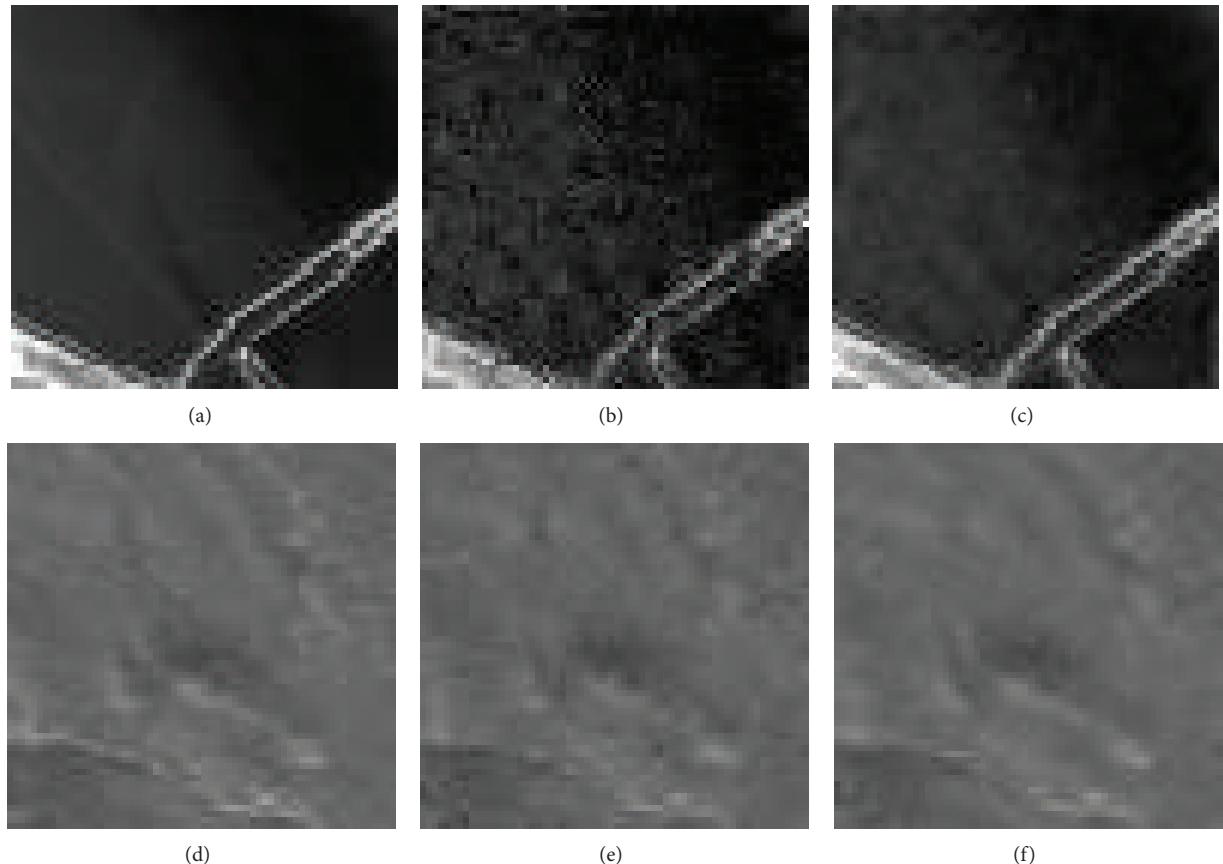


FIGURE 8: The reconstructed 199th band of Moffet Field and 5th band of Lunar Lake using wavelets or surfacelet when sampling ratio $R = 0.40$. (a) and (d) are original Moffet Field and Lunar Lake bands, (b) PSNR = 24.31 dB, SSIM = 0.9020 and (e) PSNR = 32.91 dB, SSIM = 0.9462 are recovered bands using wavelet, and (c) PSNR = 30.75 dB, SSIM = 0.9384 and (f) PSNR = 35.88 dB, SSIM = 0.9553 are reconstructed bands using surfacelet.

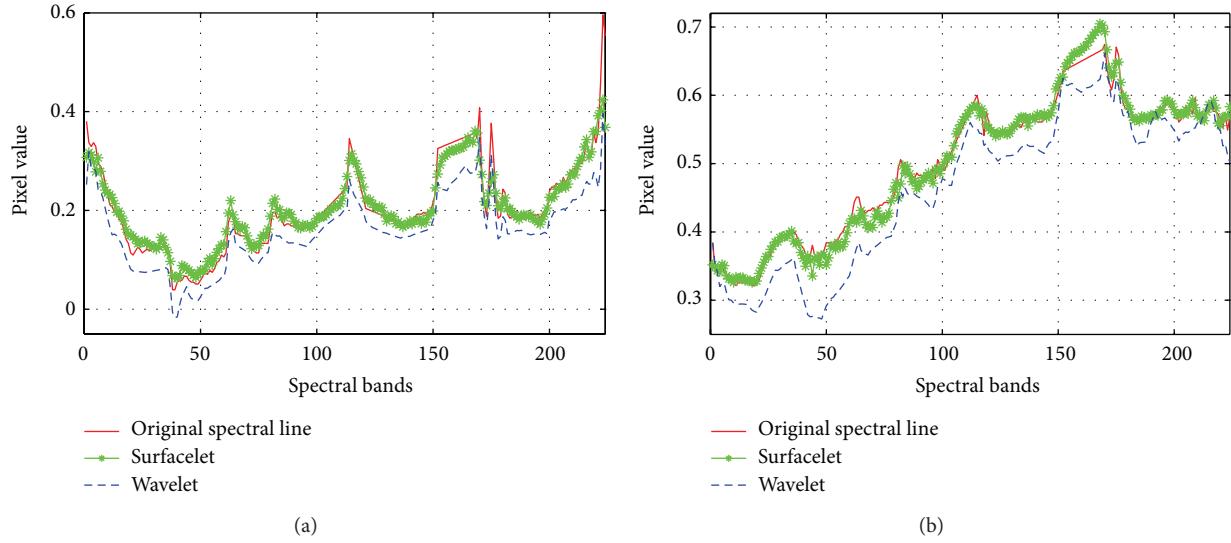


FIGURE 9: Reconstructed spectral line when sampling ratio $R = 0.40$. (a) Moffet Field dataset and (b) Lunar Lake dataset.

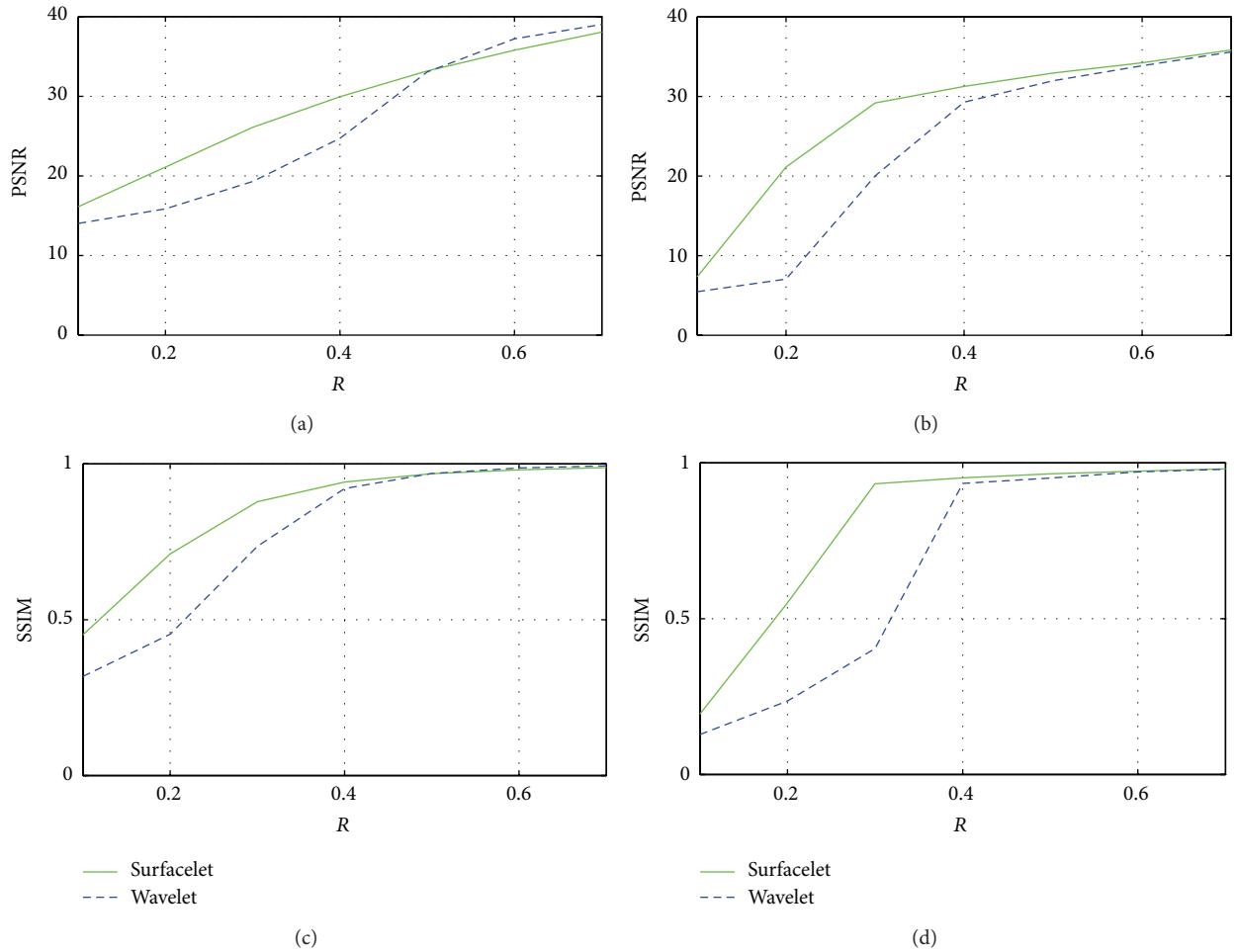


FIGURE 10: PSNR and SSIM at different sampling ratios. (a), (c) PSNR and SSIM for Moffet Field dataset and (b), (d) PSNR and SSIM for Lunar Lake dataset.

