

Noise-Robust Hyperspectral Image Classification via Multi-Scale Total Variation

Puhong Duan^{ID}, *Student Member, IEEE*, Xudong Kang^{ID}, *Senior Member, IEEE*, Shutao Li^{ID}, *Fellow, IEEE*, and Pedram Ghamisi^{ID}, *Senior Member, IEEE*

Abstract—In this paper, a novel multi-scale total variation method is proposed to extract structural features from hyperspectral images (HSIs), which consists of the following steps. First, the spectral dimension of the HSI is reduced with an averaging-based method. Then, the multi-scale structural features (MSFs), which are insensitive to image noise, are constructed with a relative total variation-based structure extraction technique. Finally, the MSFs are fused together using kernel principal component analysis (KPCA), so as to obtain the KPCA-fused MSFs for classification. Experimental results on three publicly available hyperspectral datasets, including both well-known, long-used data, and a recent dataset obtained from an international contest, demonstrate the competitive performance over several state-of-the-art classification approaches in this field. Moreover, the robustness of the proposed method to the small-sample-size problem and serious image noise is also demonstrated.

Index Terms—Dimension reduction, hyperspectral image (HSI) classification, kernel principal component analysis (KPCA), multi-scale total variation (MSTV).

I. INTRODUCTION

WITH the development of hyperspectral sensors, rich spectral and spatial information of the land covers can be easily captured with hyperspectral imaging, which has attracted much attention in various remote sensing applications [1]–[6]. Among these applications, supervised hyperspectral image (HSI) classification [7]–[9] is the most active topic based on the statistics reported in [10].

Over the past decades, a number of techniques have been developed for HSI classification [11]–[15], such as support vector machine (SVM) [11], multi-nomial logistic regression [12], random forests [16], and sparse representation [17]. These classifiers make full use of the spectral similarity among the training and test samples so as to discriminate different land covers. However, when only a limited number of labeled samples are available, most of these pixel-wise classifiers are not able to obtain satisfactory classification accuracies due to the “Hughes phenomenon.” Moreover, the classification process is usually computationally expensive due to the high spectral dimension of HSIs.

In order to overcome these problems, many feature extraction methods have been developed to reduce the spectral dimensionality while preserving the spectral discriminability [18]–[20]. For example, some classical feature extraction techniques, e.g., principal component analysis (PCA) [21], independent component analysis (ICA) [22], and linear discriminant analysis (LDA) [23], have been successfully applied in commercial remote sensing software. However, these feature extraction methods have not considered the spatial correlation among adjacent pixels, and thus, easily produce “noise”-like mislabels in the classification results.

In order to further improve the classification performance, spectral-spatial classification approaches using the spectral and spatial information of HSI have been studied extensively [10], [24]–[29]. Spectral-spatial feature extraction has been demonstrated to be one of the most effective ways in increasing the class discriminability of different ground objects. For example, Benediktsson *et al.* employed extended morphological profiles to exact spatial features from HSI by using morphological opening and closing operations [30]. Mura *et al.* used morphological attribute profiles to exact a multi-level characterization of HSIs so as to model the structural information of different kinds [31]. In recent years, various improved versions of morphological attribute profiles have been proposed, which lead to a further improvement in terms of classification accuracies [32]–[35]. Kang *et al.* proposed an edge-preserving filtering (EPF)-based method to extract the edge features of HSIs while improving the pixel separability [36]. Ghamisi *et al.* proposed extinction profiles for the classification of remote sensing data, in which a sequence of thinning and thickening transformations are used to extract multi-level decomposition of input images [37]. Besides the spectral-spatial classification methods mentioned above, other effective tools have also been proposed for HSI classification,

Manuscript received March 4, 2019; revised April 11, 2019; accepted April 27, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 61601179, 6187119, and 61890962, in part by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001, in part by the Science and Technology Plan Project Fund of Hunan Province under Grants 2017RS3024 and 2018TP1013, and in part by the Science and Technology Talents Program of Hunan Association for Science and Technology under Grant 2017TJ-Q09. (*Corresponding author: Xudong Kang*)

P. Duan, X. Kang, and S. Li are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Changsha 410082, China (e-mail: puhong_duan@hnu.edu.cn; xudong_kang@163.com; shutao_li@hnu.edu.cn).

P. Ghamisi is with the Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, 01328 Dresden, Germany (e-mail: p.ghamisi@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2019.2915272

such as tensor representation [38], intrinsic image decomposition [39], spatial kernel-based methods [40], and anisotropic diffusion [41]. All of the mentioned publications have demonstrated that the combination of spectral and spatial information in the feature extraction process is indeed an effective way in enhancing classification performance.

Furthermore, total variation techniques have been employed for classification of HSIs. In [42], Rasti *et al.* proposed an orthogonal total variation component analysis (OTVCA) for feature extraction of HSIs. In [43], Sun *et al.* introduced a spatially adaptive total variation regularization to enforce a spatially smooth classifiers. However, existing spectral-spatial feature extraction methods still face several challenges, which are listed as follows.

- 1) Most of the existing feature extraction methods have not considered the image noise that may be produced due to low-lighting or other factors. In such imperfect situations, the performance of existing spectral-spatial feature extraction methods tends to decrease seriously. Although various image denoising methods have been proposed, these methods may cause the loss of structural information, and bring extra computing burdens.
- 2) It is known that HSIs actually contain complex texture patterns, which may increase the spectral variability of pixels belonging to the same land cover. For example, spectral variations exist for pixels belonging to Tree class due to its complex textures. This phenomenon is very common when classifying HSIs with high spatial resolutions, which leads to a decrease in classification performance.

To address these challenges, a noise-robust HSI classification method is proposed based on multi-scale total variation (MSTV). Specifically, the proposed method consists of the following steps: First, the spectral dimension of the HSI is reduced with an averaging-based method. Then, the multi-scale structural features (MSFs) of the dimension-reduced image are constructed by a relative total variation (RTV) method with different parameter settings. Finally, the MSFs are stacked together and kernel principal component analysis (KPCA) is applied on the stacked MSFs so as to magnify the separability of pixels belonging to different classes. Generally, the major advantages of this paper can be concluded as follows.

- 1) RTV-based technique is explored for the first time to extract structural features from HSIs. Since RTV can decrease the influence of textures and noise, the inner-class differences can be decreased, whereas the inter-class discrepancy can be increased effectively.
- 2) An MSTV framework is introduced so as to make full use of the multi-scale information in HSIs. In the proposed framework, KPCA method is found to be an optimal solution to fuse MSFs.
- 3) Experimental results performed on several real hyperspectral datasets demonstrate the competitive performance of the proposed method with respect to several state-of-the-art methods. More importantly, the proposed method shows obvious advantages when there are only a limited number of labeled samples or the HSI is corrupted by serious image noise. These advantages make the proposed method highly effective in real applications.

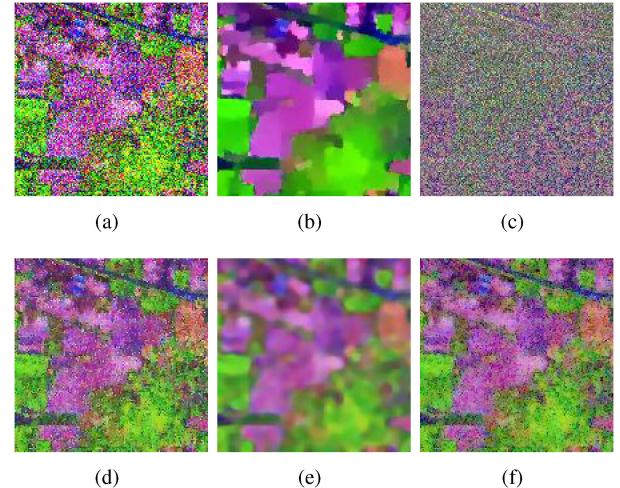


Fig. 1. Intuitive example of structure extraction. (a) Input image. (b) Structural component obtained with the RTV method [44]. (c) Texture component obtained by subtracting the structural image from the input image. (d) Resulting image obtained by recursive filtering [45]. (e) Resulting image obtained by guided filtering [46]. (f) Resulting image obtained by weighted least-squares filtering [47].

The reminder of this paper is organized as follows. The principles of structure extraction and KPCA are briefly reviewed in Section II. The proposed method is introduced in Section III. Experimental results on three real HSIs are given in Section IV. Finally, conclusions are presented in Section V.

II. BACKGROUND

A. Structure Extraction

In psychology, one perceiving function of human vision is noted by the scientist Arnheim as follows: *The overall structural features are the primary data of human perception, not the individual details* [48]. On the basis of Arnheim's observations, many structure extraction techniques have been extensively researched [44], [49]–[51]. Although EPF is designed to suppress details and preserve image edges, it cannot achieve satisfactory texture removal. Different from EPF, the objective of these structure extraction methods is to decompose input image into a structure component and a texture component, in which the structure component can well characterize the significant spatial structures of objects available in the input image. As shown in Fig. 1, it can be observed that the EPFs-based methods fail to extract the major structural feature, while the structure extraction technique can well separate the structures and textures from the complex scene.

In recent years, an RTV-based structure extraction method is proposed, which shows outstanding performances in the computer vision field [44]. Our method is the first attempt to develop the RTV method for constructing the MSFs of HSIs because of the following reasons. First, the RTV method does not need any previous knowledge such as the specific type of texture and main structure arrangement. Second, there are no complex mathematical models for the optimization-based method, and thus, it can be easily extended to process HSIs, which contain hundreds of

bands. Here, the RTV-based structure extraction method will be described briefly. For the comprehensive description of the RTV optimization algorithm, we refer readers to [44].

Formally, given an input image \mathbf{I} , the structural image \mathbf{S} can be obtained by solving the following energy function:

$$\arg \min_{\mathbf{S}} \sum_{i=1}^T (\mathbf{S}_i - \mathbf{I}_i)^2 + \lambda \cdot \left(\frac{\mathcal{D}_x(i)}{\mathcal{L}_x(i) + \varepsilon} + \frac{\mathcal{D}_y(i)}{\mathcal{L}_y(i) + \varepsilon} \right) \quad (1)$$

where T is the total number of image pixels. \mathbf{S} is the resulting image. λ is a weight that can control the smoothness of the resulting image. ε is a small positive number to avoid division by zero. On one hand, the term $(\mathbf{S} - \mathbf{I})^2$ is to make sure that the input and output should be as similar as possible. On the other hand, the term $(\frac{\mathcal{D}_x(i)}{\mathcal{L}_x(i) + \varepsilon} + \frac{\mathcal{D}_y(i)}{\mathcal{L}_y(i) + \varepsilon})$ ensures that the textures should be decreased while preserving the major structures.

