

PCA-Based Edge-Preserving Features for Hyperspectral Image Classification

Xudong Kang, *Senior Member, IEEE*, Xuanlin Xiang, Shutao Li, *Senior Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

Abstract—Edge-preserving features (EPFs) obtained by the application of edge-preserving filters to hyperspectral images (HSIs) have been found very effective in characterizing significant spectral and spatial structures of objects in a scene. However, a direct use of the EPFs can be insufficient to provide a complete characterization of spatial information when objects of different scales are present in the considered images. Furthermore, the edge-preserving smoothing operation unavoidably decreases the spectral differences among objects of different classes, which may affect the following classification. To overcome these problems, in this paper, a novel principal component analysis (PCA)-based EPFs (PCA-EPFs) method for HSI classification is proposed, which consists of the following steps. First, the standard EPFs are constructed by applying edge-preserving filters with different parameter settings to the considered image, and the resulting EPFs are stacked together. Next, the spectral dimension of the stacked EPFs is reduced with the PCA, which not only can represent the EPFs in the mean square sense but also highlight the separability of pixels in the EPFs. Finally, the resulting PCA-EPFs are classified by a support vector machine (SVM) classifier. Experiments performed on several real hyperspectral data sets show the effectiveness of the proposed PCA-EPFs, which sharply improves the accuracy of the SVM classifier with respect to the standard edge-preserving filtering-based feature extraction method, and other widely used spectral-spatial classifiers.

Index Terms—Edge-preserving filtering, hyperspectral image (HSI), image classification, principal component analysis (PCA), support vector machine (SVM).

I. INTRODUCTION

HYPERSPECTRAL imaging is able to provide images with high spectral resolution, and thus offer a unique ability to distinguish the subtle spectral signatures of different land covers. Due to this advantage, hyperspectral remote sensing has been successfully used in many applications, such as target detection [1], [2], spectral unmixing [3], and scene

Manuscript received May 13, 2017; revised July 8, 2017; accepted August 16, 2017. Date of publication September 14, 2017; date of current version November 22, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61601179, in part by the National Natural Science Fund of China for Distinguished Young Scholars under Grant 61325007, in part by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001, and in part by the Science and Technology Plan Projects Fund of Hunan Province under Grant 2015WK3001. (*Corresponding author: Shutao Li*)

X. Kang, X. Xiang, and S. Li are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: xudong_kang@163.com; advancelin@hnu.edu.cn; shutao_li@hnu.edu.cn).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: benedikt@hi.is).

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/TGRS.2017.2743102

classification [4]. Among these applications, hyperspectral image (HSI) classification has drawn a lot of attention due to its importance in precise agriculture, urban investigation, and environment surveillance [5]–[7].

Different from general land-use classification methods [8]–[10], which aim at understanding the major contents of the scene, the goal of HSI classification is to assign a unique class label to each pixel in the HSI, which is a more challenging task. To achieve this objective, classical methods, such as the Bayesian estimation method [11], the support vector machines (SVMs) [12], and the sparse representation-based techniques [13], [14], have been successfully applied in HSI classification. However, when there only exist very few labeled samples, most of the classifiers mentioned earlier are not able to obtain a satisfactory classification performance due to the curse of dimensionality. Furthermore, adjacent noise-free hyperspectral bands are usually highly correlated with each other and a high spectral dimensionality also means that the computational burden of the classification process will be increased accordingly.

In order to solve these problems, feature extraction has been demonstrated to be an effective way to reduce the dimension of the data, while preserving or increasing the class separability of different objects [15]–[18]. Some classical signal analysis tools, such as the principal component analysis (PCA) [15], [19], [20], the independent component analysis (ICA) [16], singular spectrum analysis (SSA) [21], and the manifold learning [17] methods, have been successfully applied for feature extraction of HSIs. However, most of these methods exploit only the spectral information in the feature extraction process, and thus generally demonstrate a low performance.

In order to combine the spectral and spatial information in hyperpectral imagery, spectral-spatial feature extraction methods have been studied extensively in recent years [20], [22], [23]. For instance, Benediktsson *et al.* [24] introduce the extended morphological profiles (EMPs) to model the geometrical and texture properties of objects in HSIs, which has attracted great attention in the remote sensing field. The EMPs are built by sequentially applying opening and closing operations, which are found effective for representing the multiscale variability of the structures in the image. However, the EMPs rely heavily on the shapes of structure elements, and thus are not appropriate to model all geometrical features. This limitation is addressed by the use of the extended morphological attribute profiles (EMAPs) [25], [26]

or extinction profiles [27] that both could model the spatial characteristics of the structures in a scene according to different attributes. Gu *et al.* [28] combine multiple EMPs in a nonlinear multiple kernel learning, which can effectively improve the classification performance in terms of accuracies. Zabalza *et al.* [29] develop an interesting 2-D SSA (2-D-SSA)-based method to decompose the HSI into varying trends, oscillations, and noise. Using the trends and the selected oscillations as features, the 2-D-SSA method shows excellent classification performances. Kang *et al.* [30] model the HSI as the multiplication of two intrinsic components, i.e., the reflectance and shading components. By employing prior knowledge about the correlations of adjacent pixels, the resulting reflectance features demonstrate outstanding classification performances [30], [31].

In our previous work, the spectral-spatial features of HSIs are extracted with edge-preserving filtering, in which the edge-preserving filter is used to remove noise, weak edges, and small details in the HSI, while to preserve the overall structures, strong edges, and boundaries of the image [32]. The resulting edge-preserving features (EPFs) have been demonstrated to be effective in representing the major spectral-spatial structures of HSIs. However, one limitation of the standard edge-preserving smoothing operation is that it tends to decrease the spectral differences among the objects of different classes. Furthermore, the smoothing degree of the filter also has a big influence on the classification performance. In this situation, EPFs obtained by using a single parameter setting are not able to represent the complex spatial structures in the HSIs effectively.

To tackle the problems of the standard EPF-based method mentioned earlier, we present new PCA-based EPFs (PCA-EPFs) for HSI classification. First, by using different parameter settings for the edge-preserving filter, the standard EPFs can be constructed sequentially and stacked together. Then, the PCA is performed on the stacked EPFs in order to represent the EPFs in the mean square sense and highlight the spectral separability of pixels in the EPFs. Finally, the resulting PCA-EPFs are classified with an SVM classifier. Experimental results performed on several real hyperspectral data sets demonstrate that the proposed PCA-EPFs could sharply improve the accuracy of the SVM classifier with respect to our previous proposed EPFs-based feature extraction method, especially when there only exist very few labeled samples.

This paper is organized as follows. Section II recalls the principles of edge-preserving filtering and PCA. Section III is devoted to the introduction of the proposed PCA-EPFs method. Section IV presents the experimental results on three real hyperspectral data sets. Conclusions are reported in Section V.

II. BACKGROUNDS

A. Edge-Preserving Filtering

Edge-preserving filtering is an image processing technique, which can smooth away textures and noise while retaining significant parts of the image content, typically edges, lines, or other details that are important for the interpretation

of the image. Due to this advantage, edge-preserving filtering has received considerable interest from the image processing and computer vision research communities over the last two decades [33]–[36]. At the same time, some powerful edge-preserving filters have been successfully used in hyperspectral remote sensing applications [32], [37]–[39]. For example, in [37], edge-preserving filtering is proposed for the first time to make full use of the spectral and spatial information in postprocessing of pixelwise classification results. Instead of using edge-preserving filtering as a postprocessing tool, a spectral-spatial feature extraction method is proposed in [32] to remove the insignificant parts of an HSI, so as to further improve the interpretability of the HSIs. In [38], the HSI is first decomposed with the ICA, and then processed with edge-preserving filtering, so as to extract the features for image classification.

The bilateral filter [40] and anisotropic diffusion [41] are the most popular edge-preserving filters. However, the two filters are originally designed with high computing cost. While bilateral filtering uses a space-varying weighting function computed at a space of higher dimensionality than the signal being filtered, anisotropic diffusion requires an iterative solver. Although a number of techniques have been proposed to accelerate anisotropic diffusion or bilateral filtering, most of these methods can only be applied to grayscale images or rely on approximate operations [42]–[44]. Furthermore, many alternatives of edge-preserving filtering have been investigated, such as the guided filter [45], the domain transform filter [46], and the deep edge-aware filter [47]. In this paper, the domain transform recursive filter is adopted, because it is a real-time edge-preserving filter and has been demonstrated to be quite effective in improving the performance of HSI classification [32], [46].