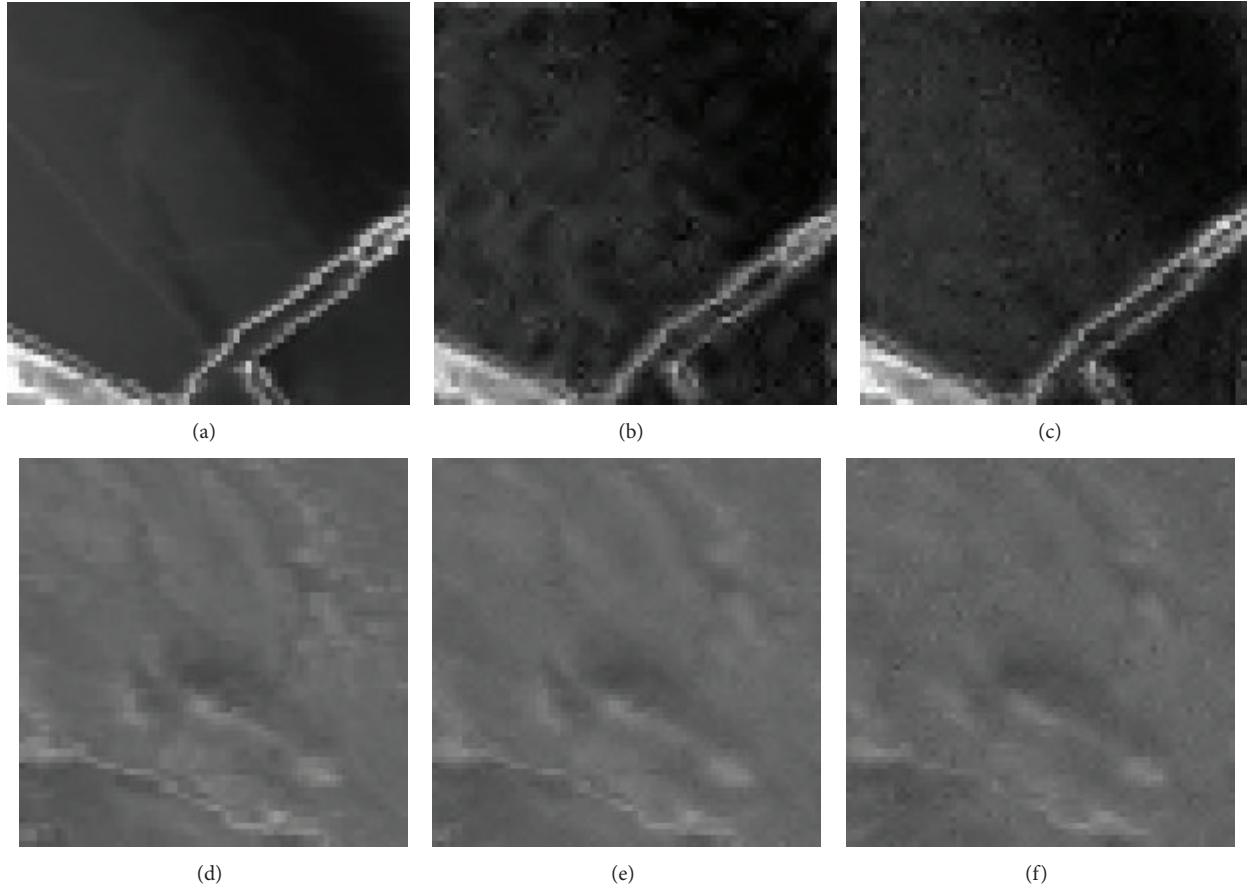


FIGURE 11: The reconstructed 199th band of Moffet Field and 5th band of Lunar Lake using 2D encoding and 3D encoding when sampling ratio $R = 0.3$. (a) and (d) are original Moffet Field and Lunar Lake bands, (b) PSNR = 26.84 dB, SSIM = 0.9120 and (e) PSNR = 33.65 dB, SSIM = 0.9330 are recovered bands using 2D encoding matrix for each band and (c) PSNR = 34.99 dB, SSIM = 0.9600 and (f) PSNR = 37.84 dB, SSIM = 0.9701 are reconstructed bands using different encoding matrix for each band.

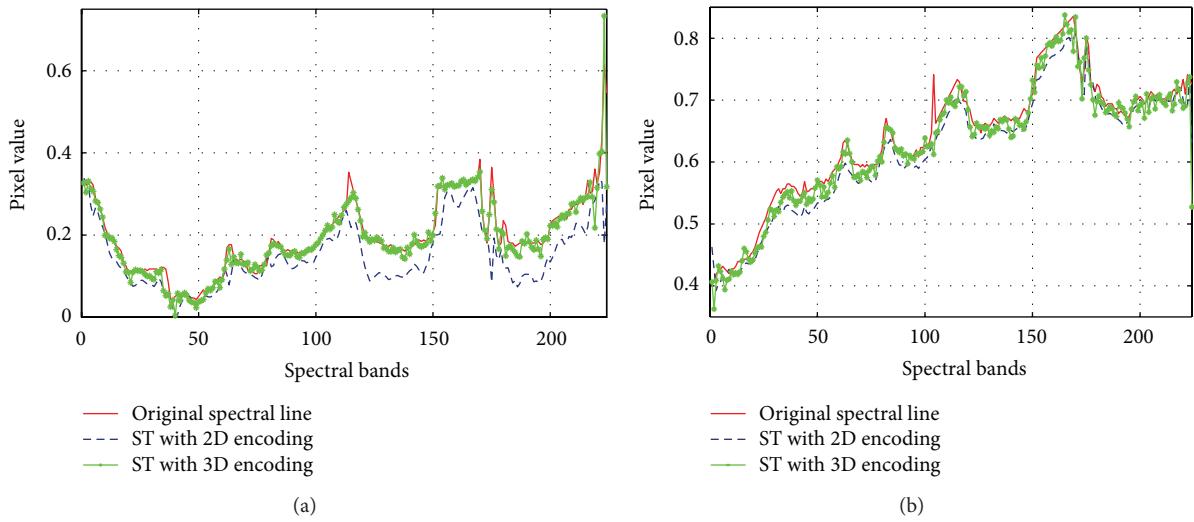


FIGURE 12: Reconstructed spectral line when sampling ratio $R = 0.3$. (a) Moffet Field dataset and (b) Lunar Lake dataset.

TABLE 1: Running time (unit: second) based on ST and WT with different sampled rate.

Sampling ratio	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ST	458.9	629.7	833.8	1002.9	1224.2	1381.8	1544.9	1758.7
WT	989.8	1115.5	1359.7	1639.9	1871.6	2089.6	2324.1	2452.9

Note: the ST is implemented in C and WT is implemented in MATLAB.

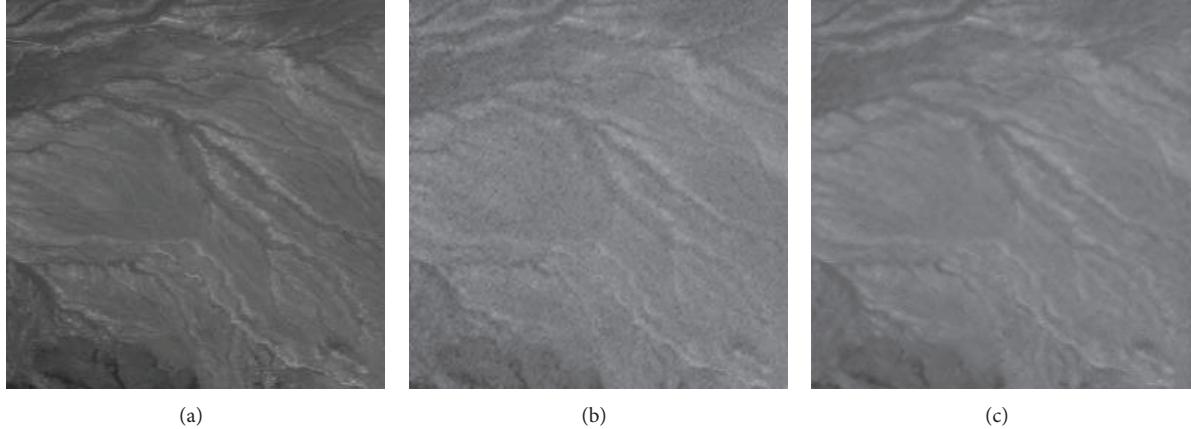


FIGURE 13: Patch-based compressed sensing ST reconstruction. (a) The original 5th band of 256×256 Lunar Lake dataset, (b) the patch-based reconstructed image using wavelet when undersampled rate $R = 0.4$ and the PSNR = 32.92 dB, SSIM = 0.9551, and (c) the patch-based reconstructed image using ST when undersampled rate $R = 0.4$ and the PSNR = 34.99 dB, SSIM = 0.9876.

The second aspect is overlapping patches compressed sensing reconstruction. Overlapping patches can reduce the “block artifacts.” But overlapping also introduces more encoded data and higher computations on image reconstructions. Trading the compressive sampling rate with image qualities using overlapping and speeding up the reconstruction algorithm [41] will be discussed in the future.

Conflict of Interests

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

Acknowledgments

The authors wish to express their thanks to all reviewers for their comments. This work was supported in part by the National Natural Science Foundation of China (no. 61201045, no. 40971206, no. 61302174), Fundamental Research Funds for the Central Universities (no. 2013SH002), Open Fund from Key Lab of Digital Signal and Image Processing of Guangdong Province (no. 54600321, no. 2013GDDSIPL-07), Academic Innovation Team Building Project of Shantou University (no. ITCI2002), and Scientific Research Foundation for the Introduction of Talent at Xiamen University of Technology (no. 90030606).

References

- [1] A. F. H. Goetz, “Three decades of hyperspectral remote sensing of the Earth: a personal view,” *Remote Sensing of Environment*, vol. 113, supplement 1, pp. S5–S16, 2009.
- [2] J. Ma, “Single-Pixel remote sensing,” *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 199–203, 2009.
- [3] Y. Zhao, J. Yang, and J. C. W. Chan, “Hyperspectral imagery super-resolution by spatial-spectral joint nonlocal similarity,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 99, pp. 2671–2679, 2013.
- [4] Y. Zhao, J. Yang, Q. Zhang, L. Song, Y. Cheng, and Q. Pan, “Hyperspectral imagery super-resolution by sparse representation and spectral regularization,” *EURASIP Journal on Advances in Signal Processing*, vol. 2011, article 87, 2011.
- [5] B. Aiazzi, L. Alparone, S. Baronti, C. Lastri, and M. Selva, “Spectral distortion in lossy compression of hyperspectral data,” *Journal of Electrical and Computer Engineering*, vol. 2012, Article ID 850637, 8 pages, 2012.
- [6] J. G. Kolo, S. A. Shanmugam, D. W. G. Lim, L.-M. Ang, and K. P. Seng, “An adaptive lossless data compression scheme for wireless sensor networks,” *Journal of Sensors*, vol. 2012, Article ID 539638, 20 pages, 2012.
- [7] Q. Du and J. E. Fowler, “Hyperspectral image compression using JPEG2000 and principal component analysis,” *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 201–205, 2007.
- [8] C.-I. Chang, B. Ramakrishna, J. Wang, and A. Plaza, “Exploitation-based hyperspectral image compression,” *Journal of Applied Remote Sensing*, vol. 4, Article ID 041760, 2010.
- [9] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [10] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [11] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen, “Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator,” *Medical Image Analysis*, vol. 18, no. 6, pp. 843–856, 2014.