More detailedly, \mathcal{D}_x and \mathcal{D}_y are the windowed total variations in the x and y directions, respectively, which capture the absolute spatial differences within the window $R(i)$. A higher \mathcal{D} indicates that the texture information in a local window is salient

$$\begin{aligned} \mathcal{D}_x(i) &= \sum_{j \in R(i)} g_{i,j} \cdot |(\partial_x \mathbf{S})_j| \\ \mathcal{D}_y(i) &= \sum_{j \in R(i)} g_{i,j} \cdot |(\partial_y \mathbf{S})_j| \end{aligned} \quad (2)$$

where j belongs to $R(i)$. $\partial_x \mathbf{S}$ and $\partial_y \mathbf{S}$ represent the partial derivatives in x and y directions of image \mathbf{S} , respectively. $g_{i,j}$ is a weighting function, which is expressed as

$$g_{i,j} \propto \exp \left(-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\sigma^2} \right) \quad (3)$$

where σ denotes the spatial scale of the window.

Different from \mathcal{D}_x and \mathcal{D}_y , \mathcal{L}_x and \mathcal{L}_y are the windowed inherent variations in the x and y directions, respectively, which calculate the overall spatial variations. A higher \mathcal{L} value indicates that the structural information in a local window is salient

$$\begin{aligned} \mathcal{L}_x(i) &= \sum_{j \in R(i)} |g_{i,j} \cdot (\partial_x \mathbf{S})_j| \\ \mathcal{L}_y(i) &= \sum_{j \in R(i)} |g_{i,j} \cdot (\partial_y \mathbf{S})_j|. \end{aligned} \quad (4)$$

The effect of texture removal is obtained by the regularizer $(\frac{\mathcal{D}_x(i)}{\mathcal{L}_x(i) + \varepsilon} + \frac{\mathcal{D}_y(i)}{\mathcal{L}_y(i) + \varepsilon})$, named as the RTV map. In the RTV map, the meaningful structures are penalized much less than textures, and thus, it is very effective to preserve significant structures (see Fig. 1). In this paper, the RTV-based structure extraction is referred as RTV($\mathbf{I}, \lambda, \sigma$), in which \mathbf{I} is the input image, and λ and σ are the parameters adjusting the degree of smoothness and spatial scale, respectively.

B. Kernel Principal Component Analysis

KPCA [52] is a classical multivariate data analysis method, which has been applied in many fields of information processing.

The main idea is to map the original data space \mathbf{R}^D into a high-dimensional feature space \mathcal{H} using a kernel mapping function Φ , and then, the mapped data are processed by PCA in the high-dimensional feature space \mathcal{H} .

Suppose a data matrix $\mathbf{S} = \{x_1, x_2, \dots, x_m\} \in \mathbf{R}^D$ and a non-linear mapping function $\Phi : \mathbf{R}^D \rightarrow \mathcal{H}$, $x_i \rightarrow \Phi(x_i)$. In this paper, the Gaussian kernel is adopted in KPCA. The non-linear transform KPCA can be expressed as follows:

$$\mathbf{P} = \mathbf{W}^T \Phi(\mathbf{S}) \quad (5)$$

where $\Phi(\mathbf{S})$ is the transformed data in the feature space \mathcal{H} . \mathbf{W} is the transform matrix, which can be obtained as follows:

$$\mu \mathbf{V} = \mathbf{C}^\Phi \mathbf{V} \quad (6)$$

where \mathbf{C}^Φ is the covariance matrix of the feature space \mathcal{H} . μ and \mathbf{V} denote the eigenvalues and eigenvectors of the covariance matrix \mathbf{C}^Φ , respectively. The fused feature \mathbf{W} is constructed by the eigenvectors \mathbf{V} . In this paper, in order to describe the proposed method in a simple way, the fused feature is referred as KPCA(\mathbf{S}, N) in the following sections, where N is the number of the selected principal components.

III. PROPOSED METHOD

The schematic of the proposed method is given in Fig. 2, which consists of the following major steps: First, the averaging-based method is applied to the original HSI to reduce the dimensionality. Then, the MSFs are constructed using the RTV-based structure extraction method with different parameters. Finally, the MSFs are stacked together and fused with the KPCA.

A. Spectral Dimension Reduction

In order to reduce the computational complexity and the influence of image noise, the spectral dimension of the original HSI is first reduced with the averaging-based method. More specifically, the M -dimensional HSI \mathbf{I} is first partitioned into K groups of subsets with equal size along the spectral dimension, in which K is the number of bands in the dimension-reduced image. The k th ($k \in \{1, 2, \dots, K\}$) subset can be obtained as follows:

$$\mathbf{P}^k = \begin{cases} (\mathbf{I}_{(k-1)\lfloor M/K \rfloor + 1}, \dots, \mathbf{I}_{k\lfloor M/K \rfloor}), & \text{if } k \lfloor M/K \rfloor \leq M \\ (\mathbf{I}_{K-\lfloor M/K \rfloor + 1}, \dots, \mathbf{I}_M), & \text{if } k \lfloor M/K \rfloor > M \end{cases} \quad (7)$$

where $\lfloor M/K \rfloor$ denotes the largest integer not greater than M/K .

Then, the contiguous bands in the k th subset are fused by the averaging rule. The mathematical expression can be described as follows:

$$\mathbf{Q}^k = \frac{\sum_{n=1}^{N_k} \mathbf{P}_n^k}{N_k} \quad (8)$$

where \mathbf{Q} is the dimension-reduced hyperspectral data. N_k denotes the total number of bands in the k th subset, and \mathbf{P}_n^k represents the n th band of the k th subset.

Compared with other dimension reduction methods such as the transform-based and feature selection based methods, the

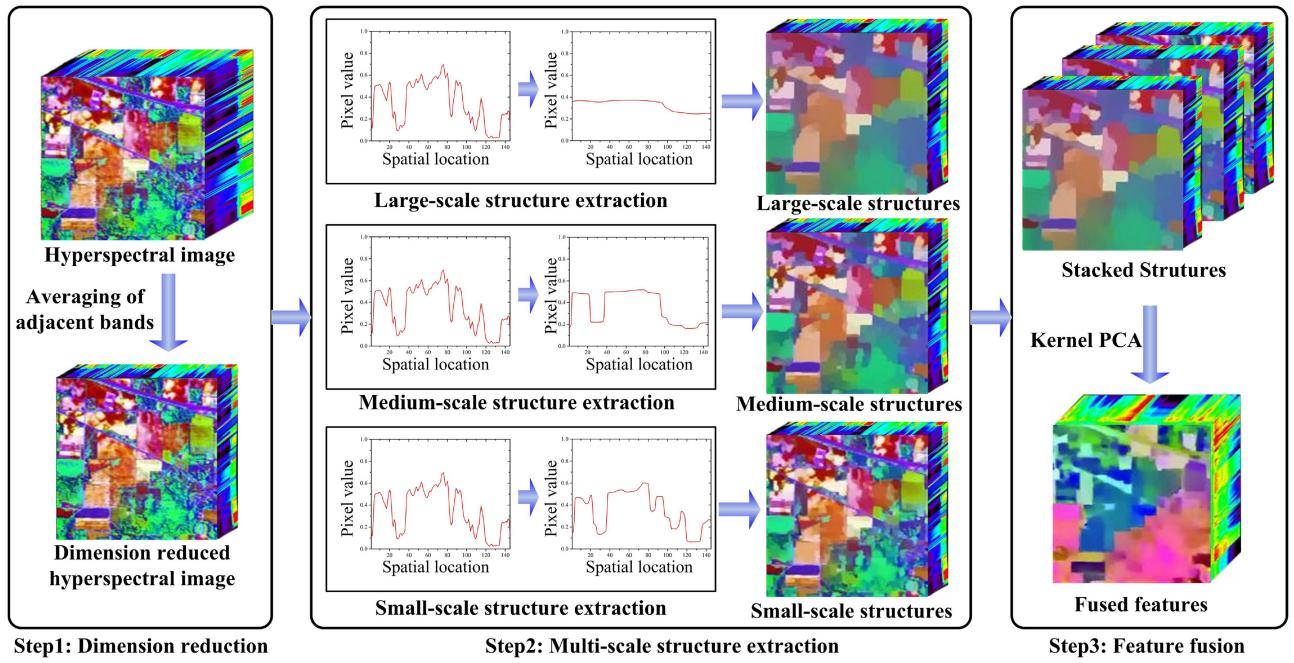


Fig. 2. Schematic of the proposed MSTV method.

averaging-based dimension reduction method has two advantages: First, the pixels of the dimension-reduced image are still highly related to the spectral information of the original image. Second, the averaging-based method can be implemented very efficiently. Therefore, the averaging-based method is adopted in this paper.

B. Multi-Scale Structure Extraction

Since the captured scene in an HSI often consists of different land covers with different scales, a single-scale structural feature cannot provide a complete characterization of objects at different scales. In this paper, MSFs are first constructed by using RTV with different parameter settings on the dimension-reduced HSI. Then, the MSFs can be stacked together as follows:

$$\mathbf{F}_l = \text{RTV}(\mathbf{Q}, \lambda_l, \sigma_l), l \in \{1, 2, \dots, L\} \quad (9)$$

$$\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_L\} \quad (10)$$

where λ_l and σ_l are the l th parameters adopted in the structure extraction. \mathbf{Q} is the dimension-reduced HSI, and \mathbf{F} denotes the stacked MSFs.

The basic motivation behind this step is that the sizes of land covers in HSI vary from small regions with a few pixels to large areas containing thousands of pixels, and the captured scene usually consists of different objects with various sizes. In order to make full use of the rich spatial information in HSI and improve the robustness of the structural feature to various objects or scenes, the multi-scale structural features are introduced in this paper. Fig. 3 presents the performance of RTV method with different parameter settings. Based on these results, the following observations can be found. First, the RTV method can

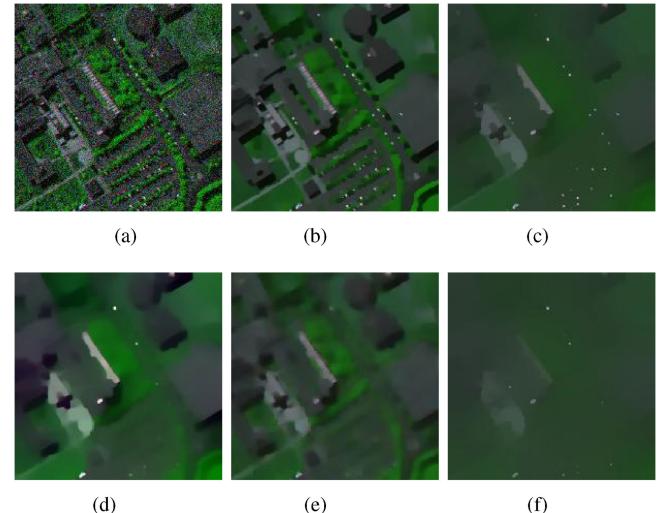


Fig. 3. Influence of the two parameters, i.e., λ and σ to the resulting structure features. (a) False color composite of the input HSI. (b)–(f) Structural features obtained by the RTV method with different parameter settings. (a) Input image. (b) $\lambda = 0.003, \sigma = 2$. (c) $\lambda = 0.02, \sigma = 1$. (d) $\lambda = 0.01, \sigma = 3$. (e) $\lambda = 0.003, \sigma = 9$. (f) $\lambda = 0.04, \sigma = 1$.

effectively remove the image noise and texture, which improves the spectral purity among pixels. Furthermore, the significant structures belonging to objects at different scales are able to be well preserved. Moreover, different structural features have complementary information. Therefore, the complementary information in the stacked MSFs is expected to be beneficial for improving the classification accuracy.