Given a 1-D signal I , a domain transform recursive filter first applies an approximate distance-preserving transformation (a simple approximation is the sum of the spatial distance and the intensity difference between every two pixels) as follows:

$$U_i = I_0 + \sum_{j=1}^i \left(1 + \frac{\delta_s}{\delta_r} |I_j - I_{j-1}| \right) \quad (1)$$

where U is the domain transformed signal, δ_s and δ_r are two parameters to adjust the amount of smoothness of the filter. Then, the input signal is processed by recursive filtering as follows:

$$J_i = (1 - a^b)I_i + a^b J_{i-1} \quad (2)$$

where J_i is the filter output of the i th pixel, $a = \exp(-\sqrt{2}/\delta_s) \in [0, 1]$ refers to the feedback coefficient, and b measures the distance between two neighbor samples U_i and U_{i-1} in the transform domain. As b increases, a^b goes to zero, which stops the propagation chain, and thus, sharp edges in the signal will be preserved and vice versa.

A domain transform recursive filter processes images by performing the above-mentioned 1-D operations along each dimension of the image. Experimental results presented in [46] show that artifact-free filtered images can be obtained by

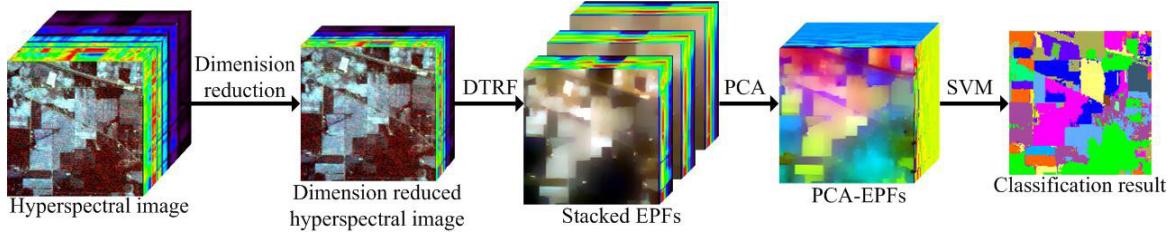


Fig. 1. Schematic of the proposed PCA-EPFs method. DTRF refers to the domain transform recursive filtering operation with different parameter settings. EPFs refer to the EPFs.

performing 1-D filtering with three iterations. In this paper, we refer the domain transform recursive filtered image with $\text{DTRF}(I, \delta_s, \delta_r)$, in which I represents the input band, and δ_s and δ_r are the parameters to adjust the amount of smoothness.

B. Principal Component Analysis

The PCA serves as a standard tool in various remote sensing image processing applications, because it is simple, nonparametric, and quite effective for exacting relevant information in the HSI [15]. In this paper, the PCA is adopted for dimension reduction and fusion of the EPFs because of the following reasons. First, the multiple EPFs obtained by different parameter settings contain not only complementary information but also a large amount of redundant information. In this situation, the PCA is a perfect solution to remove such redundant information. Second, although noise and small-scale details have been effectively removed from the HSI (the spectral differences among pixels belonging to the same class could be effectively reduced), the value differences among pixels belonging to different classes also have been decreased due to the smoothing operation (see the stacked EPFs in Fig. 1). To solve this problem, the PCA is a very good solution, since it could extract the most relevant information in the EPFs, and thus can effectively highlight the spectral differences among pixels belonging to different objects (see the PCA-EPFs in Fig. 1).

Given a data matrix \mathbf{S} , which consists of N observations and M variables (in image processing applications, N usually refers to the number of pixels in an image and M usually refers to the number of feature values for each pixel), the basic equation of the PCA-based dimension reduction is given by

$$\mathbf{P} = \mathbf{W}'\mathbf{S} \quad (3)$$

where the values of \mathbf{P} are the principal components constructed as weighted average of original sample vectors. The components in \mathbf{P} are ordered, so that the first few components retain most of the variation presenting in all of the original variables. \mathbf{W} is calculated from the covariance matrix \mathbf{C} as follows:

$$\mathbf{W} = \mathbf{E}\mathbf{A}^{-\frac{1}{2}} \quad (4)$$

where \mathbf{A} is a diagonal matrix of the eigenvalues of \mathbf{C} (arranged in decreasing order) and \mathbf{E} is the matrix of eigenvectors of \mathbf{C} . Let \mathbf{B} be the $M \times N$ matrix whose n th column is $\mathbf{S}_n - \mu$

$$\mathbf{B} = [\mathbf{S}_1 - \mu, \dots, \mathbf{S}_N - \mu] \quad (5)$$

where μ is the mean vector, which is calculated as $\mu = (1/N)(\mathbf{S}_1 + \dots + \mathbf{S}_N)$. The covariance matrix \mathbf{C} of size $M \times M$ can be calculated as

$$\mathbf{C} = \frac{1}{N-1}\mathbf{B}\mathbf{B}^T. \quad (6)$$

It is easy to observe from (4) that \mathbf{W} is simply the eigenvector matrix \mathbf{E} scaled, so that the variance of each principal component, \mathbf{P}_m , is one. For dimension reduction of the input data \mathbf{S} , only the first L principal components will be preserved for the following processing. Here, for a simple and intuitive illustration of the proposed method, the obtained principal components are simply referred as $\text{PCA}(\mathbf{S}, L)$ in the following description, where L is the number of preserved principal components. Furthermore, only the basic principle and equations of the PCA are described in this section. If further details are required, please find a detailed tutorial of the PCA in [48].

III. PROPOSED METHOD

Fig. 1 shows the schematic of the proposed method, which consists of the following steps: First, by performing edge-preserving filtering on the dimension-reduced hyperspectral data, the standard multiparametric EPFs are constructed and stacked together. Then, we use the PCA to represent the stacked EPFs in the mean square sense and highlight the spectral differences. Finally, the SVM is used for the classification of the PCA-EPFs.

A. Spectral Dimensionality Reduction

In order to reduce the computing cost spent in the following feature extraction and reduce the image noise of the original HSI, the dimension of the HSI is first reduced with a simple band averaging method. Specifically, given a M -dimensional hyperspectral data \mathbf{I} , we first segment \mathbf{I} into K subgroups of equal size along the spectral dimension, $\mathbf{I} = \mathbf{I}_1, \dots, \mathbf{I}_K$, in which K is the number of bands in the dimension-reduced data. It should be noted that when M is indivisible by K , the K th subgroup will be constructed using the last $\lceil M/K \rceil$ bands of the HSI. Here, $\lceil M/K \rceil$ calculates the smallest integer not less than M/K , which is the number of bands in each subgroup when M is divisible by K . Then, an average band $\hat{\mathbf{I}}_k$ could be calculated for each subgroup of bands \mathbf{I}_k , so as to obtain a dimension-reduced hyperspectral data $\hat{\mathbf{I}} = \hat{\mathbf{I}}_1, \dots, \hat{\mathbf{I}}_K$.

One of the advantages of such an averaging-based dimension-reduction scheme is that the pixels of the dimension-reduced data are still related to the reflectance

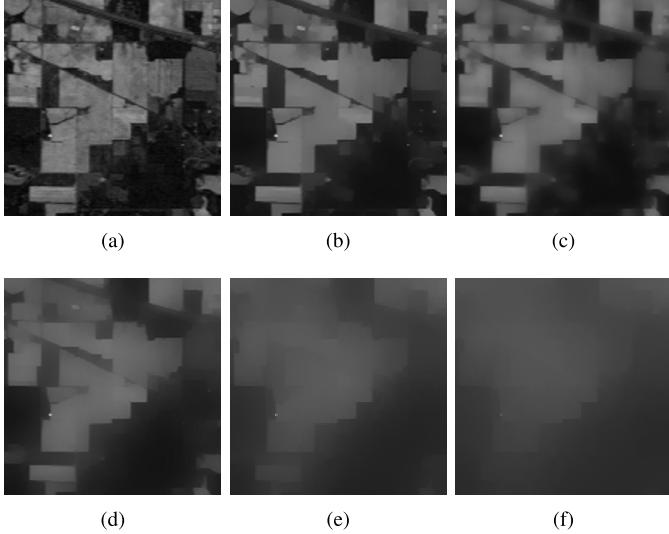


Fig. 2. Influence of the two parameters, i.e., δ_s and δ_r , to the performance of the filter. (a) Input hyperspectral band. (b)–(f) Filtered bands with different parameter settings.

of the scene. Although transform-based dimension-reduction methods, such as the PCA and ICA methods, also could serve as a dimension-reduction tool, edges and other important spatial structures in some principal and independent components may be seriously distorted, and thus decrease the performance of the following edge-preserving filtering. Aside from the transform-based methods, feature selection-based methods can be also adopted in this step. However, feature selection-based methods may require a time-consuming optimization process. In this paper, it is found that this simple averaging-based dimension-reduction scheme is already able to obtain satisfactory classification performances.