- [12] W. Hao, J. Li, X. Qu, and Z. Dong, "Fast iterative contourlet thresholding for compressed sensing MRI," *Electronics Letters*, vol. 49, no. 19, pp. 1206–1208, 2013.
- [13] J. Xiong and Q. Tang, "1-bit compressive data gathering for wireless sensor networks," *Journal of Sensors*, vol. 2014, Article ID 805423, 8 pages, 2014.
- [14] C. Caione, D. Brunelli, and L. Benini, "Compressive sensing optimization for signal ensembles in WSNs," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 382–392, 2014.
- [15] M. F. Duarte, M. A. Davenport, D. Takbar et al., "Single-pixel imaging via compressive sampling: building simpler, smaller, and less-expensive digital cameras," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [16] N. V. Aravind, K. Abhinandan, V. V. Acharya, and D. S. Sumam, "Comparison of OMP and SOMP in the reconstruction of compressively sensed hyperspectral images," in *Proceedings of the International Conference on Communications and Signal Processing (ICCP '11)*, pp. 188–192, Hamirpur, India, February 2011.
- [17] J. E. Fowler, "Compressive-projection principal component analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2230–2242, 2009.
- [18] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.
- [19] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.
- [20] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [21] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using 3D wavelets," in *Proceedings of the 32nd IEEE International Geoscience and Remote Sensing Symposium (IGARSS '12)*, pp. 1349–1352, July 2012.
- [22] X. Tang and W. A. Pearlman, "Three-dimensional wavelet-based compression of hyperspectral images," *Hyperspectral Data Compression*, pp. 273–308, 2006.
- [23] S. Prasad, W. Li, J. E. Fowler, and L. M. Bruce, "Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 9, pp. 3474–3486, 2012.
- [24] A. Karami, M. Yazdi, and G. Mercier, "Compression of hyperspectral images using discrete wavelet transform and tucker decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 444–450, 2012.
- [25] Y. M. Lu and M. N. Do, "Multidimensional directional filter banks and surfacelets," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 918–931, 2007.
- [26] X. Qu, W. Zhang, D. Guo, C. Cai, S. Cai, and Z. Chen, "Iterative thresholding compressed sensing MRI based on contourlet transform," *Inverse Problems in Science and Engineering*, vol. 18, no. 6, pp. 737–758, 2010.
- [27] P. Feng, Y. Pan, B. Wei, W. Jin, and D. Mi, "Enhancing retinal image by the contourlet transform," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 516–522, 2007.
- [28] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.
- [29] E. J. Candès and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Communications on Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219–266, 2004.
- [30] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [31] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for l_1 -minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143–168, 2008.
- [32] J.-F. Cai, S. Osher, and Z. Shen, "Linearized Bregman iterations for compressed sensing," *Mathematics of Computation*, vol. 78, no. 267, pp. 1515–1536, 2009.
- [33] Y. Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 296–308, 2015.
- [34] K. Goebel and W. A. Kirk, "A fixed point theorem for transformations whose iterates have uniform Lipschitz constant," *Studia Mathematica*, vol. 47, pp. 135–140, 1973.
- [35] AVIRIS Homepage, JPL, NASA, December 2010, <http://www.nasa.gov/>.
- [36] P. Zhong and R. Wang, "Multiple-spectral-band CRFs for denoising junk bands of hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2260–2275, 2013.
- [37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [38] E. le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 423–438, 2005.
- [39] X. Qu, D. Guo, B. Ning et al., "Undersampled MRI reconstruction with patch-based directional wavelets," *Magnetic Resonance Imaging*, vol. 30, no. 7, pp. 964–977, 2012.
- [40] 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.
- [41] J. Yang, Y. Yamaguchi, and W.-M. Boerner, "Numerical methods for solving the optimal problem of contrast enhancement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 965–971, 2000.

Research Article

Use of GMM and SCMS for Accurate Road Centerline Extraction from the Classified Image

Zelang Miao,¹ Bin Wang,¹ Wenzhong Shi,¹ Hao Wu,² and Yiliang Wan³

¹Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

²School of Resources and Environmental Engineering, Wuhan University of Technology, Wuhan, Hubei 430070, China

³School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan, Hubei 430079, China

Correspondence should be addressed to Wenzhong Shi; johnshiforscholar@gmail.com

Received 29 August 2014; Revised 22 October 2014; Accepted 23 October 2014

Academic Editor: Yongqiang Zhao

Copyright © 2015 Zelang Miao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The extraction of road centerline from the classified image is a fundamental image analysis technology. Common problems encountered in road centerline extraction include low ability for coping with the general case, production of undesired objects, and inefficiency. To tackle these limitations, this paper presents a novel accurate centerline extraction method using Gaussian mixture model (GMM) and subspace constraint mean shift (SCMS). The proposed method consists of three main steps. GMM is first used to partition the classified image into several clusters. The major axis of the ellipsoid of each cluster is extracted and deemed to be taken as the initial centerline. Finally, the initial result is adjusted using SCMS to produce precise road centerline. Both simulated and real datasets are used to validate the proposed method. Preliminary results demonstrate that the proposed method provides a comparatively robust solution for accurate centerline extraction from a classified image.

1. Introduction

Very High Resolution (VHR) satellite images have become increasingly available in recent year with the advent of modern sensors, such as QuickBird, Pleiades, and WorldView. Road centerline delineation from these images has a variety of real applications (e.g., change detection, Geographic Information System (GIS) database updating, image registration, etc.). Massive approaches have been developed over the past decades to address the road extraction issue. Mena [1] and Das et al. [2] have presented good reviews of road extraction methods. Generally speaking, road centerline extraction commonly consists of two main steps: (1) road areas extraction and (2) centerline extraction from the road areas.

The first step focuses on the extraction of road areas from the satellite images. In this field, most commonly used methods rely on the classification technology. The most traditional classification methods are pixelwise [3] that rely on spectral information only. Such pixelwise methods always meet the “salt-and-pepper” phenomenon [4] and there is large room to improve classification accuracy. In recent

years, with the development of VHR satellite images, road geometrical features are derived and explored to integrate with spectral information to meet the challenge of the “salt-and-pepper” phenomenon. To do so, a variety of road shape features are integrated with spectral information to improve road extraction accuracy, including smoothness, compactness [5], shape index [6, 7], second order moment [2, 8], and mathematical morphology [9, 10]. To improve the classification, numerous algorithms have been proposed, such as sparse representation and low-rank constraint [11] and band subset-based clustering and fusion method [12]. Except the classification methods, there are also some other road areas extraction methods, including the segmentation method [13–16], active contour (i.e., “snake” model) [17, 18], and level set [19, 20].

The second step involves centerline extraction from road areas extracted in the first step. The commonly used method is the thinning algorithm [21]. Despite its computational efficiency, the result of the thinning algorithm always produces undesired objects (i.e., spurs) that reduce the smoothness and accuracy of the centerline. To tackle this limitation to some extent, a self-organized clustering method [22] was designed.

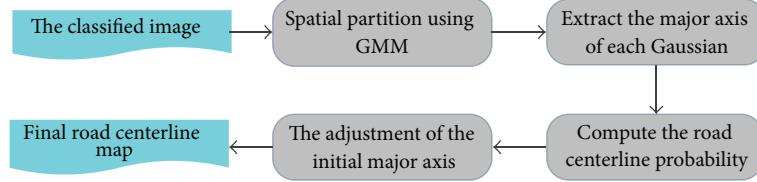


FIGURE 1: Flowchart of the proposed method.

This method firstly extracted initial points using K -medians clustering, followed by the linking algorithm to create the central line. Radon and Hough transform [23, 24] were also introduced to extract the centerline from the classified image. Strong peak locations in the transform matrix are taken to be straight line pixels in the original image. Radon and Hough transform are suitable for straight line extraction and generally fail to process curvilinear cases. Recently, in the area of computer vision, a so-called subspace constrained mean shift (SCMS) method [25–27] was proposed to extract accurate centerline from point clouds. First, the probability that a pixel is located on the centerline is computed using the kernel density estimation (KDE). Then, all discrete points are iteratively projected onto the centerline based on mean shift. Despite its good performance, when the number of points is large, SCMS becomes very slow, which limits its real applications. To improve the computational efficiency of SCMS, feature points-based SCMS (F-SCMS) [28] was proposed. F-SCMS firstly detected feature points (i.e., end and junction points) from the classified image, followed by the connection of feature points by the geodesic method to create the central lines to formulate the road network. Although F-SCMS can largely improve the speed of SCMS, it cannot process the classified image with closed form, such as circle. The aforementioned discussions indicate that the extraction of accurate road centerline from the classified image is not well resolved and this topic is still open and challenging.

There are also many other interesting studies on road extraction from a different viewpoint, such as Kalman filtering [29], road footprint [30], and dynamic programming [31]. Although most of the aforementioned studies focus on multispectral satellite images, other sources of information can also be used to extract road networks, such as Synthetic Aperture Radar (SAR) [18, 32, 33], LiDAR [34–36], and hyperspectral images [6, 37, 38]. This paper focuses on the accurate road centerline extraction from the classified image. It is worthy to point out that this focus will not weaken the proposed method, which is also suitable for centerline extraction from road areas extracted from other sources.

The main objectives of this paper are as follows:

- (i) improving the efficiency of SCMS;
- (ii) proposing a method for accurate road centerline extraction from the classified image with high generalization ability.

The remainder of this paper is organized as follows. The new approach is presented in Section 2. The experimental

results are given and discussed in Section 3. Finally, Section 4 concludes the paper.

2. Methodology

The objective of this study is to devise a computationally efficient approach with high general ability to extract accurate centerlines from classified images. Figure 1 summarizes the proposed method.

The proposed method consists of the three following steps.

- (1) The classified image is partitioned using the Gaussian mixture model (GMM) method.
- (2) The major axis is subsequently extracted from the ellipsoid of each Gaussian.
- (3) The major axis is adjusted based on SCMS to produce accurate road centerline.

In the following subsections, details of each step are described.

2.1. Image Partition Using GMM. Road pixels in the classified map can be represented as 2D discrete joint random variables. We can partition the classified image into multiple segments by taking road pixels as observations. To do so, the Gaussian mixture model (GMM) is used to implement image clustering. The reason of selecting GMM is that the Gaussian models used by the expectation-maximization (EM) algorithm are flexible and the EM result is able to accommodate clusters of variable size. This is because EM benefits from the Gaussian distribution present in the dataset. Therefore, GMM is able to capture the complex spatial structure. In other words, GMM is able to simplify the classified road map as well as retain its spatial topology. The concepts of GMM are briefly introduced as follows.