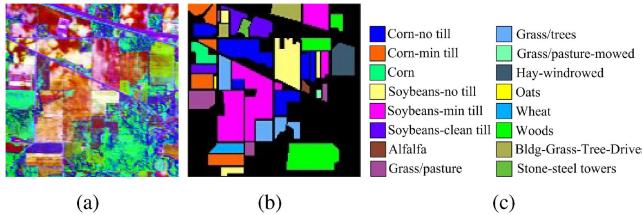


Fig. 4. Indian Pines dataset. (a) False color composite image. (b) Reference data. (c) Class names.

C. Feature Fusion With the KPCA

The stacked MSFs have a high spectral dimension ($K \times L$ spectral dimensions) and contain a great deal of redundant information, which easily increases the computing cost and leads to the Hughes phenomenon. In addition, although the spectral differences belonging to the same land cover tend to be reduced and texture information has been effectively removed, the structure extraction may decrease the spectral differences among pixels belonging to different land covers. In order to solve these problems, the KPCA is taken into account to fuse the MSFs in this paper, since the KPCA is quite an effective algorithm in both increasing the spectral separability of pixels and reducing the spectral dimension

$$\mathbf{P} = \text{KPCA}(\mathbf{F}, N) \quad (11)$$

where N is the number of principal components retained in \mathbf{P} , and \mathbf{F} is the stacked structural features. The SVM is robust to the variation of feature dimensions, and many literatures have demonstrated its effectiveness. Therefore, in this paper, the SVM is adopted to evaluate the classification performance of the resulting features \mathbf{P} .

IV. EXPERIMENTS

A. Experimental Setup

1) *Datasets*: Three real hyperspectral datasets, i.e., Indian Pines, University of Pavia, and Houston, are utilized in our experiments.

The Indian Pines image was acquired by the Airborne Visible/Infrared Imaging Spectrometer sensor with a spectral coverage ranging from 0.4 to 2.5 μm , which covers the agricultural Indian Pines test site in Northwestern Indiana. This image contains 200 spectral reflectance bands of size 145×145 pixels with a spatial resolution 20 m per pixel. The false color composite image and the reference data, which contains 16 land cover classes, are presented in Fig. 4.

The University of Pavia image was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-03) optical sensor with a spectral coverage ranging from 0.43 to 0.86 μm , which covers the campus at the University of Pavia, Pavia, Italy. This image contains 103 bands of size 610×340 pixels with a spatial resolution 1.3 m per pixel. The false color composite image and the reference data, which contains nine classes of interest, are presented in Fig. 5.

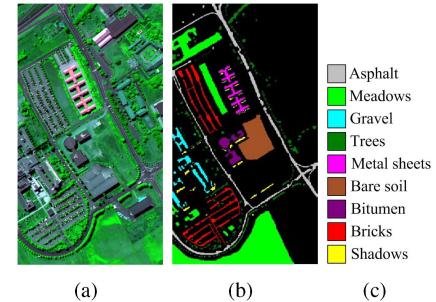


Fig. 5. University of Pavia dataset. (a) False color composite image. (b) Reference data. (c) Class names.

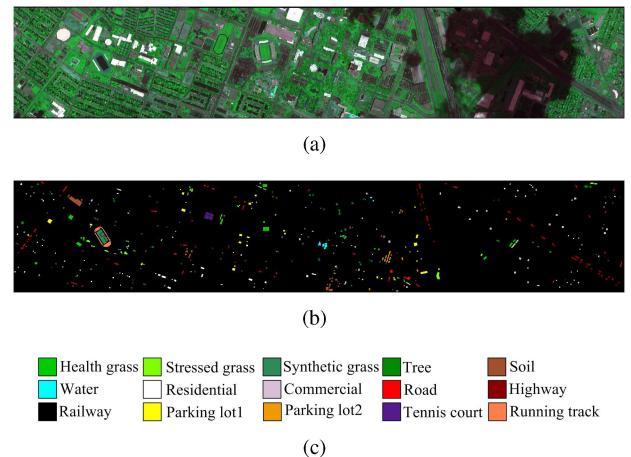


Fig. 6. Houston dataset. (a) False color composite image. (b) Reference data. (c) Class names.

The Houston image was distributed by the 2013 GRSS Data Fusion Contest. This image was acquired by the compact airborne spectrographic imager over the University of Houston campus and the neighboring urban area on June 23, 2012. Furthermore, it contains 144 bands of size 349×1905 pixels with a spatial resolution of 2.5 m, and the spectral coverage ranges from 0.38 to 1.05 μm . The false color composite image and the reference data, which contain 15 classes of interest, are presented in Fig. 6.

2) *Quantitative Indexes*: In order to evaluate the performance of HSI classification, four widely used quantitative indexes, i.e., individual class accuracy (CA), overall accuracy (OA), average accuracy (AA), and Kappa coefficient, are adopted to evaluate the classification performance. Specifically, the CA is the percentage of correctly classified pixels for each class. The OA indicates the percentage of correctly classified pixels. The AA measures the average value of the percentage of correctly classified pixels for each class. The Kappa coefficient estimates the percentage of correctly classified pixels corrected by the number of agreements that would be expected purely by chance. All experiments have been repeated ten times with randomly selected training samples so as to obtain the mean and standard deviation of CA, OA, AA, and Kappa coefficient. Moreover, the reported value of each quality index, i.e., CA, OA, AA,

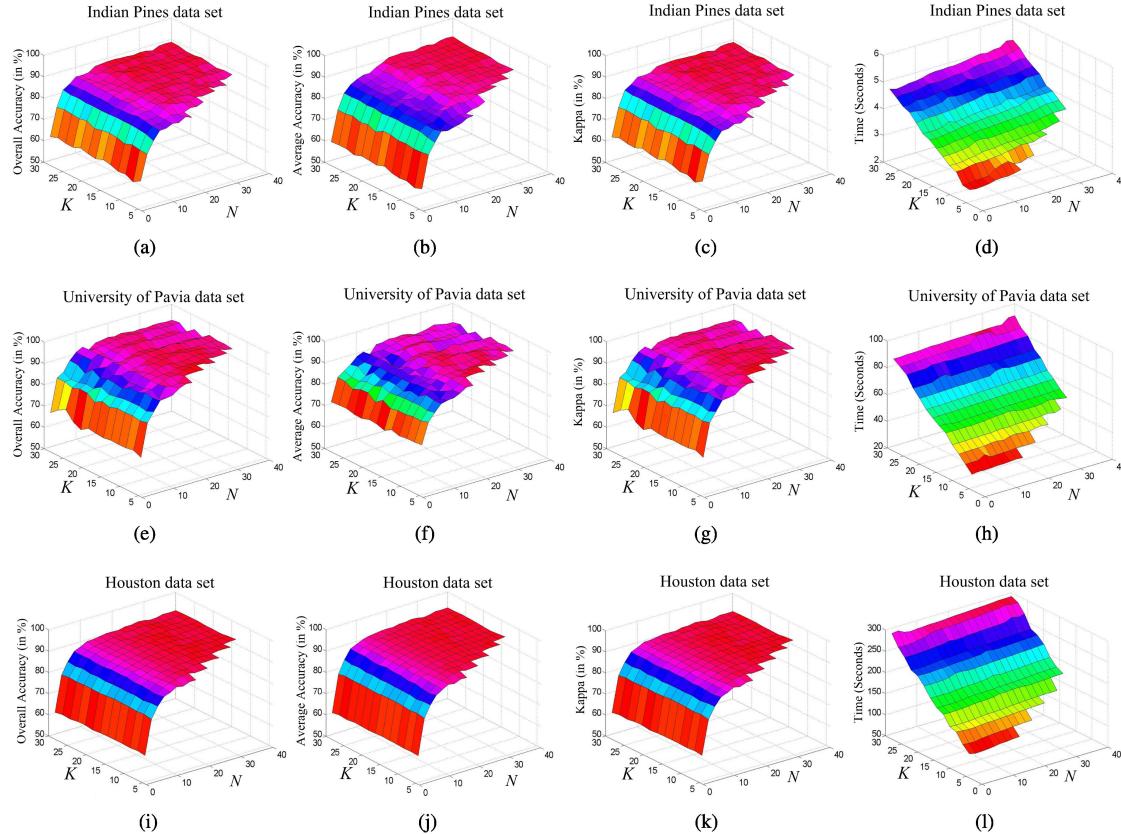


Fig. 7. Influence of parameters K and N to the performance of the proposed method on the three datasets. From top to bottom are the results for the Indian Pines, University of Pavia, and Houston datasets, respectively. Each row denotes the overall accuracy (first column), average accuracy (second column), Kappa coefficient (third column), and computing time (fourth column) of the proposed method.

and Kappa, is calculated by averaging the results obtained by ten repeated experiments with randomly selected training sets.

B. Analysis of the Influence of Parameters

In this section, the influence of two important parameters, i.e., K (the band number of the dimension-reduced image) and N (the number of principal components used for classification) to the performance of the proposed MSTV method is discussed. The experiment is performed on the Indian Pines image, and the number of training samples accounts for 1% of ground truth. As shown in Fig. 7, when the values of K and N are relatively small, the classification accuracies of the proposed method decrease dramatically [see Fig. 7(a)–(c)]. The reason is that the useful discriminative information fails to be preserved well in the dimension reduction process. When the values of K and N increase gradually, the classification accuracies become more stable. Furthermore, it can be found that the computing time of the proposed method is increasing when the values of the two parameters increase [see Fig. 7(d)]. In order to reach a balance between the computing efficiency and the classification accuracies, in this paper, the default parameters of the proposed method are set as $K = 20$ and $N = 30$.

Furthermore, the influence of the multi-scale parameters, i.e., λ_l and σ_l , $l \in \{1, 2, 3\}$, to the classification performance of the proposed method is also analyzed. The experiment is conducted

on the Indian Pines dataset with 1% training samples. When λ_1 is discussed, other parameters are fixed, i.e., $\lambda_2 = 0.02$, $\lambda_3 = 0.01$, $\sigma_1 = 2$, $\sigma_2 = 1$, and $\sigma_3 = 3$. Similarly, when the influence of σ_1 is evaluated, other parameters are fixed, i.e., $\lambda_1 = 0.003$, $\lambda_2 = 0.02$, $\lambda_3 = 0.01$, $\sigma_2 = 1$, and $\sigma_3 = 3$. Fig. 8 shows the influence of these parameters on the classification performance. It can be observed that when λ_1 and σ_1 are set to be 0.003 and 2, respectively, the proposed method will obtain the highest classification accuracies. In the same way, the optimal values of other parameters are also estimated according to the highest accuracies. On the basis of the aforementioned analysis, $(\lambda_1 = 0.003, \sigma_1 = 2)$, $(\lambda_2 = 0.02, \sigma_2 = 1)$, and $(\lambda_3 = 0.01, \sigma_3 = 3)$ are set to be the default parameters in our experiments. In the following experiments, it is demonstrated that the proposed method is able to yield satisfactory classification accuracies for different datasets with this default parameter settings. Designing an automatic method to select the optimal parameters, i.e., λ_l and σ_l , will be an interesting research topic.