B. Feature Extraction With Multiple EPFs

With the K -dimensional hyperspectral data obtained earlier, the EPFs of the HSI can be easily calculated and stacked together as follows:

$$\mathbf{F}_k^x = \text{DTRF}(\hat{\mathbf{I}}_k, \delta_s^x, \delta_r^x), \quad x = 1, \dots, X, \quad k = 1, \dots, K \quad (7)$$

$$\mathbf{F} = \{\mathbf{F}^1, \dots, \mathbf{F}^X\} \quad (8)$$

where δ_s^x and δ_r^x are the x th parameter setting adopted for the domain transform recursive filter. Specifically, for the k th band of the dimension-reduced band, the EPFs are first obtained with different parameter settings $(\delta_s^x, \delta_r^x) \subset \{(\delta_s^1, \delta_r^1), \dots, (\delta_s^X, \delta_r^X)\}$. Then, the resulting features are stacked together according to (8).

The basic motivation behind this step is to capture the multiscale structural information preserved in the EPFs obtained with different parameter settings. As shown in Fig. 2, through adjusting the parameters of the domain transform recursive filter, the filtered images may look quite different in the preservation of edges and boundaries. On the one hand, the spectral separability of pixels belonging to different objects may be decreased when the blurring degree is relatively large. On the other hand, for smooth area with noise, filtering with

a high degree of smoothness could effectively reduce the noise and increase the spectral purity of pixels in such an area. It means that the EPFs obtained with different blurring degrees have their specific advantages in representing those objects or features of different scales. Therefore, by combining these features together, the complementary information preserved in the stacked edge-preserving features is expected to be useful in improving the classification accuracy.

C. Feature Fusion With the PCA

One simple solution to make full use of the complementary information in EPFs is to directly input the stacked EPFs into a classifier. However, as we mentioned in Section I, the high dimension of the stacked EPFs ($K \times X$ spectral dimensions) may lead to the Hughes phenomenon and extra computing burden in classification. Moreover, the resulting EPFs contain not only complementary information but also a large amount of redundant information. More importantly, although edge-preserving filtering is designed to smooth an image while preserving boundaries, the smoothing operation still may decrease the absolute differences among pixels belonging to different classes. To solve these problems, PCA is a very good tool, since it can not only reduce the dimension of the data but also able to increase the spectral separability of pixels, which is the most important nature of the PCA. Therefore, in this paper, the stacked EPFs are fused together as follows:

$$\mathbf{P} = \text{PCA}(\mathbf{F}, L) \quad (9)$$

where L is the number of principal components preserved in \mathbf{P} and \mathbf{F} is the stacked EPFs. In other words, the PCA is performed directly on the stacked EPFs, and thus, the $(X \times K)$ -dimensional EPFs will be transformed into L principal components. In the experiments, the influence of this parameter will be analyzed. The resulting PCA-EPFs, \mathbf{P} , can be directly used as the inputs of a supervised classifier. In the experiments, we analyze the performance of the PCA-EPFs using the SVM classifier, so as to do a fair comparison between the proposed method and the original edge-preserving filtering-based feature extraction method, i.e., the image fusion and recursive filtering (IFRF)-based method proposed in [32].

IV. EXPERIMENTS

In the experiments, multiple HSI classification methods, i.e., the SVM on the raw data set [12], the edge-preserving filtering-based postprocessing method (EPFP) [37], the multiple feature learning (MFL)-based method [49], the local binary patterns (LBPs)-based method [50], the SVM on EMAPs [26], the weighted Markov random fields (WMRFs)-based method [51], and the SVM on EPFs extracted using IFRF [32], are considered in comparison. These methods are adopted, because they are either highly cited in the hyperspectral remote sensing community or show the state-of-the-art performances on several real hyperspectral data sets. For the EPFP, MFL, LBP, and WMRF methods, the default parameter settings given in the corresponding publications are adopted. For the EMAP method, the attribute features are built using

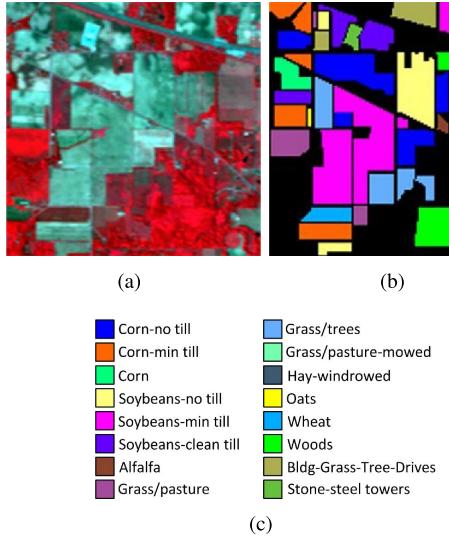


Fig. 3. Indian Pines data set. (a) Three-band color composite. (b) Reference data. (c) Class names.

threshold values in the range from 2.5% to 10% with respect to the mean of the individual features and with a step of 2.5% for the definition of the criteria based on the standard deviation attributes. Values of 100, 200, 500, and 1000 are selected as thresholds for the area attributes [49]. The SVM method is implemented using the LIBSVM library [52] by using the radial basis function kernel. Furthermore, the parameters of the SVM are determined using fivefold cross validation. All the experiments are performed using MATLAB on a computer with 3.5-GHz CPU and 8-GB Memory.

In order to address the small sample size problem, seven compared methods are evaluated with different sample sizes: the labeled samples are randomly picked out from the reference data to constitute the training set and the remaining samples are then used for test. To make the comparison fair, the reported values of overall accuracy (OA), average accuracy (AA), Kappa coefficient, and class individual accuracies are calculated by averaging the results obtained in ten repeated experiments with different randomly selected training samples. Moreover, both the mean and standard deviation of the accuracies obtained in repeated experiments are reported.

A. Data Sets

The first real data set is the Indian Pines image, which was acquired by the airborne visible infrared imaging spectrometer (AVIRIS) over the agricultural area of Northwestern Indiana in June 1992. It has a spatial dimension of 145×145 and 224 spectral bands, and its spatial resolution is 20 m per pixel. After discarding 24 low SNR and informationless bands affected by water and atmospheric absorption, 200 channels are preserved. A three-band false color image and the reference data are presented in Fig. 3. As shown in Fig. 3(b) and (c), the reference contains 16 land cover classes, most of which are different types of crops.

The second real data set is the University of Pavia image, which was acquired by the reflective optics system imaging spectrometer (ROSIS-03) optical sensor over an urban area

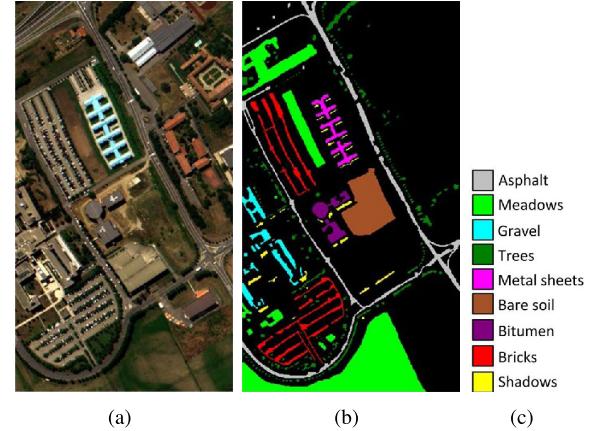


Fig. 4. University of Pavia data set. (a) Three-band color composite. (b) Reference data. (c) Class names.