GMM can be written as a linear superposition of K Gaussians that leads to a probability density in the form

$$s_{K,\nu,\Sigma,w}(\mathbf{x}) = \sum_{k=1}^K \pi_k \Phi_{\nu_k, \Sigma_k}(\mathbf{x}), \quad (1)$$

where $\{\pi_k\}_{k=1}^K$ are mixture weights subject to $0 \leq \pi_k \leq 1$, $\sum_{k=1}^K \pi_k = 1$, and $\Phi(\cdot)$ is a multivariate Gaussian probability distribution of mean ν_k and covariance matrix Σ_k . Consider

$$\Phi_{\nu,\Sigma}(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{(-1/2)(\mathbf{x}-\nu)^T \Sigma^{-1} (\mathbf{x}-\nu)}. \quad (2)$$

```

Input: The classified image.
Output: The mean  $\mu$  and covariance matrix  $\Sigma$ .
(1) Compute the spatial cluster number  $K$  using (6).
(2) Initial the parameters of EM using  $k$ -Means.
(3) Run the EM algorithm to segment the classified image until convergence.
(4) for  $i = 1, \dots, K$  do
(5)   compute the mean  $\mu_i$  and covariance matrix  $\Sigma_i$  of  $i$ th Gaussian.
(6) end for

```

ALGORITHM 1: Image partition using GMM.

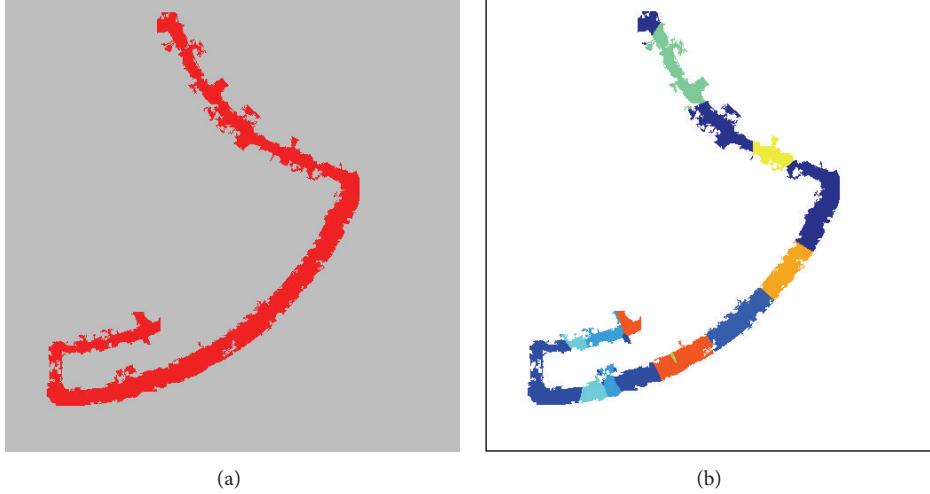


FIGURE 2: (a) The classified road map. (b) The spatial partition result using GMM.

The expected value of the indicator variables under the posterior distribution is then given by

$$\gamma(z_{ij}) = \frac{\pi_j \Phi_{\nu_j, \Sigma_j}(\mathbf{x}_i)}{\sum_{k=1}^K \pi_k \Phi_{\nu_k, \Sigma_k}(\mathbf{x}_i)}. \quad (3)$$

The log-likelihood function is defined as

$$\ln p(x, z | \Phi) = \sum_{i=1}^N \sum_{k=1}^K \gamma(z_{ij}) \ln \{\pi_k \Phi_{\nu_k, \Sigma_k}(\mathbf{x}_i)\}. \quad (4)$$

The parameters of each component are updated using the expectation-maximization (EM) algorithm as follows:

$$\begin{aligned} \pi_k &= \frac{1}{N} \sum_{i=1}^N \gamma(z_{ij}), \\ \mu_j &= \frac{1}{\sum_{i=1}^N \gamma(z_{ij})} \sum_{i=1}^N \gamma(z_{ij}) \mathbf{x}_i, \\ \Sigma_j &= \frac{1}{\sum_{i=1}^N \gamma(z_{ij})} \sum_{i=1}^N \gamma(z_{ij}) (\mathbf{x}_i - \mu_j)(\mathbf{x}_i - \mu_j)^T. \end{aligned} \quad (5)$$

The EM step is iteratively repeated until convergence. It is worthy of pointing out that initial estimates for the

parameters ($\Phi(\cdot)$) are necessary for the first EM iteration. To this end, the k -Means algorithm is selected to initialize such parameters involving running. The spatial cluster number for k -Means is estimated by

$$K = \left\lceil \frac{A}{R_W \times M_L} \right\rceil, \quad (6)$$

where $\lceil \cdot \rceil$ means to round the number to the nearest integer greater than or equal to this number, A is the area of the classified image, R_W is the average road width of the classified image, and M_L is the minimum area of a spatial cluster.

Details of the image partition using GMM are illustrated in Algorithm 1. Figure 2 gives a conceptual example. Figure 2(a) is the classified image and its corresponding spatial partition result using GMM is shown in Figure 2(b).

2.2. Extraction of Initial Road Centerlines. After the spatial partition using GMM in Section 2.1, the mean μ and covariance matrix Σ for each mixture are produced. The ellipsoids of each Gaussian can be defined using μ and Σ . In this study, the major axis of each ellipsoid is taken as the approximate initial centerline. Σ is a square matrix that can be factorized as

$$\Sigma = U \Lambda U^T, \quad (7)$$

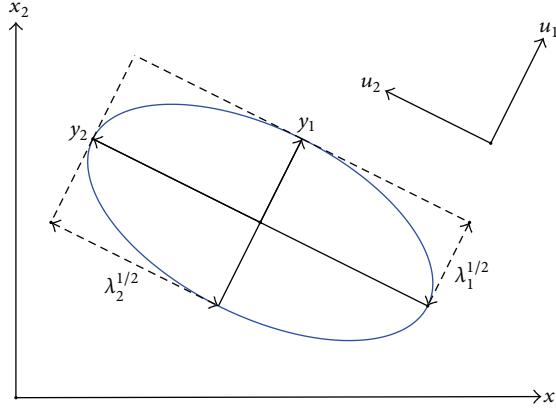


FIGURE 3: An example of the ellipsoid derived by the mean and covariance matrix.

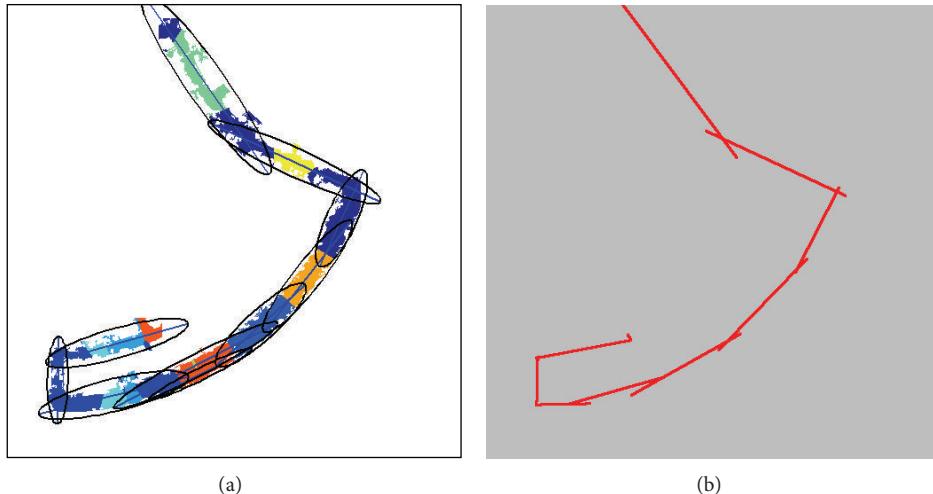


FIGURE 4: (a) Ellipse of each Gaussian. (b) The major axis which is shown in red.

where U is the eigenvector matrix and Λ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. The inverse matrix of Σ is given by

$$\Sigma^{-1} = U^{-T} \Lambda^{-1} U^{-1} = U \Lambda^{-1} U = \sum_{i=1}^p \frac{1}{\lambda_i} \vec{u}_i \vec{u}_i^T. \quad (8)$$

The inverse matrix Σ^{-1} can be rewritten as

$$\begin{aligned} (\vec{x} - \vec{u})^T \Sigma^{-1} (\vec{x} - \vec{u}) &= (\vec{x} - \vec{u})^T \left(\sum_{i=1}^p \frac{1}{\lambda_i} \vec{u}_i \vec{u}_i^T \right) (\vec{x} - \vec{u}) \\ &= \sum_{i=1}^p \frac{1}{\lambda_i} (\vec{x} - \vec{u})^T \vec{u}_i \vec{u}_i^T (\vec{x} - \vec{u}) \quad (9) \\ &= \sum_{i=1}^p \frac{y_i^2}{\lambda_i}, \end{aligned}$$

where $y_i \triangleq \vec{u}_i^T (\vec{x} - \vec{u})$. The y variables define a new coordinate system that is shifted (by \vec{u}) and rotated (by U) with respect

to the original x coordinates: $\vec{y} = U(\vec{x} - \vec{u})$. Recall that the equation for an ellipse in 2D is

$$\frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} = 1. \quad (10)$$

Equation (10) indicates that the contours of equal probability density of a Gaussian lie along ellipse. Figure 3 demonstrates an example of the ellipsoid generated by the mean and covariance matrix.

For each Gaussian mixture, the major axis of the ellipse is estimated. After that, all major axis will compose of a coarse road centerline network that has a similar spatial topology with the true one, as shown in Figure 4.

2.3. Adjustment of Initial Road Centerline Using SCMS. The major axis produced by GMM in Section 2.2 can be regarded as the initial road centerline that is commonly coarse. For the purpose of accurate road centerline result, it is promising to obtain accurate road centerline using the adjustment techniques. To this end, subspace constrained mean shift (SCMS) is selected to align the major axis. However, SCMS is iteratively implemented on the whole classified road pixels,