C. Analysis of the Influence of Different Components

First, the influence of two major steps, i.e., the multi-scale structure extraction and the KPCA-based fusion of stacked MSFs, to the classification performance of the proposed method is analyzed. An experiment is conducted on the Indian Pines dataset, and the training samples which account for 1% of

TABLE I
CLASSIFICATION PERFORMANCE OF THE PROPOSED METHOD WITH OR WITHOUT SEVERAL KEY PROCESSING COMPONENTS

Methods	Raw Data	MSF			Stacked MSFs	MSTV
		$\lambda_1 = 0.003, \sigma_1 = 2$	$\lambda_2 = 0.02, \sigma_2 = 1$	$\lambda_3 = 0.01, \sigma_3 = 3$		
OA	52.96(3.60)	73.21(3.78)	81.01(4.08)	78.41(3.14)	82.21(2.52)	89.09(1.76)
AA	51.80(3.19)	72.96(2.93)	70.74(3.95)	69.65(2.39)	79.13(3.16)	90.15(1.99)
Kappa	47.19(3.86)	69.89(4.71)	78.56(4.50)	75.63(3.56)	79.83(2.83)	87.59(1.99)

Raw data, MSF, stacked MSFs, and MSTV refer to the SVM on the raw data, the MSF obtained with a fixed parameter setting, the stacked MSFs without the KPCA-based dimension reduction step, and the proposed MSTV, respectively.

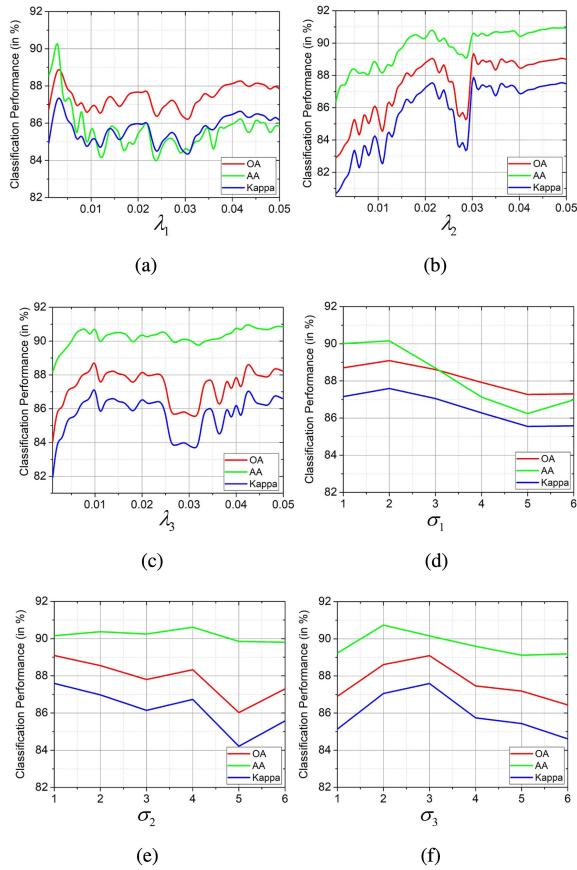


Fig. 8. Influence of the multi-scale parameters, i.e., λ_l and σ_l , $l \in \{1, 2, 3\}$, to the classification performance of the proposed method on the Indian Pines dataset.

the ground truth, are selected randomly. As shown in Table I, the K-dimensional MSFs obtained with different parameter settings, i.e., $(\lambda_1 = 0.003, \sigma_1 = 2)$, $(\lambda_2 = 0.02, \sigma_2 = 1)$, and $(\lambda_3 = 0.01, \sigma_3 = 3)$, can significantly improve the classification accuracies of the proposed method over the raw data in different levels. However, without the KPCA-based feature fusion step, the classification accuracies tend to decrease. The reason is that the stacked features contain a large amount of redundant information. By combining the KPCA and stacked MSFs, the proposed MSTV method yields the best classification performance in terms of OA, AA, and Kappa coefficient. This experiment demonstrates that the two major steps, i.e., the multi-scale structure extraction and the feature fusion steps, can both make important contributions to the proposed method.

TABLE II
CLASSIFICATION ACCURACIES (%) AND COMPUTING TIME (SECONDS) OF THE PROPOSED METHOD USING DIFFERENT METHODS FOR FEATURE FUSION, I.E., LDA [53], ICA [54], PPCA [55], PCA [56], AND KPCA [52]

Methods	LDA	ICA	PPCA	PCA	KPCA
OA	86.14	78.15	86.23	87.02	89.09
AA	82.43	81.75	87.51	88.58	90.15
Kappa	84.25	75.28	84.3	85.27	87.59
Time	1.91	1.65	22.01	2.37	4.21

Furthermore, the influence of different feature fusion methods is also analyzed. An experiment is performed on the Indian Pines dataset with the same experimental settings. Table II shows the classification performance of the proposed method with several widely used feature fusion methods, i.e., LDA [53], ICA [54], probabilistic PCA (PPCA) [55], and PCA [56]. From Table II, it can be observed that the adopted KPCA method obtains the highest classification accuracies. This experiment demonstrates that the KPCA method is an optimal solution for fusion of stacked MSFs in the proposed framework. The reason is that the KPCA considered the high-order information of the original input with a kernel mapping.

D. Effect of Structure Extraction and KPCA

In order to intuitively illustrate the advantage of the proposed method, the effect on the separability of pixels belonging to different classes is shown in Fig. 9. For an intuitive and distinct illustration, only the first eight classes are presented. Fig. 9(a), (c), and (e) show the first three bands of HSI, structural feature, and the first three principal components, respectively. Fig. 9(b), (d), and (f) present the scatter diagrams about the same three bands of 20 pixels per class.

Based on the graph examples, the following observations can be found: First, the structure extraction indeed decreases the inner-class differences while increasing the inter-class discrepancy. Taking Fig. 9(b) and (d), for example, it can be seen that the pixels belonging to different classes are intersected each other in the original HSI [see Fig. 9(b)]. However, when the structure extraction technique is performed on the HSI to obtain the structural feature, the pixel separability belonging to different classes can be increased effectively and the spectral variety of pixels belonging to the same object tends to reduce [see Fig. 9(d)]. This phenomenon verifies that the structure extraction indeed can reduce the pixel intercross among different classes.

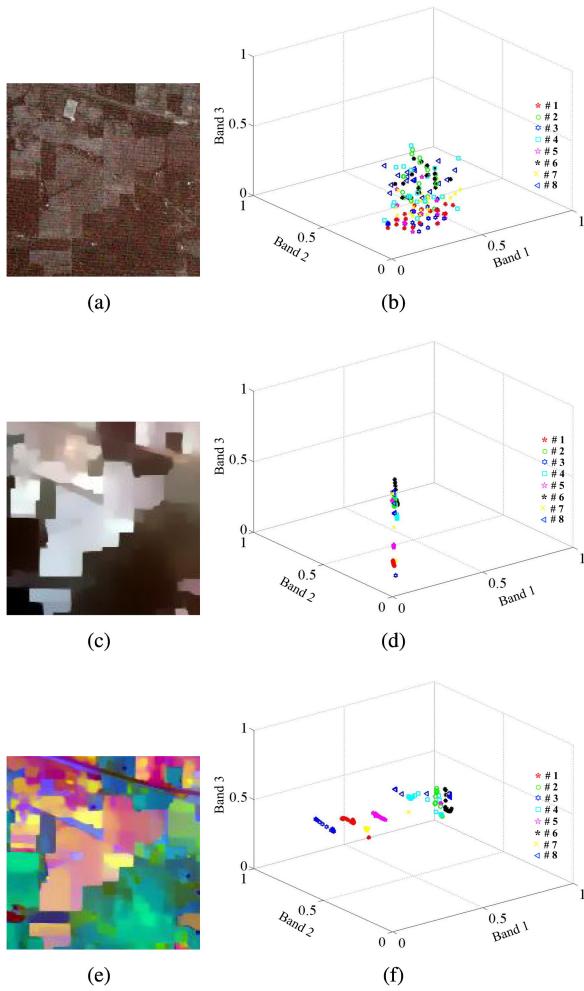


Fig. 9. Effect of structure extraction and KPCA to the pixel separability on the Indian Pines dataset (for a distinct and intuitive illustration, only the first eight classes are presented). (a) First bands of HSI. (b) Pixel separability of the first three bands (different symbols denotes pixels belonging to different classes). (c) Structural feature of the first three bands. (d) Pixel separability of the structural feature. (e) First three principal components obtained by KPCA on the structural feature. (f) Pixel separability of the first three principal components.

Furthermore, it is found that when the KPCA is performed on the structural feature, the separability of pixels belonging to different classes for the fused feature map is able to be further magnified [see Fig. 9(f)]. Moreover, by comparing Fig. 9(e) and (f), it can be observed that spectral differences among different classes indeed become more obvious. This situation demonstrates that the KPCA is helpful for highlighting the separability of pixels.

E. Comparison of Different Methods

In this section, the proposed MSTV method is compared with several state-of-the-art classification methods including the SVM [11], the extended morphological attribute profiles (EMAP) method [34], the multiple feature learning (MFL) method [57], the OTVCA method [42], the multi-hypothesis prediction (MHP) [58], the weighted Markov random fields

(WMRF) method [43], and the PCA-based edge-preserving features (PCA-EPFs) method [13], which are described briefly as follows.

- 1) SVM [11]: The SVM method is implemented using the LIBSVM library with the radial basis function (RBF) kernel, and the optimal parameters of the kernel are determined by using fivefold cross validation, in which the penalty factor c is from 10^{-2} to 10^4 and the RBF kernel width γ is from 2^{-5} to 2^5 .
- 2) EMAP [34]: The EMAP method is constructed using threshold values in the range from 2.5% to 10% with respect to the mean of the individual feature and with a step of 2.5% for the standard deviation attribute. The threshold of 100, 200, 500, and 1000 were chosen for the area attributes.
- 3) MFL [57]: The MFL method aims at integrating different types of features extracted from both linear and non-linear transformations without any regularization parameters to control the weight of each feature.
- 4) OTVCA [42]: The OTVCA method solves a minimization problem, which finds the optimal representation of the original HSI in a low-dimensional space while controlling the piecewise smoothness of the extracted features by using a total variation regularization.
- 5) MHP [58]: The MHP method uses a distance-weighted Tikhonov regularization to find the optimal linear combination of hypotheses, which aims to make original image have less intra-class variability and greater spatial smoothness by using MHP.
- 6) WMRF [43]: The WMRF method models the spatial prior on the hidden marginal probability, in which the labels are totally determined using an MRF prior with an adaptive TV-induced potential function.
- 7) PCA-EPFs [13]: The SVM method is performed on the PCA-EPFs.