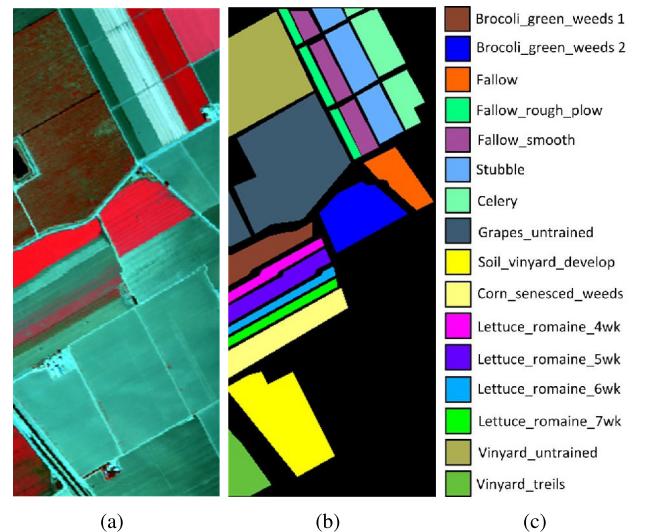


Fig. 5. Salinas data set. (a) Three-band color composite. (b) Reference data. (c) Class names.

surrounding the University of Pavia, Italy. It has a spatial dimension of 610×340 and 115 spectral bands, and its spatial resolution is 1.3 m per pixel. After 12 noisy bands are omitted, the image kept for processing contains 103 bands. A three-band false color image and the reference data are presented in Fig. 4. As shown in Fig. 4(b) and (c), the reference contains nine classes, most of which are man-made building objects.

The third real data set is the Salinas image, which was captured with AVIRIS over the agricultural area of Salinas Valley, California. After discarding 20 water absorption bands, this image contains 204 bands of size 512×217 . Fig. 5 shows the three-band false color image and the reference of the Salinas data set. It can be seen that the reference image contains 16 different classes, which represent mostly different types of crops.

B. Parameter and Component Analysis

First, a sensitivity analysis of the two important parameters of the proposed method, i.e., K (the band number of the dimension-reduced data) and L (the number of principal

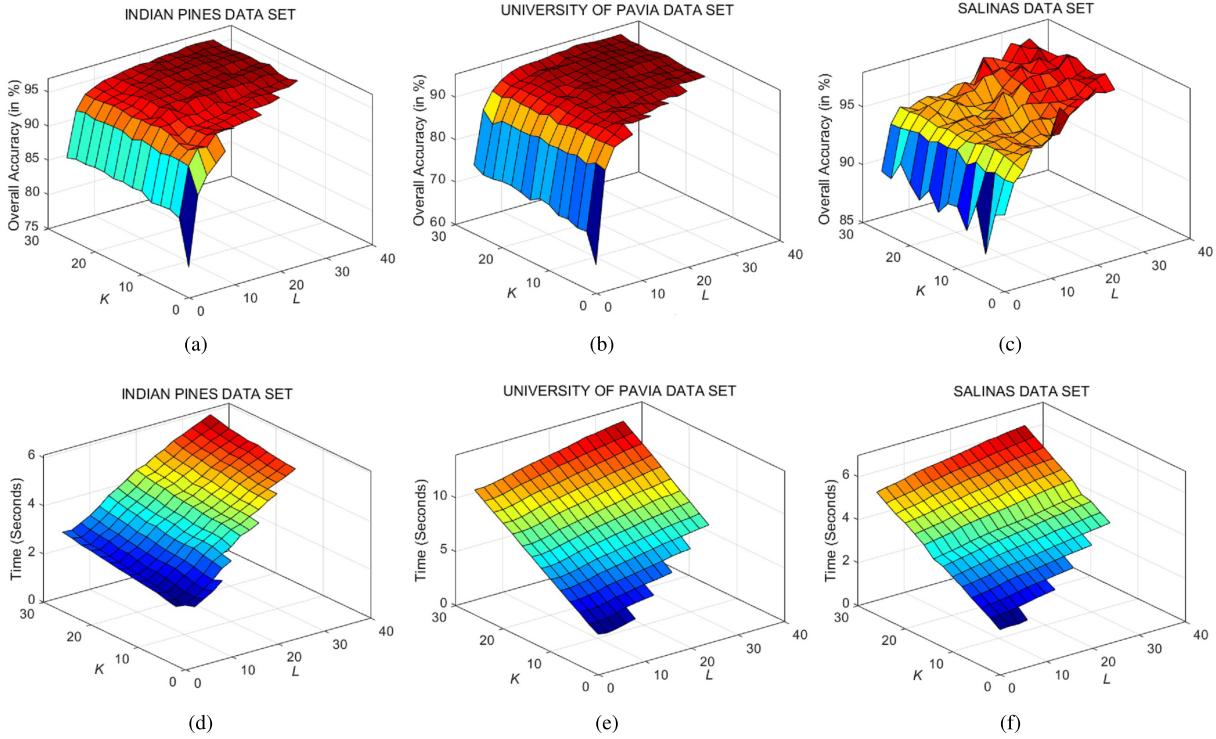


Fig. 6. Influence of parameters K and L to the performance of the proposed PCA-EPFs method. OAs recorded on (a) Indian Pines, (b) University of Paiva, and (c) Salinas data sets. Computing time recorded on (d) Indian Pines, (e) University of Pavia, and (f) Salinas data sets.

TABLE I
CLASSIFICATION PERFORMANCE OF THE SVM ON THE RAW DATA, THE EPFs OBTAINED WITH DIFFERENT PARAMETER SETTINGS, THE STACKED EPFs WITHOUT THE PCA-BASED FUSION STEP, AND THE PROPOSED PCA-EPFs

Methods	Raw Data	The K -dimensional EPFs obtained with different parameter settings			Stacked EPFs	PCA-EPFs
		$\delta_s = 30, \delta_r = 0.3$	$\delta_s = 115, \delta_r = 0.6$	$\delta_s = 200, \delta_r = 0.9$		
OA	63.15(2.56)	87.77(2.14)	81.68(2.72)	78.51(3.26)	90.77(2.11)	95.00(1.52)
AA	62.46(2.25)	83.82(3.10)	74.58(2.95)	71.79(3.11)	87.33(2.57)	91.86(2.15)
Kappa	54.15(2.78)	84.05(2.73)	76.20(3.37)	72.31(3.86)	87.96(2.65)	93.43(1.96)

components used for classification) is performed. It can be seen from Fig. 6 that OA generally becomes more stable as the two parameters increase. Moreover, it should be noted that the computing time of the proposed method also increases linearly as the two parameters increase. Specifically, for the Indian Pines data set, the computing time depends majorly on the parameter L [see Fig. 6(a)], since the feature extraction step will be quite efficient when the image is of a small size. However, for the University of Pavia and Salinas data sets, the computing time of the proposed method depends majorly on the parameter K [see Fig. 6(b) and (c)], since the feature extraction step will be much more time-consuming as the size of the image increases. Considering the tradeoff between computing efficiency and classification accuracy, K and L are set as 15 and 30 in the experiments. For the parameters used to construct the stacked EPFs, i.e., $\{(\delta_s^1, \delta_r^1), \dots, (\delta_s^X, \delta_r^X)\}$, they are set as $(\delta_s^1 = 30, \delta_r^1 = 0.3)$, $(\delta_s^2 = 115, \delta_r^2 = 0.6)$, and $(\delta_s^3 = 200, \delta_r^3 = 0.9)$ in the experiments. By varying the number of EPFs X and the parameters δ_s and δ_r , which control the smoothness of the resulting EPFs, we find that this fixed parameter setting given earlier is able to obtain satisfactory

classification performances for all three real data sets used in the experiments. Designing an automatic way to select the parameters, X , δ_s , and δ_r will be an interesting future research direction.

Then, we analyze the effects of two major components, i.e., the multiparameter EPF extraction and the PCA-based fusion of stacked EPFs, over the classification accuracies of the proposed method. The experiment is performed on the University of Pavia data set, and the number of training samples is set to be 20 for each class. As shown in Table I, the K -dimensional EPFs obtained with different parameter settings, i.e., $(\delta_s = 30, \delta_r = 0.3)$, $(\delta_s = 115, \delta_r = 0.6)$, and $(\delta_s = 200, \delta_r = 0.9)$, are all able to improve the classification accuracies in varying degrees. By stacking these features together, the classification accuracy can be further improved. Finally, as a combination of the PCA and stacked EPFs, the proposed PCA-EPFs show the best classification performance in terms of different accuracy indexes. This experiment demonstrates that the multiparameter EPF extraction and the PCA-based fusion of stacked EPFs both contribute significantly to the proposed PCA-EPFs method.