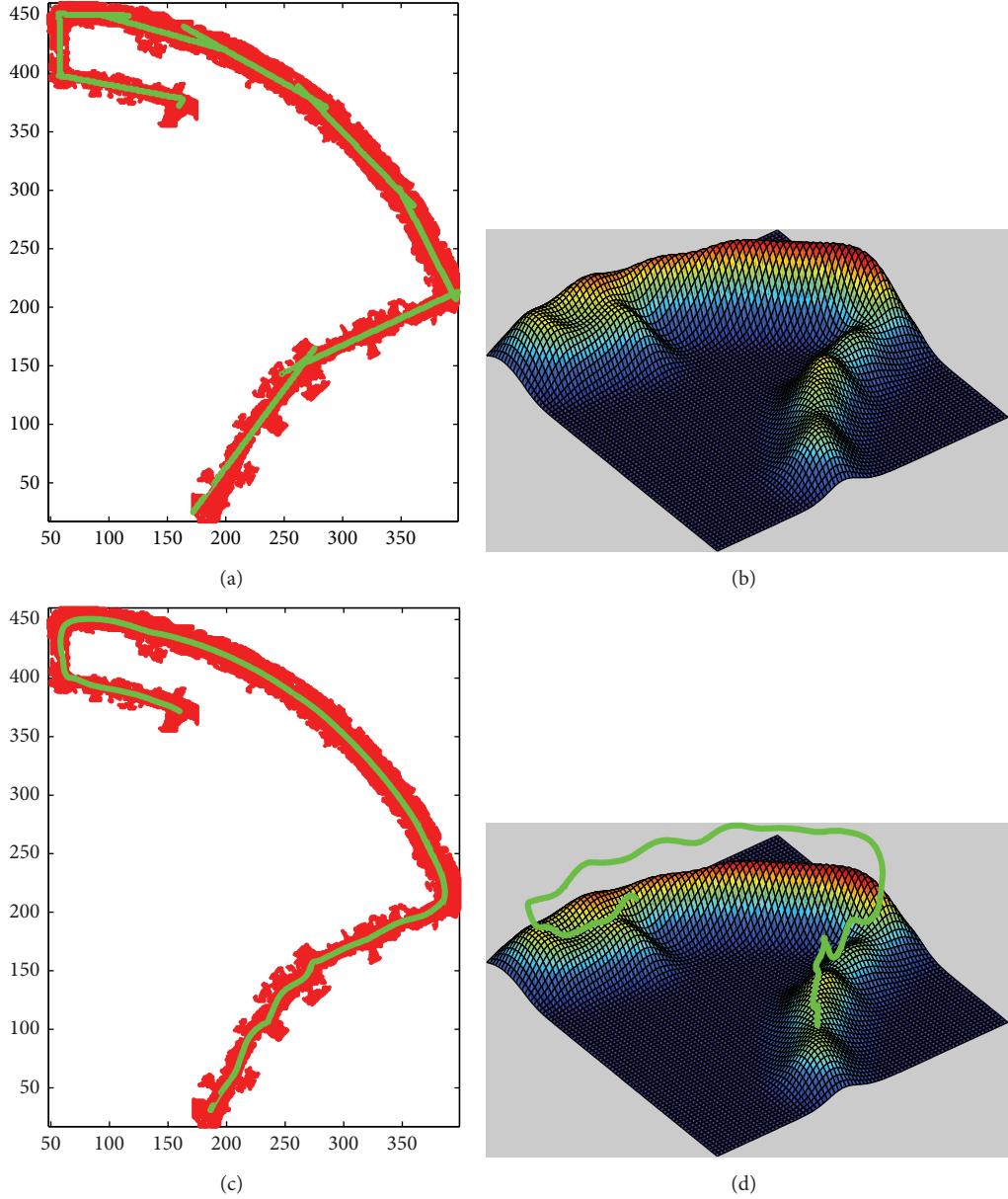


FIGURE 5: An example of a principal curve. (a) Noisy input points. (b) The estimated probability density map. (c) The principal curve over the probability density map where the principal curve is shown in green. (d) The principal curve projected back onto the plane of original points.

which leads to high computational load. On the other hand, the major axis can also be taken as the simplification case of the classified road map that only accounts for small size pixels. In other words, GMM finds the representative points of the classified road pixels to be used in principal curve projection; and therefore, GMM followed by SCMS, termed as GMM-SCMS, provides a new means to lower the computational load. The SCMS method is briefly introduced as follows.

Considering a d -variate random sample X_1, X_2, \dots, X_n drawn from a population with its density function f , the kernel density estimator (KDE) is defined as

$$\hat{f}(x; \Sigma) = n^{-1} \sum_{i=1}^n K_\Sigma(x - X_i), \quad (11)$$

where $x = (x_1, x_2, \dots, x_d)^T$ and $X_i = (X_{i1}, X_{i2}, \dots, X_{id})^T$, $i = 1, 2, \dots, n$. In addition, the symmetric positive definite $d \times d$ matrix Σ is the bandwidth matrix, which critically determines the performance of the KDE function \hat{f} . K is a d -variate kernel function defined as

$$K_\Sigma(x) = |\Sigma|^{-1/2} K(\Sigma^{-1/2}x) \quad (12)$$

which satisfies $\int K(x)dx = 1$. Here, $|\cdot|$ indicates the determinant operation, and the common choice of Gaussian KDE functions is adopted; $K(x) = (2\pi)^{-d/2} \exp(-(1/2)x^T x)$.

By generalizing Scott's rule of thumb, bandwidth matrix $\Sigma^{-1} = n^{2/(d+4)} \hat{\Sigma}^{-1}$ is chosen here (see Ahamada and Flachaire

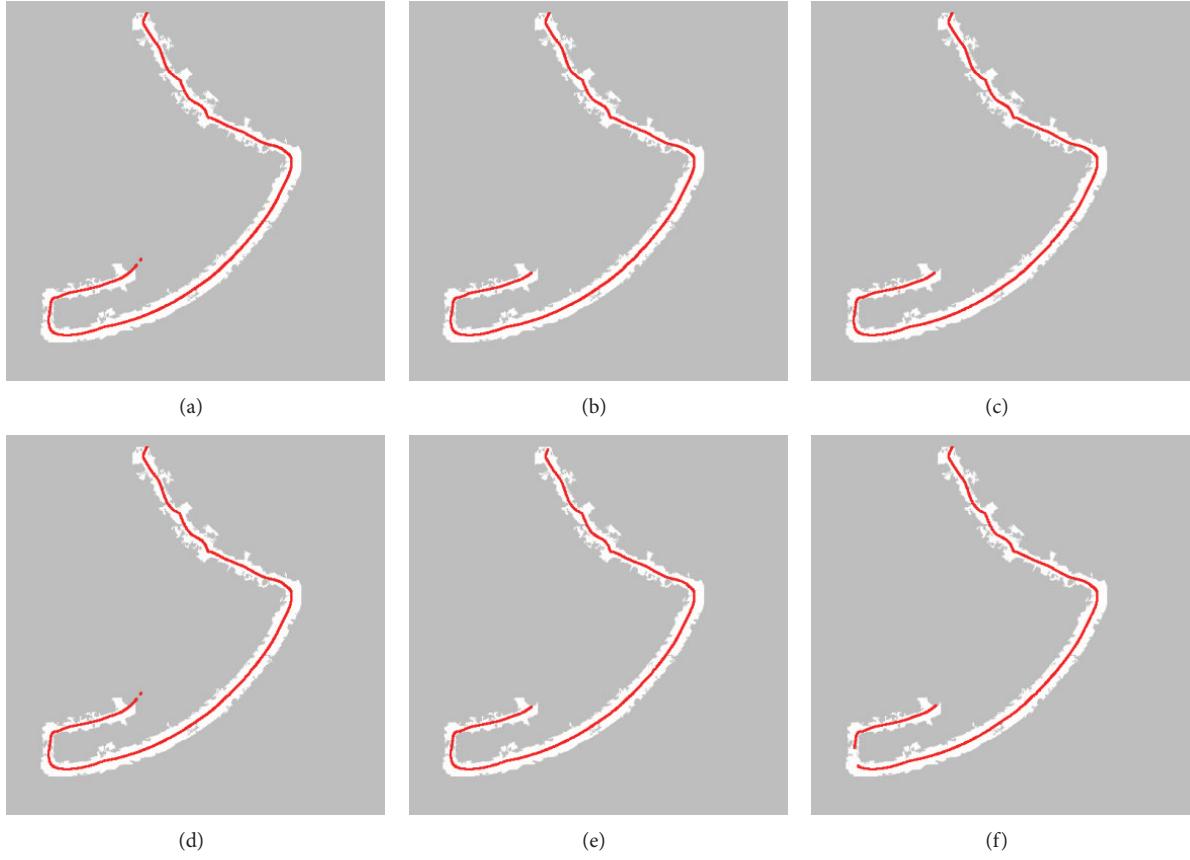


FIGURE 6: Road centerline extraction results with different minimum area values. (a) 100. (b) 120. (c) 140. (d) 160. (e) 180. (f) 200. The road centerline extraction results are shown in red.

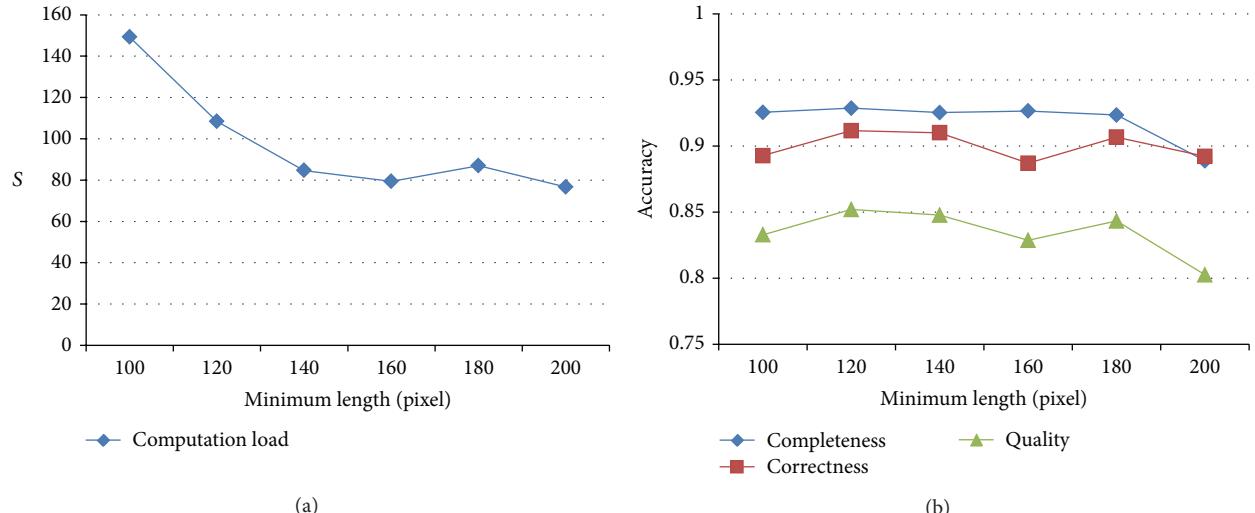


FIGURE 7: The evaluation results of the minimum area influence. (a) Computational load. (b) Accuracy.