For all comparison methods, the default parameters settings given in the corresponding literatures are adopted for a fair comparison. In order to make the proposed method reproducible, the code will be available online.¹

The experiment is first conducted on the Indian Pines dataset. The training samples, which account for 1% of the ground truth, are selected randomly. The classification results, i.e., CA, OA, AA, and Kappa, obtained by different classification methods are shown in Table III. It can be observed from Table III that the proposed method obtains the highest classification accuracies in terms of OA, AA, and Kappa coefficient. Moreover, the proposed method obtains the highest CA in nine classes. Fig. 10 shows the classification maps obtained by different methods associated with the corresponding OAs. From this figure, it can be seen that the proposed method yields satisfactory performance in the edges and boundaries of the resulting map. Besides, compared with other methods, the proposed method provides the highest OA reported in Fig. 10. By contrast, the SVM method appears obvious misclassification on the homogenous

¹<https://github.com/PuhongDuan>

TABLE III
CLASSIFICATION PERFORMANCE OF THE SVM [11], EMAP [34], MFL [57], OTVCA [42], MHP [58], WMRF [43], PCA-EPFs [13],
AND MSTV METHODS FOR THE INDIAN PINES DATASET WITH 1% TRAINING SAMPLES

Class	train	test	Classification accuracies obtained by different methods (in %)							
			SVM	EMAP	MFL	OTVCA	MHP	WMRF	PCA-EPFs	MSTV
Alfalfa	6	40	32.57(12.7)	94.26(8.42)	95.50(1.05)	14.70(5.52)	48.16(16.5)	98.00(1.97)	99.76(0.77)	83.73(23.0)
Corn_N	7	1421	46.83(6.18)	60.15(8.29)	65.96(9.99)	47.24(7.65)	56.43(11.0)	60.42(8.82)	74.20(8.46)	87.98(5.27)
Corn_M	6	824	34.13(14.6)	57.06(13.8)	65.67(9.50)	36.35(8.78)	51.49(12.9)	62.73(19.0)	69.14(13.3)	73.81(3.98)
Corn	6	231	26.79(6.53)	37.65(8.03)	77.36(9.36)	14.51(2.60)	31.42(9.46)	91.86(17.1)	68.04(14.6)	71.53(11.6)
Grass_M	6	477	61.19(11.1)	76.88(11.1)	78.43(7.13)	48.78(7.03)	63.53(18.4)	79.18(10.3)	95.53(4.20)	98.07(2.76)
Grass_T	6	724	79.44(8.38)	93.49(3.69)	92.50(6.91)	83.41(6.15)	76.12(4.71)	97.80(2.35)	92.79(5.83)	99.55(0.73)
Grass_P	6	22	32.15(14.8)	59.86(23.8)	97.27(2.35)	39.56(12.5)	35.48(18.2)	99.55(1.44)	66.32(27.9)	94.96(3.70)
Hay_W	7	471	95.62(2.13)	100.0(0.00)	99.49(0.35)	95.31(1.69)	98.14(2.01)	99.15(2.61)	100.0(0.00)	100.0(0.00)
Oats	6	14	16.45(7.79)	41.57(10.8)	99.29(2.26)	8.910(3.32)	11.26(6.41)	96.43(6.07)	70.41(27.5)	93.5(20.55)
Soybean_N	7	965	41.77(5.27)	52.49(11.7)	74.17(7.46)	29.50(2.50)	48.47(10.6)	70.59(12.8)	78.46(12.3)	81.42(8.67)
Soybean_M	8	2447	60.24(8.49)	75.88(11.6)	77.03(8.48)	51.39(4.02)	68.74(9.44)	55.19(20.4)	88.68(5.53)	93.06(4.01)
Soybean_C	6	587	28.89(5.65)	45.10(5.73)	69.88(9.48)	31.85(4.54)	30.45(9.73)	68.14(19.2)	78.27(11.7)	78.85(13.0)
Wheat	6	199	79.58(3.04)	88.30(10.5)	99.35(0.24)	64.99(16.9)	82.51(6.24)	100.0(0.00)	99.27(1.21)	100.0(0.00)
Woods	6	1259	88.56(5.33)	91.07(6.16)	91.93(4.56)	84.13(2.62)	87.85(5.95)	93.96(8.85)	99.43(0.59)	99.01(1.59)
Buildings	6	380	33.07(7.34)	68.14(13.2)	76.42(5.21)	37.51(12.1)	35.22(6.83)	66.53(17.6)	95.32(6.75)	91.82(7.63)
Stone	7	86	71.57(24.3)	92.02(8.57)	91.28(4.69)	89.40(9.79)	96.88(3.85)	98.14(3.25)	97.31(3.12)	95.06(2.49)
OA			52.96(3.60)	68.71(4.73)	78.62(3.12)	46.76(3.22)	58.24(4.93)	72.59(5.91)	84.17(4.10)	89.09(1.76)
AA			51.80(3.19)	70.87(3.90)	84.47(1.82)	48.60(1.45)	57.63(4.07)	83.60(2.30)	85.81(3.64)	90.15(1.99)
Kappa			47.19(3.86)	64.73(5.17)	75.77(3.48)	40.64(2.95)	53.24(5.36)	69.31(6.29)	82.03(4.60)	87.59(1.99)

Value in the each parentheses indicates the standard deviation of the accuracies obtained in repeated experiments. The highest accuracy in each row is highlighted in bold.

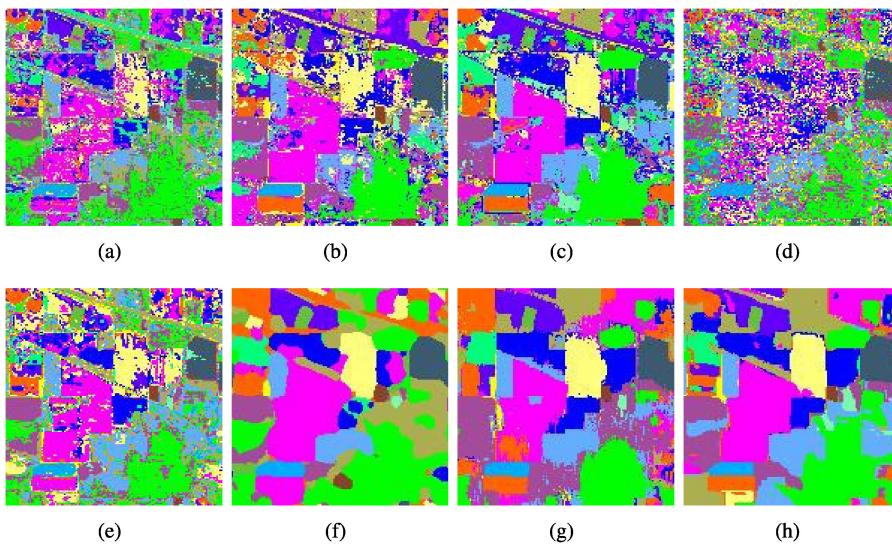


Fig. 10. Classification maps obtained by different methods on the Indian Pines dataset when the number of training samples is 1% of the ground truth. (a) SVM, OA = 52.96%. (b) EMAP, OA = 68.71%. (c) MFL, OA = 78.62%. (d) OTVCA, OA = 46.76%. (e) MHP, OA = 58.24%. (f) WMRF, OA = 72.59%. (g) PCA-EPFs, OA = 84.17%. (h) MSTV, OA = 89.13%.

regions. The reason is that this classifier takes only the spectral information into account. By combining the spatial information into feature extraction, the EMAP method improves the classification performance over the SVM method, but some noise-like mislabeled pixels can still be observed in the classification map. By considering the complementary information of different features, the MFL method produces a smoother classification map than the SVM and EMAP methods. However, the resulting map still has some noise, such as misclassifications. Owing to the limited training samples, the classification map obtained by the OTVCA method has serious mislabeled pixels. The MHP method also tends to yield the “noise”-like mislabels in the

classification map. For the WMRF method, a spatially adaptive total variation regularization is used to enforce a more spatially smooth classification, which can improve the classification accuracy. However, the classification map appears over-smoothed when the number of training samples is limited. The PCA-EPFs method improves the classification performance with respect to the other compared methods, since the multi-scale information is used to increase the separability of pixels. However, since some important spatial structures are not well preserved in the features, the boundaries of the objects in the classification result are not aligned well with those in the original HSI. Compared with the methods presented above, the proposed method shows

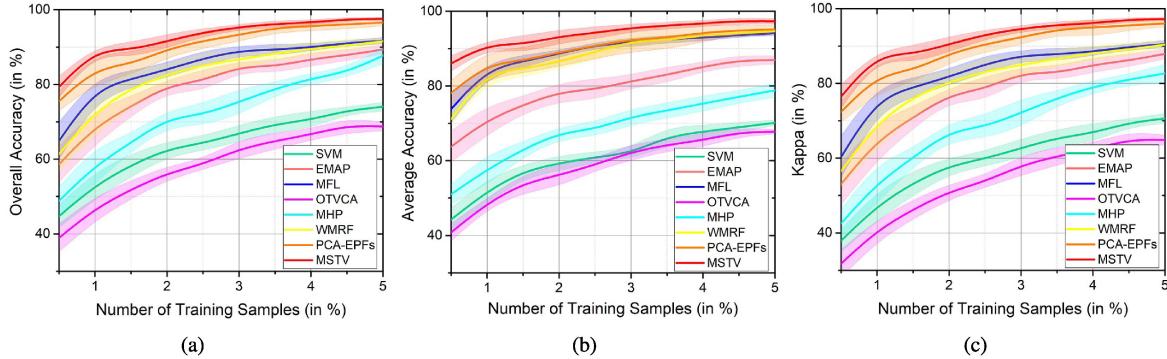


Fig. 11. Classification accuracies of different methods on the Indian Pines dataset with different numbers of training samples (varying from 1% to 5%). (a) OA. (b) AA. (c) Kappa Coefficient. The widths of the line areas refer to the standard deviation of accuracies obtained in ten repeated experiments.