TABLE II

PERFORMANCE OF THE PROPOSED METHOD USING DIFFERENT METHODS FOR FEATURE FUSION. THE OA, AA, AND KAPPA COEFFICIENT ARE RECORDED IN %. COMPUTING TIME IS RECORDED IN SECONDS. THE EXPERIMENT IS PERFORMED ON THE INDIAN PINES DATA SET AND THE NUMBER OF TRAINING SAMPLES IS SET TO BE 1% OF THE REFERENCE

Methods	LDA [53]	ICA [54]	PPCA [55]	KPCA [56]	PCA [48]
OA	76.68	80.27	82.37	83.24	83.57
AA	72.69	81.83	87.13	82.95	88.23
Kappa	73.82	77.63	80.03	81.07	81.41
Time	5.37	2.68	233	633	2.59

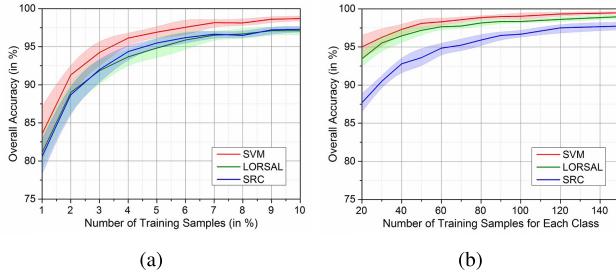


Fig. 7. OAs of the SVM [12], LORSAL [57], and SRC [13] classifiers on (a) Indian Pines and (b) University of Pavia data sets as functions of the number of labeled samples per class (varying from 1% to 10% for the Indian Pines data set and varying from 20 to 150 for the University of Pavia data set). The widths of the line areas refer to the standard variances of accuracies obtained in ten repeated experiments.

Next, an experiment is performed to demonstrate the effectiveness of the PCA in fusion of stacked EPFs. This experiment is performed on the Indian Pines data set, and the number of training samples is set to be 1% of the reference. As shown in Table II, compared with other widely used dimension-reduction methods, i.e., the linear discriminant analysis [53], ICA [54], probabilistic PCA [55], and kernel PCA (KPCA) [56] methods, the adopted PCA method is simple, efficient, and, more importantly, able to obtain the best classification performance in terms of classification accuracies. This experiment further demonstrates that although advancing dimension-reduction tools may be more effective in analyzing the original hyperspectral data with noise, textures, and other complex spatial and spectral structures, the original PCA method is more suitable for analyzing EPFs, in which insignificant structures, such as noise and texture, have been removed.

At last, the performance of the proposed PCA-EPFs method is also evaluated using different spectral classifiers, i.e., the LOgistic Regression via Splitting and Augmented Lagrangian (LORSAL) [57], and the sparse representation-based classifier (SRC) [13]. Fig. 7(a) and (b) shows the OAs of different classifiers on the Indian Pines and University of Pavia data sets, respectively. Specifically, when the number of training samples per each class is varying from 1% to 10% for the Indian Pines data set and from 20 to 150 for the University of Pavia data set, it can be seen that the PCA-EPFs method can lead to high OAs for different classifiers, while the SVM classifier always shows the best performance. The reason is that the SVM method usually performs better than the

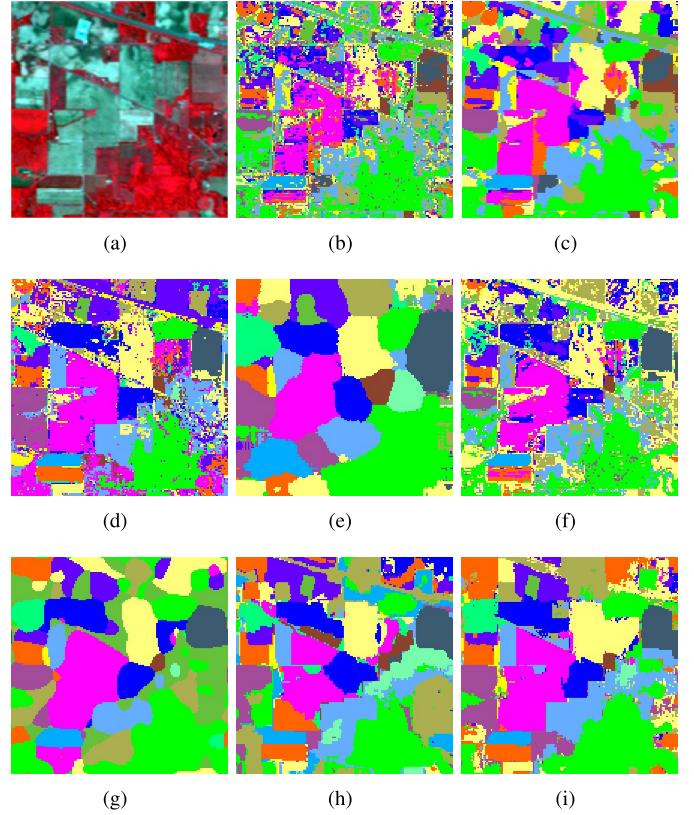


Fig. 8. Classification maps obtained by different methods on the Indian Pines data set. (a) False color composite of the hyperspectral image. Classification maps obtained by (b) SVM, OA = 53.32%, (c) EPFP, OA = 64.33%, (d) MFL, 76.67%, (e) LBP, OA = 72.20%, (f) EMAP, OA = 66.38%, (g) WMRF, OA = 78.89%, (h) IFRF, OA = 74.53%, and (i) PCA-EPFs OA = 86.28%.

LORSAL and SRC methods on several real hyperspectral data sets [14], [57]. Therefore, the SVM classifier is adopted for the proposed PCA-EPFs method in the following experiments.

C. Comparison of Different Methods

In this section, experiment is first performed on the Indian Pines data set. When the training samples account for 1% of the ground truth, the classification accuracies for each class, the OA, AA, and Kappa coefficient calculated using different methods, are shown in Table III. From Table III, it can be seen that the results yielded by the proposed PCA-EPFs method are better than those obtained by other methods in terms of the highest OA, AA, and Kappa coefficient. Fig. 8 shows the false color composite of the HSI and the classification maps obtained by different methods. It can be visually observed that the edges and boundaries in the map of the PCA-EPFs are in better accordance with the false color composite. By contrast, the LBP and WMRF methods may lead to oversmoothed classification maps. For the LBPs-based method, a slack variable is introduced so as to enable the difference between the central pixel and the neighboring ones varying in a small interval. This step could greatly decrease the impact of spectral variability and noise. However, when the number of training samples is limited, this step may lead to oversmoothed classification results. For the WMRF method,

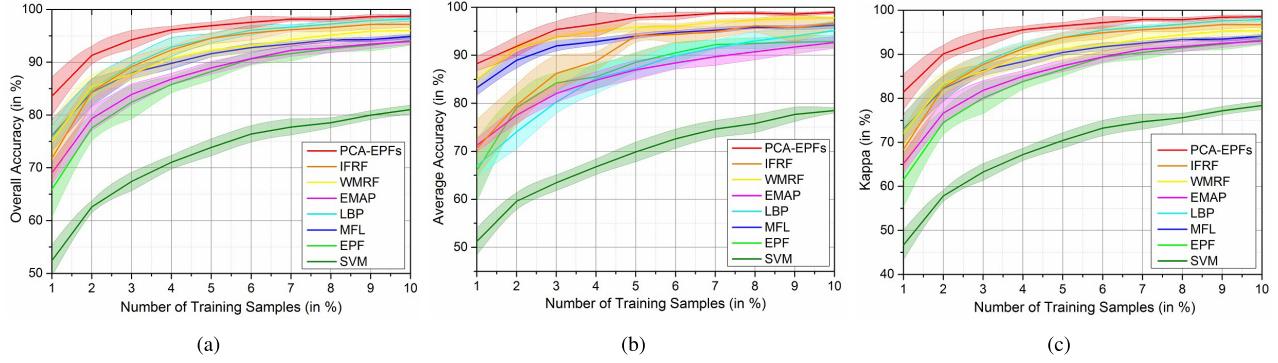


Fig. 9. Performance of the compared methods on the Indian Pines data set with different number of training samples (varying from 1% to 10%). (a) OA. (b) AA. (c) Kappa coefficient. The widths of the line areas refer to the standard variances of accuracies obtained in ten repeated experiments.