[39] for details). Consequently, the kernel density estimator is constructed above as

$$p(x; \Sigma) = \frac{1}{n(2\pi)^{d/2} |\Sigma|^{1/2}} \sum_{i=1}^n \exp\left(-\frac{1}{2} (x - X_i)^T \Sigma^{-1} (x - X_i)\right). \quad (13)$$

In consideration of the numerical stability when solving the principle curve, the logarithm of the density function $p(x; \Sigma)$ is involved:

$$f(p(x)) = \log(p(x)). \quad (14)$$

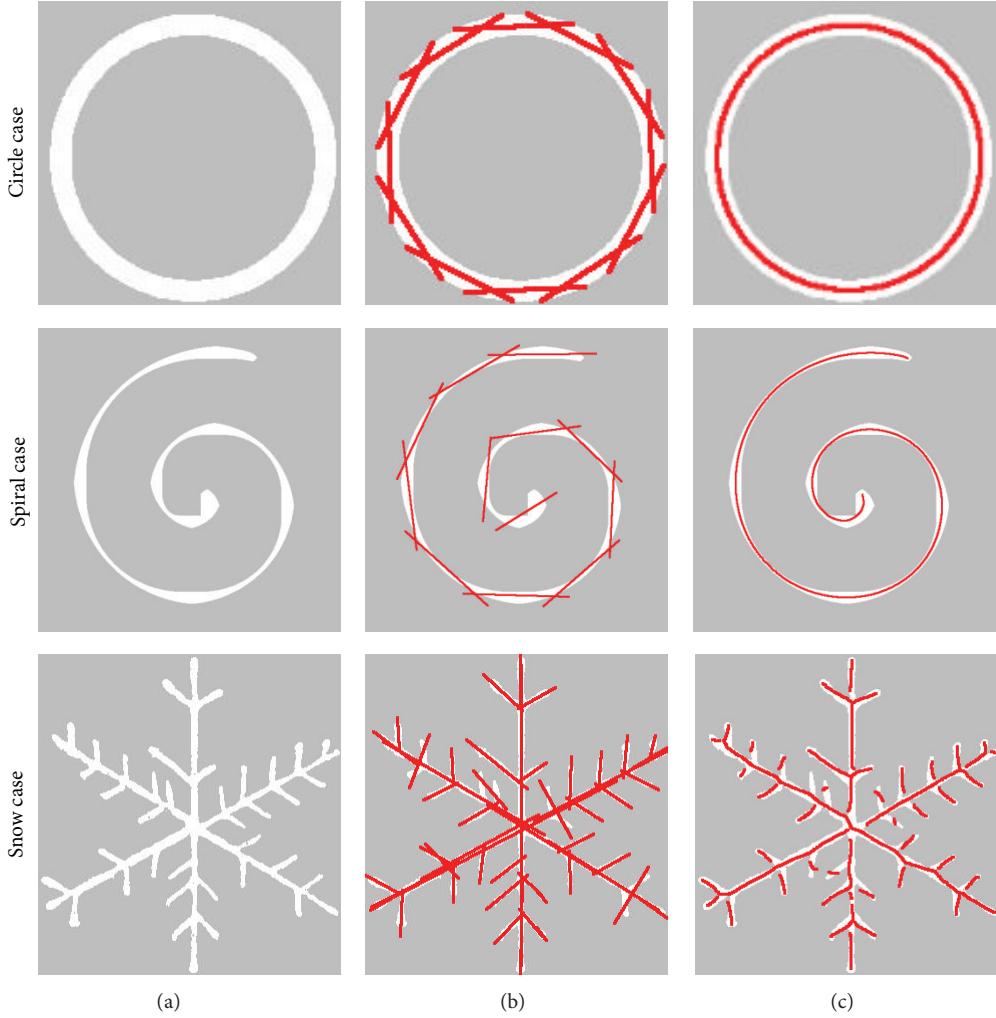


FIGURE 8: (a) The test images. (b) The major axis of GMM partitioning result. (c) The adjustment results using SCMS.

The corresponding gradient and Hessian matrices for KDE are

$$\begin{aligned} g_f(x) &= f'(p(x)) \nabla p(x)^T \\ &= p(x)^{-1} g(x), \end{aligned}$$

$$\begin{aligned} H_f(x) &= f''(p(x)) \nabla p(x)^T \nabla p(x) f'(p(x)) \nabla^2 p(x) \\ &= f''(p(x)) g(x) g(x)^T + f'(p(x)) H(x) \\ &= -p(x)^{-2} g(x) g(x)^T p(x)^{-1} H(x). \end{aligned} \quad (15)$$

Thus, we get the fixed-point update rule for adjusting the coarse initial point:

$$m_{\Sigma}(x) = \frac{\sum_{i=1}^n \exp(-(1/2)(x - X_i)^T \Sigma^{-1}(x - X_i)) X_i}{\sum_{i=1}^n \exp(-(1/2)(x - X_i)^T \Sigma^{-1}(x - X_i))} - x. \quad (16)$$

On account of the similarity of the above iteration formula to the mean-shift algorithm, thus it has the name of subspace constrained mean shift (SCMS) method, which is presented in Algorithm 2. An example of adjusting coarse points to produce accurate centerline utilizing SCMS is illustrated in Figure 5.

3. Experiment

In this section, several experiments that test the proposed method are described. The proposed method is also compared with other methods in the literature to show the advantages and disadvantages of the proposed method. In this study, MATLAB was used as the coding environment on a PC that has an Intel Core2Quad processor with 2.83 GHz clock speed.

3.1. Tests of Parameters. To test the minimum area M_L influence on the road centerline extraction accuracy, we adjusted its value automatically from 5 to 30 with an increment of 5 for each step. The test results are presented in Figure 6. The

```

Input: The coarse road centerline derived from the major axis  $X_1, X_2, \dots, X_n$ 
Output: The adjusted discrete points  $X_1^{\text{new}}, X_2^{\text{new}}, \dots, X_n^{\text{new}}$ 
(1) Provide the band width matrix  $\Sigma^{-1} = n^{2/(d+4)} \hat{\Sigma}^{-1}$  (i.e., the kernel covariance).
(2) for  $i = 1, \dots, K$  do
(3)      $P = I - qq^T$  and  $s = Pm_{\Sigma}(x)$ 
(4)     while  $s < e_{\text{thre}}$  do
(5)          $X_i^{t+1} = X_i^t + s$ 
(6)     end while
(7)      $X_i^{\text{new}} = X_i^{t+1} + s$ 
(8) end for

```

ALGORITHM 2: Subspace constrained mean shift.

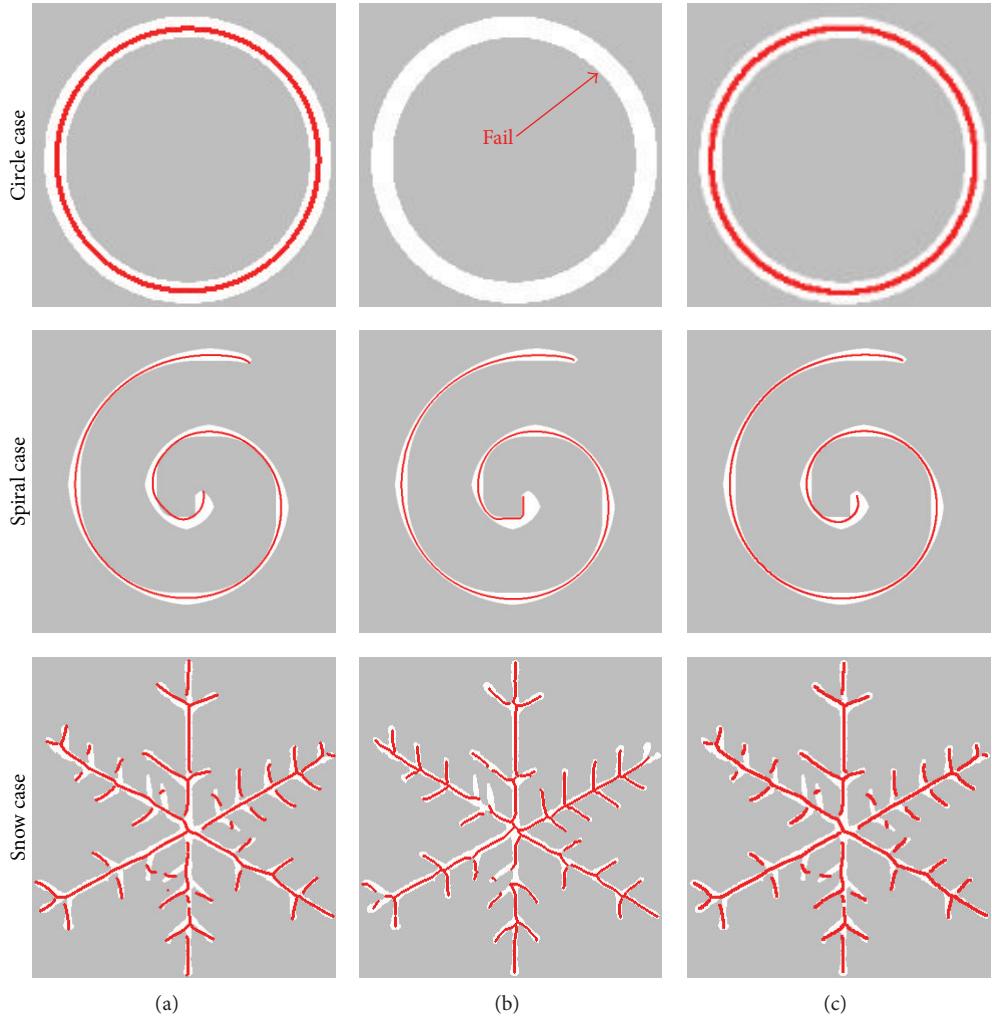


FIGURE 9: (a), (b), and (c) show SCMS, F-SCMS, and GMM-SCMS results, respectively. The centerline extracted are shown in red.

performance of the M_L influence is quantitatively evaluated in terms of the computational time and accuracy. Here, three accuracy measures, (1) Completeness, (2) Correctness, and (3) Quality [40], were used.

$$\text{Correctness} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{Quality} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}},$$

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (17)$$

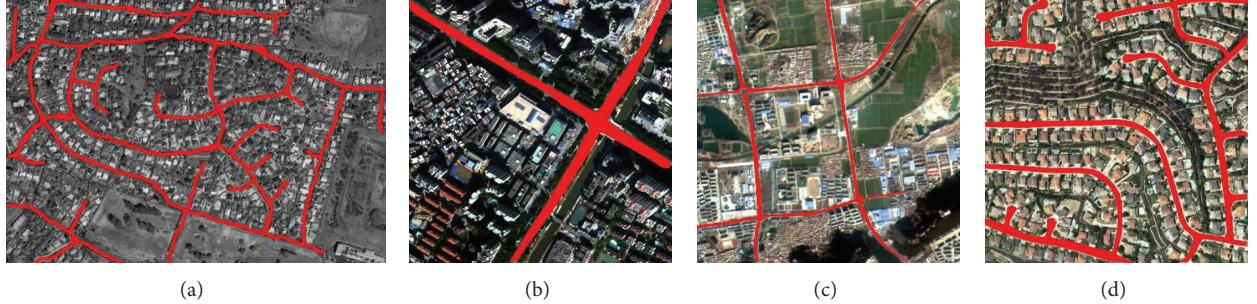


FIGURE 10: (a), (b), (c), and (d) show four test images, respectively. The ground truth datasets are shown in red.