TABLE IV
CLASSIFICATION PERFORMANCE OF THE SVM [11], EMAP [34], MFL [57], OTVCA [42], MHP [58], WMRF[43], PCA-EPFs [13],
AND MSTV METHODS FOR THE UNIVERSITY OF PAVIA DATASET WITH TEN LABELED SAMPLES PER CLASS

Class	train	test	Classification accuracies obtained by different methods (in %)							
			SVM	EMAP	MFL	OTVCA	MHP	WMRF	PCA-EPFs	MSTV
Asphalt	10	6621	93.18(3.70)	93.25(3.42)	80.84(7.92)	93.09(3.01)	93.22(3.90)	66.48(18.2)	86.71(6.05)	95.60(4.39)
Meadows	10	18639	84.90(3.46)	93.91(1.85)	80.34(10.4)	97.18(0.76)	89.79(4.21)	81.70(11.5)	98.07(1.30)	98.66(0.80)
Gravel	10	2089	38.88(6.59)	46.10(7.28)	67.45(12.5)	61.19(9.25)	52.53(17.6)	79.57(17.7)	81.63(5.61)	90.78(12.4)
Tress	10	3054	56.09(8.72)	74.26(13.5)	94.77(1.83)	87.51(9.00)	64.50(11.5)	87.84(5.48)	80.49(7.73)	90.63(8.90)
Sheets	10	1335	86.83(15.4)	98.93(0.94)	99.39(0.29)	100.0(0.00)	94.69(3.31)	99.97(0.07)	90.97(8.60)	91.74(11.8)
Soil	10	5019	42.21(8.33)	55.95(8.30)	73.15(6.31)	86.97(7.17)	49.36(10.2)	70.27(12.1)	92.49(3.93)	96.95(2.42)
Bitumen	10	1320	44.76(6.68)	58.8(18.46)	94.86(2.91)	87.93(8.73)	51.93(17.5)	97.07(4.07)	79.29(7.85)	99.82(0.30)
Bricks	10	3672	67.04(13.1)	80.22(7.31)	75.56(16.5)	72.64(7.77)	74.72(8.98)	76.12(20.5)	64.99(10.5)	82.60(5.80)
Shadows	10	937	99.93(0.17)	99.77(0.14)	96.14(5.66)	96.43(4.36)	99.85(0.07)	48.03(10.4)	78.05(8.88)	99.42(0.98)
OA			67.63(4.16)	78.30(3.96)	80.96(4.63)	89.20(2.39)	73.59(6.64)	78.16(4.22)	87.79(1.56)	95.00(1.79)
AA			68.20(3.71)	77.91(2.15)	84.72(1.95)	86.99(2.89)	74.51(4.60)	78.56(1.50)	84.17(1.93)	94.02(2.81)
Kappa			59.33(4.32)	72.43(4.49)	75.64(5.39)	85.91(3.06)	66.79(7.82)	71.83(0.05)	83.63(2.21)	93.41(2.33)

Value in the each parentheses indicates the standard deviation of the accuracies obtained in repeated experiments. The highest accuracy in each row is highlighted in bold.

the best performance in terms of classification accuracy and visual appearance.

Furthermore, the influence of different numbers of training samples to the classification performance is also investigated. Fig. 11 presents the OA, AA, and Kappa coefficient of different classification methods as functions of the number of training samples (varying from 0.5% to 5% with a step size of 0.5%). As shown in Fig. 11, with the increase of the number of training samples, the classification performance of all methods gradually tends to increase. More importantly, the proposed MSTV method always yields the highest accuracies, which also demonstrates the effectiveness of the proposed method.

The second experiment is performed on the University of Pavia dataset. The number of training samples is ten for each class. The classification results, i.e., CA, OA, AA, and Kappa, obtained by different classification methods are shown in Table IV. From this table, it can be seen that the proposed method shows the best performance in terms of OA, AA, and Kappa coefficient. Moreover, the influences of different numbers of training samples to the performance of all classification methods are shown in Fig. 12. As can be seen in this figure, the classification accuracies obtained by the proposed MSTV method are still the highest among all compared methods with the increase of the

number of training samples per class (ranging from 5 to 80). Moreover, when the number of training samples is very small, the proposed method shows very obvious advantages with respect to other compared methods.

The third experiment is performed on the Houston dataset. The number of training samples is 20 for each class. Table V shows the classification accuracies, i.e., OA, AA, and Kappa coefficient, obtained by different methods. As before, it can be observed that the proposed method is still superior than other compared methods, and obtains the highest accuracies for most classes. Furthermore, the influences of various numbers of training samples (varying from 10 to 50 with a step size of 5 per class) to the performance of all compared methods are shown in Fig. 13. Similarly, the classification accuracies tend to increase as the number of training samples increases, and the proposed method still performs better than other compared methods. This is in accordance with the previous experiments on the Indian Pines and University of Pavia datasets. It should be mentioned that the PCA-EPFs method shows obvious decrease in the classification accuracy since the captured scene contains many imperfect situations, such as shadow and noise. Compared with the PCA-EPFs, the proposed method still works well in such situations.

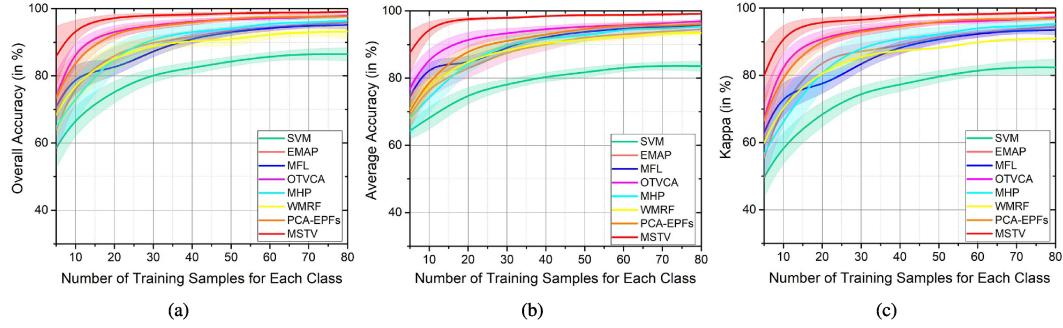


Fig. 12. Classification accuracies of different methods on the University of Pavia dataset with different numbers of training samples per class (varying from 5 to 80). (a) OA. (b) AA. (c) Kappa Coefficient. The widths of the line areas refer to the standard deviation of accuracies obtained in ten repeated experiments.

TABLE V
CLASSIFICATION PERFORMANCE OF THE SVM [11], EMAP [34], MFL [57], OTVCA [42], MHP [58], WMRF [43], PCA-EPFs [13], AND MSTV METHODS FOR THE HOUSTON DATASET WITH 20 LABELED SAMPLES PER CLASS

Class	train	test	Classification accuracies obtained by different methods (in %)							
			SVM	EMAP	MFL	OTVCA	MHP	WMRF	PCA-EPFs	MSTV
Grass_healthy	20	1053	92.87(4.24)	94.28(4.41)	81.36(6.87)	92.48(4.69)	79.67(31.9)	93.2(3.85)	80.67(7.79)	90.11(4.84)
Grass_stressed	20	790	94.19(4.27)	90.66(3.85)	90.62(5.65)	96.46(4.43)	91.23(6.52)	95.37(4.00)	72.00(9.33)	91.06(6.69)
Grass_synthetic	20	6767	97.06(2.74)	99.83(0.55)	100.0(0.00)	99.77(0.46)	92.35(12.4)	99.87(0.15)	99.58(0.62)	100.0(0.00)
Tree	20	1033	93.95(3.98)	92.41(4.18)	90.37(4.64)	94.24(2.64)	92.71(4.38)	96.59(2.44)	64.76(7.72)	86.41(5.43)
Soil	20	1222	93.04(2.90)	93.40(1.90)	97.36(2.90)	96.58(1.49)	95.25(2.32)	98.52(0.66)	95.94(3.36)	97.17(2.33)
Water	20	305	90.81(9.24)	99.31(1.14)	90.16(6.17)	99.86(0.43)	82.49(25.0)	92.07(6.74)	77.04(10.6)	95.67(5.71)
Residential	20	958	78.09(6.75)	94.67(3.11)	87.50(4.15)	88.14(5.39)	84.54(7.53)	80.72(5.90)	71.44(5.55)	85.99(6.75)
Commercial	20	604	76.28(7.13)	79.71(6.71)	51.14(5.92)	80.95(7.06)	67.00(22.6)	63.55(8.09)	81.32(8.67)	87.41(6.52)
Road	20	1011	70.95(4.76)	74.98(3.69)	57.39(6.53)	81.84(4.52)	80.06(8.48)	69.19(6.47)	64.36(7.75)	89.22(4.05)
Highway	20	362	76.43(6.55)	81.13(5.99)	68.80(8.96)	85.21(6.72)	80.20(6.77)	77.31(6.21)	86.65(5.40)	87.70(4.34)
Railway	20	94	75.76(6.92)	82.20(5.08)	74.67(5.34)	88.20(4.50)	57.12(30.2)	70.12(7.32)	91.96(5.40)	96.28(5.05)
Parking_lot1	20	1213	67.82(3.75)	78.34(7.05)	55.58(8.60)	85.36(4.80)	77.54(7.70)	60.99(7.31)	90.55(6.12)	95.18(3.17)
Parking_lot2	20	429	42.65(10.9)	84.71(8.04)	73.56(5.32)	68.30(14.1)	46.87(15.9)	62.52(4.01)	70.43(9.57)	82.09(7.33)
Tennis_court	20	408	87.24(7.10)	81.14(7.42)	97.99(1.40)	99.14(1.32)	85.78(11.0)	98.95(1.04)	99.35(0.87)	100.0(0.00)
Running_track	20	640	98.62(1.08)	86.59(1.80)	99.45(0.48)	98.98(1.98)	84.14(0.51)	98.72(0.57)	98.89(1.60)	97.90(1.68)
OA			82.31(2.14)	86.54(1.13)	78.56(1.98)	89.38(1.61)	77.53(6.23)	82.46(1.19)	81.23(2.14)	91.09(1.55)
AA			82.39(1.98)	87.56(0.89)	81.06(1.92)	90.37(1.34)	79.80(4.87)	83.84(1.05)	83.00(1.98)	92.15(1.10)
Kappa			80.87(2.32)	85.45(1.23)	76.82(2.13)	88.582(1.74)	75.72(6.73)	81.05(1.29)	79.72(2.31)	90.37(1.68)

Value in the each parentheses indicates the standard deviation of the accuracies obtained in repeated experiments. The highest accuracy in each row is highlighted in bold.

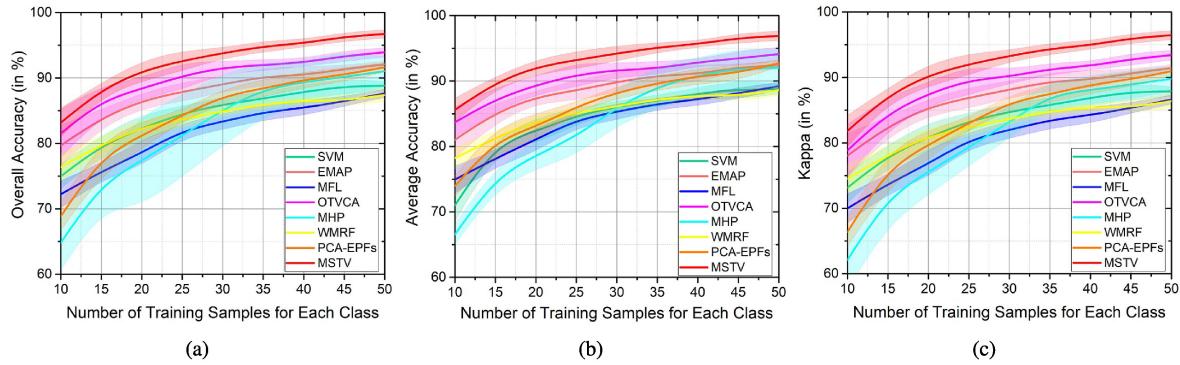


Fig. 13. Classification accuracies of different methods on the Houston dataset with different numbers of training samples per class (varying from 10 to 50). (a) OA. (b) AA. (c) Kappa Coefficient. The widths of the line areas refer to the standard deviation of accuracies obtained in ten repeated experiments.