TABLE III
CLASSIFICATION PERFORMANCE OF THE SVM, EPF, MFL, LBP, EMAP, WMRF, IFRF, AND PCA-EPFs METHODS FOR THE INDIAN PINES DATA SET WITH 1% TRAINING SAMPLES. NUMBERS IN THE PARENTHESIS INDICATE THE STANDARD VARIANCE OF THE ACCURACIES OBTAINED IN REPEATED EXPERIMENTS

Class	train	test	Classification accuracies obtained by the compared methods (in %)							
			SVM [12]	EPF [37]	MFL [49]	LBP [50]	EMAP [26]	WMRF [51]	IFRF [32]	PCA-EPFs
Alfalfa	6	40	27.98(13.4)	48.87(30.4)	95.74(1.22)	28.83(15.1)	91.92(20.5)	95.99(4.41)	49.14(32.7)	99.52 (1.51)
Corn_N	7	1421	47.19(8.50)	62.06(16.4)	63.29(6.81)	76.10 (10.3)	57.15(9.11)	64.46(7.96)	65.64(10.1)	74.95(10.5)
Corn_M	6	824	40.81(7.47)	61.73(14.9)	66.02(6.35)	74.93 (8.87)	63.22(7.5)	65.87(6.78)	54.99(9.02)	72.34(12.6)
Corn	6	231	26.49(6.65)	40.41(13.8)	73.02(12.9)	67.10(11.5)	38.81(10.1)	95.02 (12.2)	58.86(20.4)	76.78(12.8)
Grass_M	6	477	59.66(17.9)	79.83(22.0)	77.14(9.16)	82.18(14.1)	75.03(18.7)	77.12(7.48)	80.8(14.08)	89.66 (12.7)
Grass_T	6	724	78.89(4.53)	80.94(7.78)	92.33(5.67)	64.08(10.0)	94.8(2.11)	96.33 (7.83)	86.91(8.23)	92.75(5.67)
Grass_P	6	22	26.49(15.1)	59.71(2.20)	96.82(2.20)	65.08(29.3)	57.89(19.3)	100.0 (3.83)	49.08(30.2)	84.30(26.7)
Hay_W	7	471	93.66(4.77)	98.34(5.24)	98.54(2.71)	89.25(7.27)	100.0 (0.00)	97.39(11.9)	100.0 (0.00)	99.98(0.07)
Oats	6	14	10.20(5.07)	32.80(27.7)	100.0 (0.00)	22.73(14.8)	41.06(9.35)	100.0 (9.04)	32.56(20.8)	86.86(29.3)
Soybean_N	7	965	43.07(7.97)	56.30(11.6)	75.16 (10.5)	73.68(12.7)	51.91(9.64)	74.42(8.37)	59.73(16.8)	68.53(18.3)
Soybean_M	8	2447	62.56(6.33)	74.62(6.57)	75.04(15.1)	90.45(4.90)	80.20(7.38)	61.98(13.4)	81.02(10.7)	91.55 (3.86)
soybean_C	6	587	26.87(7.24)	39.31(15.5)	72.85(13.2)	79.58(14.0)	44.84(9.98)	76.72(13.1)	69.23(12.1)	87.70 (10.5)
Wheat	6	199	78.33(3.02)	95.02(4.56)	99.45(0.16)	46.74(7.76)	93.39(5.49)	100.0 (0.29)	71.23(19.6)	100.0 (0.00)
Woods	6	1259	84.28(5.48)	91.13(7.49)	86.99(6.43)	93.58(2.83)	92.45(2.75)	88.47(13.0)	96.81(2.65)	98.65 (1.51)
Buildings	6	380	28.94(8.30)	55.13(19.5)	72.50(6.03)	76.49(10.3)	62.62(12.6)	67.43(9.29)	79.67(7.68)	94.33 (8.32)
Stone	7	86	83.64(19.6)	79.38(8.90)	91.02(6.24)	40.67(9.81)	94.78(2.68)	98.05 (6.97)	93.39(13.4)	93.79(9.21)
OA			52.42(2.94)	65.91(5.25)	76.12(3.02)	75.65(2.21)	69.02(2.14)	75.05(3.75)	71.82(4.28)	83.57 (3.60)
AA			51.19(2.77)	65.97(6.51)	83.24(1.55)	66.96(2.22)	71.25(1.38)	84.95(1.60)	70.57(5.95)	88.23 (1.47)
Kappa			46.67(3.37)	61.50(6.12)	73.03(3.29)	72.63(2.43)	65.14(2.29)	71.99(3.90)	68.37(4.87)	81.41 (4.00)

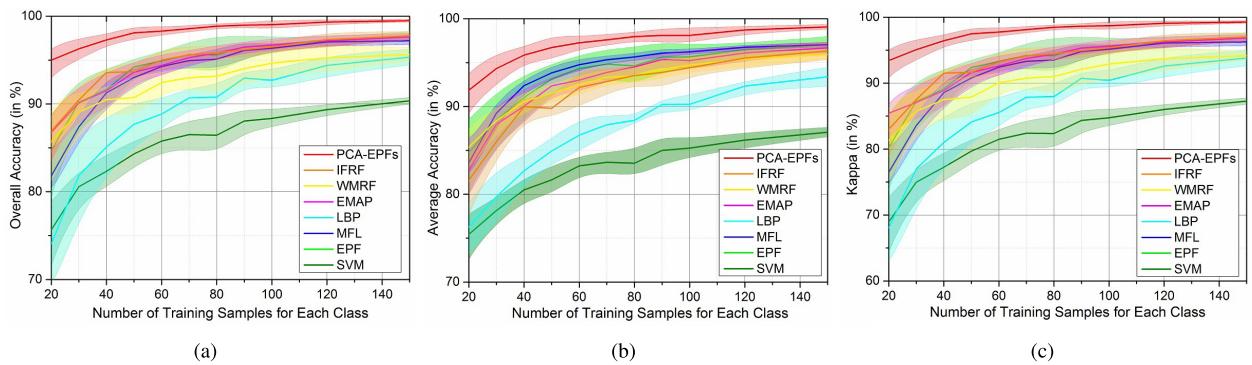


Fig. 10. Performance of the compared methods on the University of Pavia data set as functions of the number of labeled samples per class (varying from 20 to 150). (a) OA. (b) AA. (c) Kappa coefficient. The widths of the line areas refer to the standard variances of accuracies obtained in ten repeated experiments.

the true labels of training samples serve as an additional constraint in the WMRF model. By doing this, the training samples will spread their labels from their positions to the whole data until a stable status is achieved. This constraint

is helpful in improving classification accuracy while it may produce oversmoothed classification maps when the number of training samples is limited. Furthermore, it can be found that the SVM and EMAP methods tend to produce noisy labels

TABLE IV

CLASSIFICATION PERFORMANCE OF THE SVM, EPF, MFL, LBP, EMAP, WMRF, IFRF, AND PCA-EPFs METHODS FOR THE UNIVERSITY OF PAVIA DATA SET WITH 20 LABELED SAMPLES PER CLASS AS THE TRAINING SET. NUMBERS IN THE PARENTHESIS INDICATES THE STANDARD VARIANCE OF THE ACCURACIES OBTAINED IN REPEATED EXPERIMENTS

Class	train	test	Classification accuracies obtained by the compared methods (in %)							
			SVM [12]	EPF [37]	MFL [49]	LBP [50]	EMAP [26]	WMRF [51]	IFRF [32]	PCA-EPFs
Asphalt	20	6611	93.29(2.46)	96.94 (1.67)	80.19(2.12)	92.36(4.89)	95.26(1.12)	83.30(6.36)	82.05(5.50)	93.52(4.76)
Meadows	20	18629	88.51(1.74)	92.43(3.15)	84.49(3.27)	99.08(0.70)	94.73(2.05)	87.64(7.11)	97.98(0.90)	99.60 (0.15)
Gravel	20	2079	57.31(7.63)	81.88(13.2)	67.03(9.05)	69.69(17.7)	63.21(7.92)	82.10(12.2)	74.67(7.90)	90.00 (3.75)
Tress	20	3044	71.85(11.8)	78.61(14.6)	94.28 (2.42)	30.05(7.03)	87.94(10.3)	91.34(3.34)	78.17(7.78)	89.36(6.33)
Sheets	20	1325	92.68(4.79)	94.37(4.14)	98.35(1.01)	86.29(3.10)	98.58(0.91)	99.95 (0.09)	99.64(0.74)	92.04(7.35)
Soil	20	5009	48.79(7.25)	61.77(10.2)	76.63(4.93)	89.50(8.76)	68.06(8.19)	81.00(17.1)	86.72(5.42)	97.33 (4.48)
Bitumen	20	1310	49.39(4.60)	71.67(9.42)	89.83(4.18)	91.37(7.65)	74.97(14.7)	99.27 (0.36)	82.07(9.66)	89.41(6.54)
Bricks	20	3662	77.27(4.61)	86.85(6.43)	63.28(8.71)	68.89(13.0)	86.56(2.87)	79.88(14.6)	73.02(7.10)	87.26 (6.13)
Shadows	20	927	99.88 (0.12)	98.25(1.15)	98.22(2.19)	59.35(20.9)	99.85(0.08)	64.03(12.4)	60.12(14.8)	88.23(8.35)
OA			75.66(3.31)	84.70(4.11)	87.38(2.17)	74.07(5.51)	86.78(2.38)	85.74(4.46)	86.93(2.43)	95.00 (1.52)
AA			75.44(2.39)	84.75(3.81)	89.25(2.02)	76.29(2.90)	82.80(3.01)	85.39(2.60)	81.60(3.50)	91.86 (2.15)
Kappa			68.96(3.88)	80.26(5.07)	83.59(2.68)	68.04(6.06)	85.46(2.47)	81.36(6.37)	83.03(3.04)	93.43 (1.96)