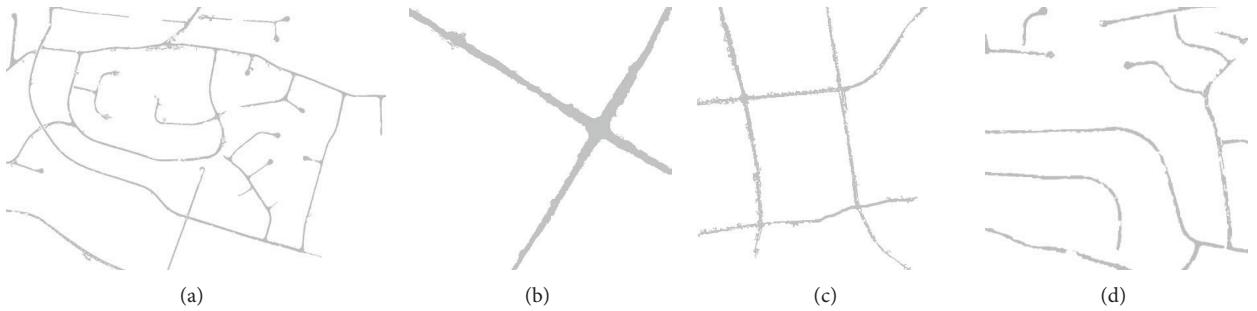


FIGURE 11: (a), (b), (c), and (d) show classified images of the test images in Figure 9, respectively.

where TP, FN, and FP represent true positive, false negative, and false positive, respectively. In this study, the ground truth dataset is produced by hand drawing method and the buffer width is fixed to 8 pixels.

Figure 7 reports quantitative evaluation results with different minimum area values. Figure 7(a) indicates that the computational efficiency steadily improved with the increase of the minimum area value. This is because large M_L will result in initial road centerline with small number of points, which in turn reduce the computational load of the adjustment using SCMS. Figure 7(b) illustrates the influence of M_L on the accuracy. It can be seen that when M_L is increased from 100 to 180, there is a slight change of the accuracy, which indicates that M_L has a weak effect on the accuracy in this range of values. However, the accuracy declines dramatically when M_L is exceeding 180. The reason for this phenomenon is that the large M_L cannot produce the initial road centerline with the correct spatial topology as the classified image, as shown in Figure 6(f). To achieve the best balance between the computational efficiency and the accuracy, this study fixed M_L to 180 throughout the tests.

3.2. Experiments on Simulated Images. The proposed method is tested on three simulated images, including circle image, spiral image, and snow image, as shown in Figure 8(a). The intermediate results using GMM are shown in Figure 8(b) and the corresponding precise centerline results using SCMS are given in Figure 8(c).

The proposed method is compared with two existing methods in literature: (1) SCMS and (2) F-SCMS. The comparison results are illustrated by Figure 9. From Figure 9,

TABLE 1: Comparison of computation time for different centerline extraction methods.

Experiment	Number of points	Computation time (s)		
		SCMS	F-SCMS	GMM-SCMS
Circle case	5953	20.38	Fail	1.76
Spiral case	18828	114.52	11.24	8.53
Snow case	11807	146.78	8.36	16.42

it can be seen that F-SCMS well works in spiral and snow cases but fails to process circle case. By contrast, both SCMS and GMM-SCMS are capable of extracting centerline from three test images, which indicates that SCMS and GMM-SCMS are more general than F-SCMS.

The comparison results are quantitatively evaluated in terms of computation time and three accuracy measures. Table 1 reports the computational time of three methods on test images. It can be seen that F-SCMS and GMM-SCMS generally achieve higher computational efficiency than SCMS. F-SCMS is more efficient than GMM-SCMS in snow case. In spiral case, GMM-SCMS, however, is more efficient than F-SCMS. It seems that the efficiency of GMM-SCMS and F-SCMS is influenced by points number. F-SCMS is more efficient with the large points number while GMM-SCMS in small points number. The accuracy results are given in Table 2. From the practical prospective, circle and spiral cases are too easy and the snow case is more similar to the real-world issue. Therefore, the accuracy is just compared in snow case. It can be seen that F-SCMS produces the highest

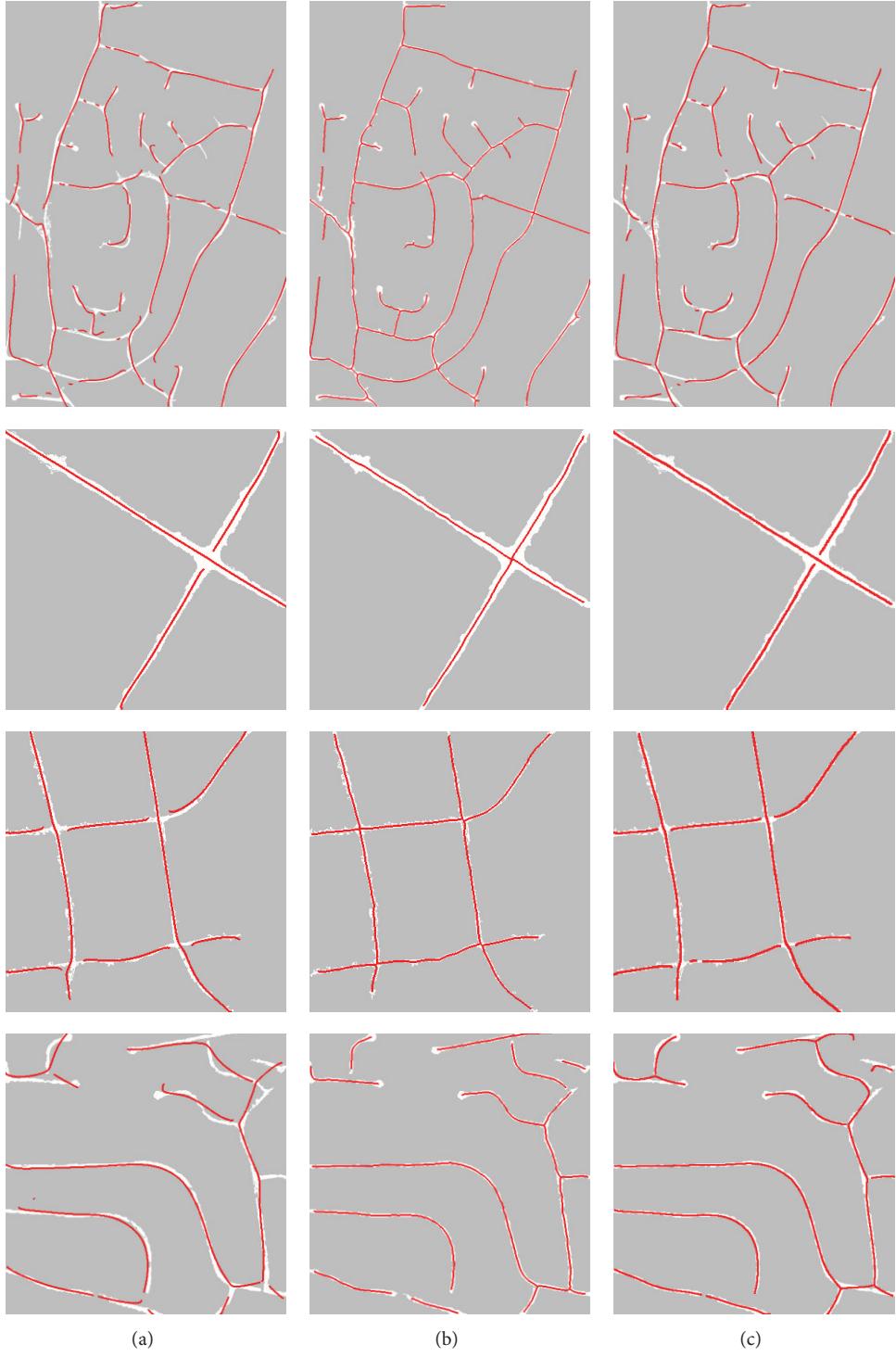


FIGURE 12: Comparison results of different road centerline extraction methods. (a) SCMS results. (b) F-SCMS results. (c) GMM-SCMS results. The centerlines are shown in red for display.

accuracy among these three methods and SCMS and GMM-SCMS achieve a similar accuracy.

3.3. Experiments on Real Satellite Images. The proposed method was tested on four multispectral satellite images used in a previous work [28]. The test images are shown

in Figure 10, where the ground truth datasets are shown in red. The corresponding classified images are presented in Figure 11. The proposed GMM-SCMS method is compared with two existing methods in literature: SCMS and F-SCMS. The comparison results are reported in Figure 12. The performances of three methods are quantitatively evaluated in terms

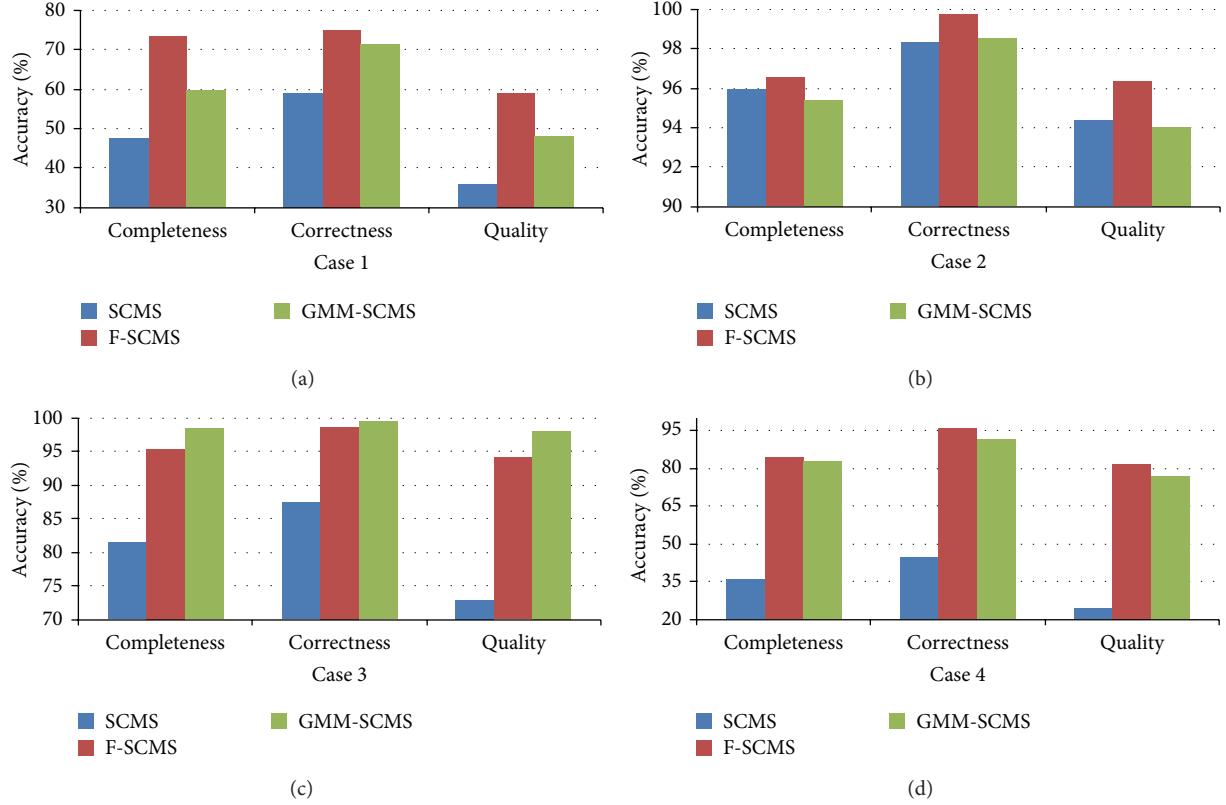


FIGURE 13: (a), (b), (c), and (d) show quantitative evaluation results of different centerline extraction methods for Cases 1, 2, 3, and 4, respectively.