Finally, in order to evaluate the efficiency of the eight techniques on the three real HSIs, the computing time (in seconds) of all methods is shown in Table VI. The training samples are set to be the same as those presented in Tables III–V. All experiments are performed on a Laptop with 2.6 GHz and 8 GB

RAM using MATLAB 2014. From Table VI, it can be observed that the computational cost of the proposed MSTV method performed on the Indian Pines image is quite competitive among all compared methods (takes about 4.21 s for classifying the Indian Pines image of size 145 × 145 pixels). However, when the

TABLE VI

COMPUTING TIME OF THE SVM, EMAP, MFL, OTVCA, MHP, WMRF, PCA-EPFs, AND MSTV METHODS APPLIED TO THE INDIAN PINES, UNIVERSITY OF PAVIA, AND HOUSTON DATASETS WHEN THE NUMBER OF TRAINING SAMPLES IS SET TO THE SAME AS THOSE SHOWN IN TABLES III–V

Data sets	Computing time of the compared methods (in seconds)							
	SVM [11]	EMAP [34]	MFL [57]	OTVCA [42]	MHP [58]	WMRF [43]	PCA-EPFs [13]	MSTV
Indian Pines	4.83	6.28	4.28	27.09	178.72	25.23	2.78	4.21
University of Pavia	8.24	18.46	13.88	246.51	680.97	105.15	10.84	40.04
Houston	51.49	58.62	45.86	698.45	1221.03	212.82	41.65	156.06

TABLE VII

CLASSIFICATION ACCURACIES (IN %) OF DIFFERENT METHODS, I.E., SVM, EMAP, EPF, MFL, WMRF, IFRF, PCA-EPFs, AND MSTV, ON THE INDIAN PINES DATASET CORRUPTED BY ZERO-MEAN GAUSSIAN NOISE WITH VARIANCE $\sigma \in \{0.1, 0.3, 0.5\}$

Methods	$\sigma = 0.1$			$\sigma = 0.3$			$\sigma = 0.5$		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
SVM [11]	24.12(0.00)	1.510(0.00)	0.000(0.00)	24.12(0.00)	1.510(0.00)	0.000(0.00)	24.12(0.00)	1.510(0.00)	0.000(0.00)
EMAP [34]	24.12(0.00)	1.510(0.00)	0.000(0.00)	24.12(0.00)	1.510(0.00)	0.000(0.00)	24.12(0.00)	1.510(0.00)	0.000(0.00)
MFL [57]	18.18(1.53)	14.00(1.96)	17.75(2.79)	11.86(0.98)	8.484(1.02)	8.031(1.04)	10.17(0.91)	6.762(0.61)	5.421(1.83)
OTVCA [42]	28.33(2.53)	27.05(2.34)	20.48(2.16)	24.71(3.08)	22.09(1.53)	16.57(2.21)	22.56(1.87)	19.46(1.54)	14.67(1.64)
MHP [58]	44.70(2.72)	46.58(2.82)	37.99(2.70)	41.84(2.97)	41.23(1.72)	34.37(2.91)	40.33(2.25)	38.42(1.93)	33.37(2.18)
WMRF [43]	0.881(0.02)	7.370(0.75)	0.041(0.02)	0.881(0.02)	7.301(0.72)	0.030(0.02)	0.870(0.01)	7.270(0.70)	0.030(0.02)
PCA-EPFs [13]	60.89(2.23)	63.08(2.26)	55.60(2.38)	53.23(2.16)	54.37(1.23)	47.51(2.46)	51.59(2.48)	50.19(1.37)	45.79(2.63)
MSTV	80.38(1.72)	80.90(3.66)	77.78(1.88)	77.42(2.08)	75.91(4.41)	74.46(2.31)	76.27(3.17)	72.98(2.35)	73.13(3.51)

Value in the each parentheses indicates the standard deviation of the accuracies obtained in repeated experiments.

TABLE VIII

CLASSIFICATION ACCURACIES (IN %) OF DIFFERENT METHODS, I.E., KSVD [59], LRTA [60], ITSR [61], NLM3D [62], LRMF [63], PARAFAC [64], BM4D [65], AND MSTV, ON THE INDIAN PINES DATASET CORRUPTED BY ZERO-MEAN GAUSSIAN NOISE WITH VARIANCE $\sigma = 0.1$

Class	train	test	Classification accuracies obtained by different methods (in %)							
			KSVD	LRTA	ITSR	NLM3D	LRMF	PARAFAC	BM4D	MSTV
Alfalfa	6	40	46.63(15.7)	14.00(4.18)	42.99(11.9)	58.98(19.2)	8.370(1.40)	26.10(9.99)	91.58(12.0)	95.36(5.33)
Corn_N	7	1421	52.90(15.2)	39.09(9.72)	45.79(10.7)	51.60(11.0)	31.55(7.81)	48.58(14.3)	52.32(22.5)	78.74(5.40)
Corn_M	6	824	56.14(27.4)	21.58(6.24)	31.36(7.74)	41.81(10.8)	15.28(2.88)	31.33(5.60)	69.45(27.3)	64.68(6.86)
Corn	6	231	57.87(23.9)	19.02(7.81)	40.09(14.6)	50.19(17.4)	13.24(3.51)	26.98(8.02)	61.44(26.5)	32.53(7.40)
Grass_M	6	477	70.46(19.5)	31.45(8.58)	57.22(16.3)	62.23(12.8)	21.48(5.92)	46.10(8.94)	58.94(13.4)	96.91(5.02)
Grass_T	6	724	82.78(4.76)	80.71(6.30)	81.52(3.35)	80.74(5.20)	58.39(10.4)	74.54(7.10)	85.57(5.61)	91.90(5.90)
Grass_P	6	22	36.31(11.0)	14.89(3.37)	48.71(13.3)	55.47(16.6)	2.980(1.28)	44.11(15.1)	74.80(13.6)	84.14(20.7)
Hay_W	7	471	83.10(8.75)	91.22(5.83)	99.46(0.91)	94.15(3.35)	55.46(14.2)	80.46(8.46)	96.42(3.00)	100.00(0.00)
Oats	6	14	9.390(3.32)	6.740(2.05)	34.13(23.6)	18.74(7.90)	1.660(0.53)	40.97(17.8)	94.64(9.00)	65.74(29.6)
Soybean_N	7	965	37.61(14.1)	27.39(5.63)	58.93(5.92)	41.28(9.85)	20.15(4.32)	38.32(7.31)	57.36(9.68)	75.59(10.3)
Soybean_M	8	2447	33.96(4.36)	53.65(7.51)	65.88(7.77)	60.42(6.57)	45.77(5.14)	57.28(5.02)	57.80(10.0)	88.27(3.41)
Soybean_C	6	587	49.09(24.5)	19.44(3.06)	33.33(5.33)	37.15(6.15)	14.80(6.15)	34.59(5.60)	48.14(16.3)	50.99(9.16)
Wheat	6	199	59.12(17.1)	68.46(8.18)	88.98(7.53)	80.17(12.0)	39.75(9.61)	77.58(11.5)	96.38(3.35)	99.75(0.49)
Woods	6	1259	89.88(4.26)	76.73(3.95)	81.66(5.43)	90.07(6.28)	70.17(3.62)	82.21(5.40)	88.42(6.31)	98.58(1.42)
Buildings	6	380	44.51(14.1)	39.43(12.6)	36.24(6.14)	53.92(9.98)	17.10(2.48)	46.14(14.3)	57.77(8.98)	93.10(5.50)
Stone	7	86	99.39(0.65)	78.82(25.8)	25.68(4.66)	97.94(0.88)	48.21(23.2)	57.58(20.9)	99.41(0.62)	78.14(25.4)
OA			44.12(5.84)	41.35(3.34)	55.11(3.60)	58.31(4.31)	31.73(5.54)	50.88(3.27)	61.23(4.69)	80.38(1.72)
AA			56.82(2.01)	42.66(2.46)	54.50(3.15)	60.93(4.04)	29.02(2.90)	50.80(3.21)	74.40(3.37)	80.90(3.66)
Kappa			31.57(8.06)	34.77(3.45)	49.54(3.76)	52.95(4.64)	23.63(4.84)	44.90(3.34)	55.37(5.50)	77.78(1.88)

Value in the each parentheses indicates the standard deviation of the accuracies obtained in repeated experiments.

number of pixels in the spatial domain is relatively large, it is not computationally efficient with respect to other methods such as SVM, EMAP, and MFL. The reason is that the RTV-based structure extraction-method is a pixel-wise processing algorithm. To solve this problem, designing a faster implementation of MSTV with GPU programing could be investigated.

F. Noise Robustness

In this section, the influence of image noise to the performance of different methods is discussed. An experiment is performed on the Indian Pines dataset with 1% samples for training, in

which each band has been corrupted by the zero-mean Gaussian noise of different variances $\sigma \in \{0.1, 0.3, 0.5\}$. The classification accuracies obtained by different methods are presented in Table VII. It can be observed that the performance of the compared methods, i.e., SVM, EMAP, EPF, and WMRF, are dramatically downgraded when the original HSI is corrupted by a serious noise. The classification accuracies obtained by the IFRF and PCA-EPFs methods also decrease dramatically with respect to the two methods performed on the original HSI without adding noise. By contrast, the proposed method is still able to obtain very high classification accuracies, i.e., OA = 80.38%, AA = 80.90%, Kappa = 77.78%, when the original HSI is

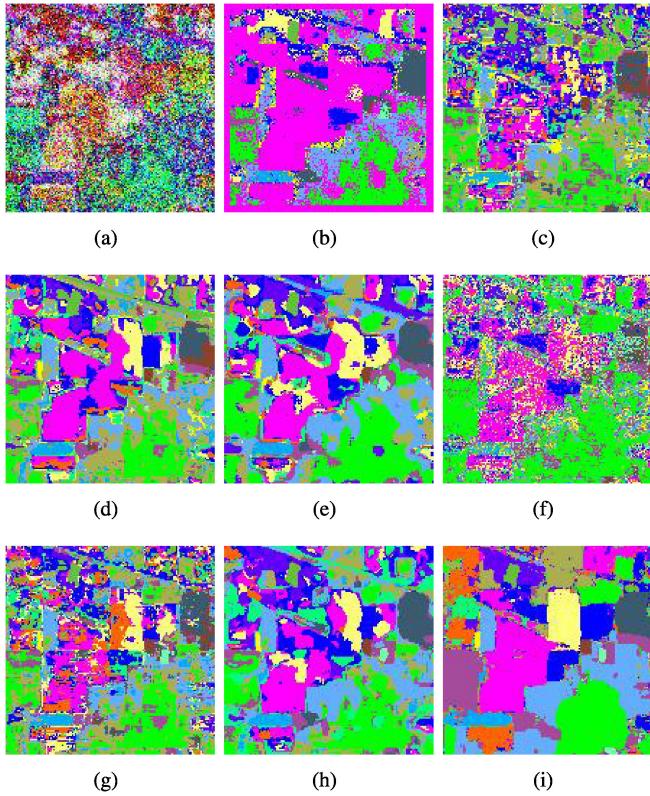


Fig. 14. Classification maps obtained by different methods on the Indian Pines dataset when the number of training samples is 1% of the ground truth. (a) Original image corrupted by the zero-mean Gaussian noise with variance $\sigma = 0.1$. (b) KSVD, OA = 44.12%. (c) LRTA, OA = 41.35%. (d) ITSR, OA = 55.11%. (e) NLM3D, OA = 58.31%. (f) LRMF, OA = 31.73%. (g) PARAFAC, OA = 50.88%. (h) BM4D, OA = 61.23%. (i) MSTV, OA = 80.38%.

corrupted by the zero-mean Gaussian noise with variance $\sigma = 0.1$. This experiment demonstrates that the major advantage of the proposed method is its robustness against serious image noise, which may be caused by low-lighting or other imperfect factors.