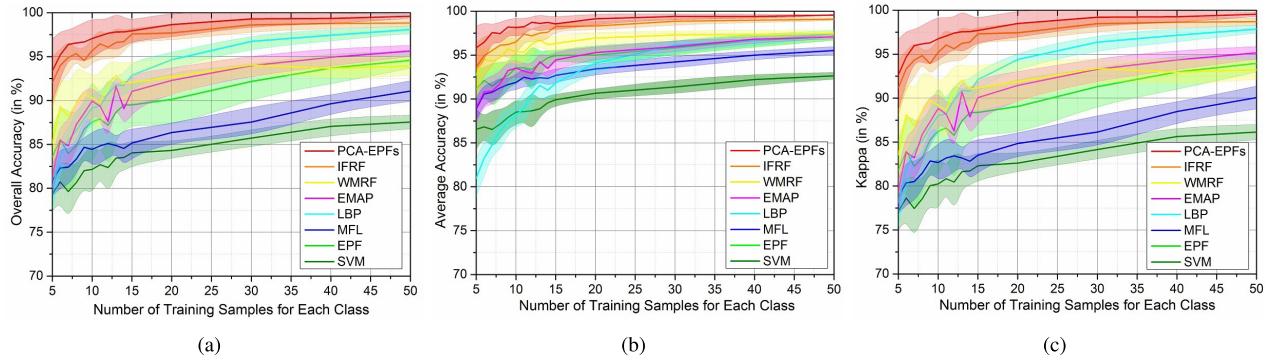


Fig. 11. Performance of the compared methods on the Salinas data set as functions of the number of labeled samples per class (varying from 5 to 50). (a) OA. (b) AA. (c) Kappa coefficient. The widths of the line areas refer to the standard variances of accuracies obtained in then repeated experiments.

TABLE V

CLASSIFICATION PERFORMANCE OF THE SVM, EPF, MFL, LBP, EMAP, WMRF, IFRF, AND PCA-EPFs METHODS FOR THE SALINAS DATA SET WITH TEN LABELED SAMPLES PER CLASS AS THE TRAINING SET. NUMBERS IN THE PARENTHESIS INDICATE THE STANDARD VARIANCE OF THE ACCURACIES OBTAINED IN REPEATED EXPERIMENTS

Class	train	test	Classification accuracies obtained by the compared methods (in %)							
			SVM [12]	EPF [37]	MFL [49]	LBP [50]	EMAP [26]	WMRF [51]	IFRF [32]	PCA-EPFs
Weeds_1	10	1999	98.96(1.10)	100.0 (0.00)	99.85(0.06)	97.42(2.22)	99.95(0.05)	99.80(1.03)	98.52(4.02)	100.0 (0.00)
Weeds_2	10	3716	98.92(0.41)	99.93(0.19)	99.58(0.13)	99.14(0.84)	99.92(0.12)	100.0 (0.79)	99.99(0.02)	99.79(0.26)
Fallow	10	1966	87.44(3.64)	92.51(4.15)	99.79 (0.39)	82.40(13.0)	96.30(1.05)	91.65(3.36)	99.64(0.52)	98.58(2.74)
Fallow_P	10	1384	96.75(1.04)	97.39(0.33)	96.62(2.65)	65.01(6.66)	97.13(0.56)	99.05 (0.41)	92.39(5.22)	95.81(1.94)
Fallow_S	10	2668	98.30(1.18)	99.91(0.07)	98.65(0.07)	95.26(3.58)	99.57(0.21)	97.96(2.47)	99.26(1.34)	99.99 (0.02)
Stubble	10	3949	99.98(0.03)	99.98(0.02)	99.17(0.72)	91.56(4.66)	99.46(0.49)	100.0 (0.08)	99.99(0.02)	99.74(0.28)
Celery	10	3569	96.49(2.85)	97.47(3.02)	99.75(0.07)	93.11(5.37)	97.65(2.81)	99.95 (0.24)	98.50(2.28)	99.50(0.91)
Grapes	10	11261	68.51(5.91)	79.19(11.1)	61.84(12.4)	96.18(3.13)	84.49(5.66)	78.35(8.17)	98.94(1.65)	99.43 (0.57)
Soil	10	6193	98.70(0.86)	99.26(0.42)	99.66(0.31)	99.51(1.09)	98.86(0.61)	99.92 (0.77)	99.72(0.27)	99.57(0.40)
Corn	10	3268	78.68(7.55)	87.62(5.52)	87.93(6.36)	97.96(2.07)	84.02(8.75)	93.07(4.34)	99.68(0.38)	99.73 (0.33)
Lettuce_4	10	1058	85.43(6.45)	94.29(4.75)	90.95(1.90)	84.16(5.29)	87.45(8.17)	99.86 (2.56)	97.28(2.10)	99.83(0.19)
Lettuce_5	10	1917	93.64(5.67)	97.67(5.78)	99.78 (0.11)	96.68(2.83)	93.91(2.86)	99.79(2.85)	97.87(2.03)	99.32(1.17)
Lettuce_6	10	906	89.82(5.32)	97.49(1.92)	99.18(0.30)	74.94(7.99)	96.17(3.23)	99.49 (1.08)	91.19(12.7)	97.05(4.78)
Lettuce_7	10	1060	84.07(15.5)	95.14(6.43)	89.06(3.29)	73.94(9.42)	97.00(1.87)	99.35 (3.34)	92.47(7.98)	96.32(5.90)
Vinyard_U	10	7258	49.67(5.81)	61.36(14.5)	56.24(9.81)	76.35(5.95)	71.50(8.12)	68.41(16.6)	81.45(4.01)	85.95 (9.23)
Vinyard_T	10	1797	90.87(7.96)	97.82(4.54)	91.98(7.53)	86.12(12.3)	92.75(11.0)	99.25(2.42)	97.26(6.03)	99.98 (0.07)
OA			82.15(2.94)	87.58(5.21)	84.43(1.92)	89.16(1.68)	89.94(1.62)	90.31(2.43)	95.69(1.30)	97.06 (2.19)
AA			88.51(1.47)	93.56(1.89)	91.88(0.87)	88.11(1.31)	93.51(0.96)	95.37(1.34)	96.51(1.49)	98.16 (1.05)
Kappa			80.23(3.18)	86.25(5.72)	82.69(2.09)	88.01(1.83)	88.83(1.78)	89.21(2.76)	95.21(1.44)	96.74 (2.42)

in the classification results. The OA of each classification map is also reported in Fig. 8. It can be seen that the proposed PCA-EPFs method obtains the highest OA for this example.

Furthermore, the influence of the number of training samples to the performance of different methods is also analyzed. Fig. 9 shows the OA, AA, and Kappa coefficient of different

methods as functions of the number of training samples (ranging from 1% to 10%). As shown in Fig. 9, as expected, the increase in the training size has a favorable effect on the performance of all classification methods. More importantly, compared with the IFRF method, the proposed PCA-EPFs show obvious improvements, especially when the number of training samples is quite limited.