TABLE 2: Quantitative evaluation of different centerline extraction methods.

Method	Circle case			Spiral case			Snow case		
	Completeness (%)	Correctness (%)	Quality (%)	Completeness (%)	Correctness (%)	Quality (%)	Completeness (%)	Correctness (%)	Quality (%)
SCMS	100	100	100	84.13	85.61	73.71	81.03	90.53	74.70
F-SCMS	Fail	Fail	Fail	93.56	94.75	88.95	90.58	98.84	86.92
GMM-SCMS	100	100	100	96.94	97.74	94.82	78.77	92.36	73.96

TABLE 3: Comparison of computation time for different centerline extraction methods.

Experiment	Number of points	Computation time (s)		
		SCMS	F-SCMS	GMM-SCMS
1	35177	1142.95	73.48	110.68
2	11578	58.77	2.18	6.51
3	8009	35.73	1.97	8.84
4	16503	172.99	67.93	15.36

of the computational efficiency and the accuracy, as shown in Table 3 and Figure 13.

From Table 3, it can be seen that the computational load of SCMS is highest among these three methods. By contrast, both F-SCMS and GMM-SCMS produce higher computation efficiency than SCMS. In general, F-SCMS achieves

better computational performance than GMM-SCMS, such as Cases 1–3. However, Table 3 also shows an interesting phenomenon that GMM-SCMS shows higher computational efficiency than F-SCMS [28] in Case 4. The phenomenon indicates that the number of connected components (CCs) influences the F-SCMS performance. For instance, the numbers of CCs in Cases 1–4 are 3, 1, 1, and 8, respectively. It seems that when the number of connected components is high, GMM-SCMS is more efficient than F-SCMS. The main reason is that F-SCMS detects feature points (i.e., end and junction points) and then links these points iteratively for each connected component. Detection of feature points from the classified image costs a lot of time. Therefore, high number of CCs decreases the performance of F-SCMS.

Figure 13 demonstrates the accuracy of three methods. As can be seen, SCMS produces the lowest accuracy while F-SCMS the highest. In general, GMM-SCMS achieves the modest accuracy among these three methods.

By comparing the proposed method with two existing methods on simulated and real datasets, it can be seen that the proposed method produces the best balance between the computational complexity, accuracy, and generalization ability among these three methods. Thus, GMM-SCMS is more practical for accurate and efficient road centerline extraction from classified images.

4. Conclusions

In conclusions, an automatic method, so-called Gaussian mixture model based subspace constrained mean shift (GMM-SCMS), was developed for road centerline extraction from the classified image. The GMM-SCMS method can extract smooth road centerlines with high accuracy. Benefited from GMM, the proposed method can represent the classified images using a small number of points while retaining the correct spatial topology, which in turn improves the computational efficiency. By using SCMS, the approximate points are iteratively adjusted to produce the accurate road centerline.

The proposed method was compared with cutting-edge SCMS and feature points-based SCMS (F-SCMS). The experimental results indicate that (1) GMM-SCMS is more efficient than SCMS and (2) GMM-SCMS has a higher generalization ability than F-SCMS. In other words, GMM-SCMS can process the classified image with arbitrary shape. By full consideration of efficiency, accuracy, and generalization ability, it can be concluded that GMM-SCMS can provide a practical solution for delineating road centerlines from classified images.

Conflict of Interests

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

Acknowledgments

The authors would like to thank Professor D. Erdogmus from Northwestern University for his kind help on the subspace constrained mean shift method, Prof. H. Mayer from the Universität der Bundeswehr München for kind discussion on road extraction, and Dr. C. Ünsala from Yeditepe University for kindly providing GeoEye-1 images. This work was supported in part by the National Natural Science Foundation of China (41201451, 40901214), Ministry of Science and Technology of China (Project nos. 2012BAJ15B04, 2012AA12A305).

References

- [1] J. B. Mena, "State of the art on automatic road extraction for GIS update: a novel classification," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 3037–3058, 2003.
- [2] S. Das, T. T. Mirnalinee, and K. Varghese, "Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3906–3931, 2011.
- [3] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on parti-tional clustering techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2973–2987, 2009.
- [4] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
- [5] M. Song and D. Civco, "Road extraction using SVM and image segmentation," *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 12, pp. 1365–1371, 2004.
- [6] Z. Miao, W. Shi, H. Zhang, and X. Wang, "Road center-line extraction from high-resolution imagery based on shape features and multivariate adaptive regression splines," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 583–587, 2013.
- [7] X. Huang and L. Zhang, "Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines," *International Journal of Remote Sensing*, vol. 30, no. 8, pp. 1977–1987, 2009.
- [8] W. Shi, Z. Miao, and J. Debayle, "An integrated method for urban main-road centerline extraction from optical remotely sensed imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3359–3372, 2014.
- [9] S. Valero, J. Chanussot, J. A. Benediktsson, H. Talbot, and B. Waske, "Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images," *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1120–1127, 2010.
- [10] W. Shi, Z. Miao, Q. Wang, and H. Zhang, "Spectral-spatial classification and shape features for urban road centerline extraction," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 788–792, 2014.
- [11] Y. Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 296–308, 2015.
- [12] Y.-Q. Zhao, L. Zhang, and S. G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 747–756, 2011.
- [13] D. Chaudhuri, N. K. Kushwaha, and A. Samal, "Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 5, pp. 1538–1544, 2012.
- [14] D. Guo, A. Weeks, and H. Klee, "Robust approach for suburban road segmentation in high-resolution aerial images," *International Journal of Remote Sensing*, vol. 28, no. 2, pp. 307–318, 2007.
- [15] A. Mohammadzadeh and M. J. V. Zanjani, "A Self-Organizing Fuzzy Segmentation (SOFS) method for road detection from high resolution satellite images," *Photogrammetric Engineering and Remote Sensing*, vol. 76, no. 1, pp. 27–35, 2010.
- [16] W. Wei and Y. Xin, "Feature extraction for man-made objects segmentation in aerial images," *Machine Vision and Applications*, vol. 19, no. 1, pp. 57–64, 2008.
- [17] I. Laptev, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner, "Automatic extraction of roads from aerial images based on scale space and snakes," *Machine Vision and Applications*, vol. 12, no. 1, pp. 23–31, 2000.
- [18] L. Bentabet, S. Jodouin, D. Ziou, and J. Vaillancourt, "Road vectors update using SAR imagery: a snake-based method,"

- [18] *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1785–1803, 2003.
- [19] K. Karantzalos and D. Argialas, “A region-based level set segmentation for automatic detection of man-made objects from aerial and satellite images,” *Photogrammetric Engineering and Remote Sensing*, vol. 75, no. 6, pp. 667–677, 2009.
- [20] M. Rochery, I. H. Jermyn, and J. Zerubia, “Higher order active contours,” *International Journal of Computer Vision*, vol. 69, no. 1, pp. 27–42, 2006.
- [21] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Prentice-Hall, 2003.
- [22] P. Doucette, P. Agouris, A. Stefanidis, and M. Musavi, “Self-organised clustering for road extraction in classified imagery,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 55, no. 5-6, pp. 347–358, 2001.
- [23] C. Poullis and S. You, “Delineation and geometric modeling of road networks,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 2, pp. 165–181, 2010.
- [24] Q. Zhang and I. Couloigner, “Accurate centerline detection and line width estimation of thick lines using the radon transform,” *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 310–316, 2007.
- [25] U. Ozertem and D. Erdogmus, “Nonparametric snakes,” *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2361–2368, 2007.
- [26] U. Ozertem and D. Erdogmus, “Principal curve time warping,” *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2041–2049, 2009.
- [27] U. Ozertem and D. Erdogmus, “Locally defined principal curves and surfaces,” *Journal of Machine Learning Research*, vol. 12, pp. 1249–1286, 2011.
- [28] Z. Miao, B. Wang, W. Shi, and H. Wu, “A method for accurate road centerline extraction from a classified image,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014.
- [29] S. Movaghati, A. Moghaddamjoo, and A. Tavakoli, “Road extraction from satellite images using particle filtering and extended Kalman filtering,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2807–2817, 2010.
- [30] J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka, “Road network extraction and intersection detection from aerial images by tracking road footprints,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4144–4157, 2007.
- [31] A. P. Dal Poz, R. A. B. Gallis, J. F. C. da Silva, and É. F. O. Martins, “Object-space road extraction in rural areas using stereoscopic aerial images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, pp. 654–658, 2012.
- [32] F. Dell’Acqua and P. Gamba, “Detection of urban structures in SAR images by robust fuzzy clustering algorithms: the example of street tracking,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2287–2297, 2001.
- [33] M. Negri, P. Gamba, G. Lisini, and F. Tupin, “Junction-aware extraction and regularization of urban road networks in high-resolution SAR images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2962–2971, 2006.
- [34] Y.-W. Choi, Y.-W. Jang, H.-J. Lee, and G.-S. Cho, “Three-Dimensional LiDAR data classifying to extract road point in urban area,” *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 725–729, 2008.
- [35] S. Clode, F. Rottensteiner, P. Kootsookos, and E. Zelniker, “Detection and vectorization of roads from lidar data,” *Photogrammetric Engineering and Remote Sensing*, vol. 73, no. 5, pp. 517–535, 2007.
- [36] P. Kumar, C. P. McElhinney, P. Lewis, and T. McCarthy, “An automated algorithm for extracting road edges from terrestrial mobile LiDAR data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 85, pp. 44–55, 2013.
- [37] R. B. Gomez, “Hyperspectral imaging: a useful technology for transportation analysis,” *Optical Engineering*, vol. 41, no. 9, pp. 2137–2143, 2002.
- [38] S. Roessner, K. Segl, U. Heiden, and H. Kaufmann, “Automated differentiation of urban surfaces based on airborne hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1525–1532, 2001.
- [39] I. Ahamada and E. Flachaire, *Non-Parametric Econometrics*, Oxford University Press, Oxford, UK, 2010.
- [40] C. Wiedemann, C. Heipke, and H. Mayer, “Empirical evaluation of automatically extracted road axes,” in *Proceedings of the CVPR Workshop on Empirical Evaluation Techniques in Computer Vision*, Los Alamitos, Calif, USA, 1998.