Furthermore, in order to further demonstrate the advantage of the proposed method, several representative denoising methods, i.e., K-singular value decomposition (KSVD) [59], lower rank tensor approximation (LRTA) [60], intrinsic tensor sparsity regularization (ITSR) [61], three-dimensional (3-D) non-local means filtering (NLM3D) [62], low-rank matrix factorization (LRMF) [63], parallel factor analysis (PARAFAC) [64], and block matching four-dimensional (4-D) filtering (BM4D) [65], are adopted to remove the zero-mean Gaussian noise containing the original HSI, and then, the SVM classifier is performed on the denoised image for classification. The classification maps are shown in Fig. 14. It can be seen that these methods, i.e., KSVD, LRTA, LRMF, and PARAFAC, produce serious misclassified labels in the resulting maps. The classification maps obtained by the NLM3D and BM4D methods are slightly improved. However, some areas near structures are still misclassified. By contrast, the proposed method can yield a satisfactory result compared to other studied methods. Moreover, Table VIII presents the classification accuracies of different methods. It

can be observed that the proposed method still obtains the best performance in terms of OA, AA, and Kappa coefficient.

V. CONCLUSION

In this paper, a novel HSI classification method based on multi-scale structure extraction is proposed. This method mainly consists of two components, i.e., the RTV-based multi-scale structure extraction and the KPCA-based feature fusion. The advantage of structure extraction is that it can remove the useless spatial information such as noise and texture. The advantage of KPCA is that it cannot only increase the class separability in MSFs, but also effectively fuse the complementary information in the MSFs obtained with different parameter settings. Experimental results on three real hyperspectral datasets demonstrate the advantages of the proposed MSTV method. First, the proposed method outperforms other recently proposed HSI classification methods in terms of classification accuracy, especially when the number of training samples is limited. Second, the proposed method works well when the original image has been corrupted by serious noise. However, a disadvantage of the proposed method is that its parameters in the multi-scale structure extraction need to be set manually, which influences on the operational applicability in real applications. How to automatically select the optimal parameters for MSFs will be the focus of our future research.

ACKNOWLEDGMENT

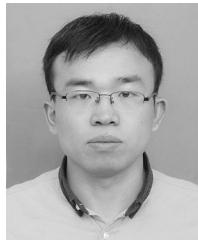
The authors would like to thank the editor and anonymous reviewers for their detailed review, valuable comments, and useful suggestions.

REFERENCES

- [1] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.
- [2] G. Gao and Y. Gu, "Multitemporal landsat missing data recovery based on tempo-spectral angle model," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3656–3668, Jul. 2017.
- [3] X. Kang, Y. Huang, S. Li, H. Lin, and J. A. Benediktsson, "Extended random walker for shadow detection in very high resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 867–876, Feb. 2018.
- [4] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [5] X. Kang, P. Duan, S. Li, and J. A. Benediktsson, "Decolorization-based hyperspectral image visualization," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4346–4360, Aug. 2018.
- [6] X. Lu, B. Wang, X. Zheng, and X. Li, "Exploring models and data for remote sensing image caption generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2183–2195, Apr. 2018.
- [7] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [8] G. Cheng, J. Han, L. Guo, Z. Liu, S. Bu, and J. Ren, "Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4238–4249, Aug. 2015.
- [9] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

- [10] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.
- [11] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [12] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
- [13] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017.
- [14] X. Kang, B. Zhuo, and P. Duan, "Dual-path network-based hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 447–451, Mar. 2019.
- [15] S. Li, Q. Hao, G. Gao, and X. Kang, "The effect of ground truth on performance evaluation of hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7195–7206, Dec. 2018.
- [16] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.
- [17] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.
- [18] B. Mojarradi, H. Abrishami-Moghaddam, M. J. V. Zanjani, and R. P. W. Duin, "Dimensionality reduction of hyperspectral data via spectral feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2091–2105, Jul. 2009.
- [19] N. He *et al.*, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 755–769, Feb. 2019.
- [20] J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction in remote sensing using hyperspectral imaging [applications corner]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 149–154, Jul. 2014.
- [21] G. Licciardi, P. R. Marpu, J. Chanussot, and J. A. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 447–451, May 2012.
- [22] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.
- [23] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.
- [24] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.
- [25] P. Ghamisi, M. D. Mura, and J. A. Benediktsson, "A survey on spectral-spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, May 2015.
- [26] X. Kang, P. Duan, X. Xiang, S. Li, and J. A. Benediktsson, "Detection and correction of mislabeled training samples for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5673–5686, Oct. 2018.
- [27] P. Ghamisi *et al.*, "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.
- [28] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.
- [29] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2019.2893180.
- [30] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [31] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
- [32] M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.
- [33] Y. Gu, T. Liu, X. Jia, J. A. Benediktsson, and J. Chanussot, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.
- [34] P. R. Marpu, M. Pedernana, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "Automatic generation of standard deviation attribute profiles for spectral-spatial classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 293–297, Mar. 2013.
- [35] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1641–1645, Nov. 2016.
- [36] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014.
- [37] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. A. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, Oct. 2016.
- [38] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.
- [39] X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image decomposition for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2241–2253, Apr. 2015.
- [40] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
- [41] J. M. Duarte-Carvajalino, P. E. Castillo, and M. Velez-Reyes, "Comparative study of semi-implicit schemes for nonlinear diffusion in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1303–1314, May 2007.
- [42] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976–6985, Dec. 2016.
- [43] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.
- [44] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 139:1–139:10, Nov. 2012.
- [45] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
- [46] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [47] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 67:1–67:10, Aug. 2008.
- [48] R. Arnheim, *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA, USA: Univ. California Press, 1956.
- [49] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition-modeling, algorithms, and parameter selection," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 111–136, Apr. 2006.
- [50] X. Wu, J. Zheng, C. Wu, and Y. Cai, "Variational structure-texture image decomposition on manifolds," *Signal Process.*, vol. 93, no. 7, pp. 1773–1784, Jul. 2013.
- [51] Y. Chen *et al.*, "Curve-like structure extraction using minimal path propagation with backtracking," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 988–1003, Feb. 2016.
- [52] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for feature reduction in hyperspectral images analysis," in *Proc. Int. Signal Process. Symp.*, Jun. 2006, pp. 238–241.
- [53] A. J. Izenman, "Linear discriminant analysis," in *Modern Multivariate Statistical Techniques*. New York, NY, USA: Springer, 2013, pp. 237–280.

- [54] S. Roweis, "Kurtosis maximization ICA." Accessed: Aug. 30, 2007. [Online]. Available: <http://www.cs.nyu.edu/roweis/kica.html>
- [55] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Statist. Soc. B (Statist. Methodol.)*, vol. 61, no. 3, pp. 611–622, Jan. 1999.
- [56] L. I. Smith, "A tutorial on principal components analysis," *Inf. Fusion*, vol. 51, Feb. 2002, Art. no. 52.
- [57] J. Li *et al.*, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.
- [58] C. Chen, W. Li, E. W. Tramel, M. Cui, S. Prasad, and J. E. Fowler, "Spectral-spatial preprocessing using multihypothesis prediction for noise-robust hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1047–1059, Apr. 2014.
- [59] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [60] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.
- [61] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, Jun. 2016, pp. 1692–1700.
- [62] Y. Qian, Y. Shen, M. Ye, and Q. Wang, "3-D nonlocal means filter with noise estimation for hyperspectral imagery denoising," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 1345–1348.
- [63] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-i.i.d. noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, Mar. 2018.
- [64] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.
- [65] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013.



Puhong Duan (S'17) received the B.Sc. degree from Suzhou University, Suzhou, China, in 2014, and the M.S. degree from the Hefei University of Technology, Hefei, China, in 2017. He is currently working toward the Ph.D. degree with the Laboratory of Vision and Image Processing, Hunan University, Changsha, China.

His research interests include image fusion and hyperspectral image visualization and classification.



Xudong Kang (S'13–M'15–SM'17) received the B.Sc. degree from Northeast University, Shenyang, China, in 2007, and the Ph.D. degree from Hunan University, Changsha, China, in 2015.

In 2015, he joined the College of Electrical Engineering, Hunan University, Changsha, China. His research interest includes hyperspectral feature extraction, image classification, image fusion, and anomaly detection.

Dr. Kang was awarded the Second Prize in the Student Paper Competition in IGARSS 2014. In IGARSS 2017, he was selected as the Best Reviewer for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2016. He serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.



Shutao Li (M'07–SM'15–F'19) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology. From November 2002 to November 2003, he was a Postdoctoral Fellow with the Royal Holloway College, University of London. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored more than 200 refereed papers. His current research interests include image processing, pattern recognition, and artificial intelligence.

Dr. Li gained two 2nd-Grade State Scientific and Technological Progress Awards of China in 2004 and 2006. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT. He is a Member of the Editorial Board of *Information Fusion* and *Sensing and Imaging*.



Pedram Ghamisi (S'12–M'15–SM'18) received the B.Sc. degree in civil (survey) engineering from the Tehran South Campus of Azad University, Tehran, Iran, in 2008, the M.Sc. degree (first class hons.) in remote sensing from the K. N. Toosi University of Technology, Tehran, Iran, in 2012, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015.

From 2013 to 2014, he was with the School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane, QLD, Australia. He was a Postdoctoral Research Fellow with the University of Iceland. He has been a Postdoctoral Research Fellow with the Technical University of Munich, Munich, Germany, and Heidelberg University, Heidelberg, Germany, since 2015. He was a Research Scientist with the German Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, Germany, from 2015 to 2018. He was the Head of the Machine Learning Group, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany. His research interests include interdisciplinary research on remote sensing and machine (deep) learning, image, and signal processing, and multisensory data fusion.

Dr. Ghamisi was a recipient of the prestigious Alexander von Humboldt Fellowship in 2015, the prestigious High Potential Program in 2018, the Best Researcher Award for M.Sc. students in K. N. Toosi University of Technology, from 2010 to 2011, the 2013 IEEE International Geoscience and Remote Sensing Symposium, Melbourne, Australia, in 2013, the IEEE Mikio Takagi Prize for winning the Student Paper Competition, competing with almost 70 submissions, and the Data Fusion Contest 2017 organized by the Image Analysis and Data Fusion Technical Committee of the Geoscience and Remote Sensing Society. He was also the winner of the 2017 Best Reviewer Prize of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. In 2016, he was selected as the Talented International Researcher by Iran's National Elites Foundation. He serves as an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and *Remote Sensing*.