TABLE VI

COMPUTING TIME OF THE SVM, EPF, MFL, LBP, EMAP, WMRF, IFRF, AND PCA-EPFs METHODS FOR THE INDIAN PINES, UNIVERSITY OF PAVIA, AND SALINAS DATA SETS WHEN THE NUMBER OF TRAINING SAMPLES IS SET ACCORDING TO AS THOSE SHOWN IN TABLES III–V

Data sets	Computing time of the compared methods (in seconds)							
	SVM [12]	EPF [37]	MFL [49]	LBP [50]	EMAP [26]	WMRF [51]	IFRF [32]	PCA-EPFs
Indian Pines	3.40	3.61	1.98	67.16	3.44	28.90	1.66	2.47
University of Pavia	8.41	9.87	15.93	710.25	21.03	101.70	15.96	13.38
Salinas	12.71	13.37	7.97	396.88	10.66	7.80	7.80	7.60

The second experiment is performed on the University of Pavia data set. Fig. 10 shows the OA, AA, and Kappa coefficient as functions of the training size for the samples per class (from 20 to 150). Likewise, the classification accuracies improve as the number of training samples increases and the results obtained by the PCA-EPFs are better than those of other compared methods. This is in accordance with the previous experiment performed on the Indian Pines data set. Moreover, Table IV shows the classification accuracies, OA, AA, and Kappa coefficient obtained with different methods when the number of training samples is 20 for each class. It can be seen that the proposed method obtains the highest accuracies for four different classes and the highest OA, AA, and Kappa coefficient.

The third experiment is performed on the Salinas data set. As before, the classification results using different methods as functions of the number of training samples per class (from 5 to 50) are given in Fig. 11. The accuracies of different methods with ten labeled samples per class as training set are summarized in Table V. From Fig. 11 and Table V, it can be seen that the proposed method still shows competitive performances in terms of OA, AA, and Kappa coefficient.

Finally, the computing time (in seconds) of the eight techniques on the three real hyperspectral data sets is presented in Table VI. The computation time was recorded by a computer with a two-core processor of 2.8 GHz and a 8-GB RAM, and the results are obtained when the number of training samples is set as those presented in Tables III–V. It can be observed from Table VI that the computational cost of the proposed PCA-EPFs method is quite competitive even when it is compared with the original single parameter-based EPF extraction method, i.e., the IFRF method.

V. CONCLUSION

In this paper, a novel PCA-EPF extraction method is proposed for the classification of HSIs. The proposed method consists of two stages, i.e., standard edge-preserving filtering-based feature extraction and the PCA-based feature fusion. The advantage of PCA is that it can not only effectively fuse the EPFs obtained with different parameter settings but also increase the class separability in EPFs. Experimental results reported in this paper show that, compared with the standard EPFs and other recently proposed HSI classification methods, the proposed PCA-EPFs method is able to obtain much higher classification accuracies on three real hyperspectral data sets, especially when the number of training samples is limited. However, the disadvantage of the proposed method is that it selects the parameters of the EPFs manually, which decreases the operational applicability of the proposed method

in real application. Therefore, how to automatically decide the optimal parameters for the edge-preserving filters will be the focus of our future work.

ACKNOWLEDGMENT

The authors would like to thank Prof. J. Li for providing the code of the multiple feature learning-based method, Dr. L. Sun for providing the code of the weighted Markov random fields-based method, and Prof. W. Li for providing the code of the local binary patterns-based method.

REFERENCES

- [1] Y. Liu, G. Gao, and Y. Gu, “Tensor matched subspace detector for hyperspectral target detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1967–1974, Apr. 2017.
- [2] 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.
- [3] F. Kizel, M. Shoshany, N. S. Netanyahu, G. Even-Tzur, and J. A. Benediktsson, “A stepwise analytical projected gradient descent search for hyperspectral unmixing and its code vectorization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4925–4943, Sep. 2017.
- [4] J. Li, I. Dpido, P. Gamba, and A. Plaza, “Complementarity of discriminative classifiers and spectral unmixing techniques for the interpretation of hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2899–2912, May 2015.
- [5] T. Lu, S. Li, L. Fang, L. Bruzzone, and J. A. Benediktsson, “Set-to-set distance-based spectral-spatial classification of hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7122–7134, Dec. 2016.
- [6] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, “Extended random walker-based classification of hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.
- [7] B. Sun, X. Kang, S. Li, and J. A. Benediktsson, “Random-walker-based collaborative learning for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 212–222, Jan. 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] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, “Semantic annotation of high-resolution satellite images via weakly supervised learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3660–3671, Jun. 2016.
- [10] 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.
- [11] Y. Bazi and F. Melgani, “Gaussian process approach to remote sensing image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 186–197, Jan. 2010.
- [12] 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.
- [13] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Hyperspectral image classification using dictionary-based sparse representation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [14] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, “Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.

- [15] 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.
- [16] 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.
- [17] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.
- [18] B. Mojarradi, H. Abrishami-Moghaddam, M. J. V. Zoj, 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] 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.
- [20] B. Demir and S. Erturk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4071–4084, Nov. 2010.
- [21] J. Zabalza, J. Ren, Z. Wang, S. Marshall, and J. Wang, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1886–1890, Nov. 2014.
- [22] M. Faivel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] J. Zabalza *et al.*, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418–4433, Aug. 2015.
- [30] 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.
- [31] X. Jin and Y. Gu, "Superpixel-based intrinsic image decomposition of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4285–4295, Aug. 2017.
- [32] 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.
- [33] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 504–511, Feb. 2013.
- [34] 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.
- [35] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 626–632, May 2012.
- [36] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
- [37] 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.
- [38] J. Xia, L. Bombrun, T. Adal, Y. Berthoumieu, and C. Germain, "Spectral-spatial classification of hyperspectral images using ICA and edge-preserving filter via an ensemble strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4971–4982, Aug. 2016.
- [39] B. Pan, Z. Shi, and X. Xu, "Hierarchical guidance filtering-based ensemble classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4177–4189, Jul. 2017.
- [40] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th IEEE Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.
- [41] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [42] J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," *ACM Trans. Graph.*, vol. 26, no. 3, p. 103, Jul. 2007.
- [43] A. Criminisi, T. Sharp, C. Rother, and P. Pérez, "Geodesic image and video editing," *ACM Trans. Graph.*, vol. 29, no. 5, pp. 134:1–134:15, Oct. 2010.
- [44] Q. Yang, K.-H. Tan, and N. Ahuja, "Real-time O(1) bilateral filtering," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 557–564.
- [45] 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.
- [46] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, p. 69, 2011.
- [47] L. Xu, J. Ren, Q. Yan, R. Liao, and J. Jia, "Deep edge-aware filters," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1669–1678.
- [48] L. I. Smith *et al.* (Feb. 2002). A tutorial on principal components analysis. Cornell Univ., Ithaca, NY, USA. [Online]. Available: <http://kybele.psych.cornell.edu/~Psych-465-Spring-2003/PCA-tutorial.pdf>
- [49] 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.
- [50] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.
- [51] 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.
- [52] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.
- [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, 2017. [Online]. Available: <http://www.cs.nyu.edu/~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, 1999.
- [56] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Proc. Int. Conf. Artif. Neural Netw.*, 1997, pp. 583–588.
- [57] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.



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. His research interests include hyperspectral feature extraction, image classification, image fusion, and anomaly detection.

Dr. Kang received the Second Prize in the Student Paper Competition in IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2014. In IGARSS 2017, he was selected as a Best Reviewer for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2016.



Xuanlin Xiang received the B.Sc. degree from Guangxi University, Nanning, China, in 2015. She is currently pursuing the M.S. degree in electrical engineering with Hunan University, Changsha, China.

Her research interests include hyperspectral image classification.



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

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. In 2001, he was also a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong. From 2002 to 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, Surrey, U.K. In 2005, he was a Visiting Professor with the Department of Computer Science, The Hong Kong University of Science and Technology. In 2013, he was granted the National Science Fund for Distinguished Young Scholars in China. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He is also a Chang-Jiang Scholar Professor appointed by the Ministry of Education of China. He has authored or co-authored over 180 refereed papers. His research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li was a recipient of two Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006. He is a member of the Editorial Board of the *Information Fusion, Sensing and Imaging*. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.



Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984, and the M.S.E.E. and Ph.D. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

He is currently the Rector and a Professor of electrical and computer engineering with the University of Iceland. He is also a Co-Founder of the biomedical start-up company Oxymap, Reykjavik.

His research interests include remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing. He has published extensively in those fields.

Dr. Benediktsson is a fellow of the SPIE. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica, and Tau Beta Pi. He received the Stevan J. Kristof Award from Purdue University in 1991 as an Outstanding Graduate Student in remote sensing. He was a recipient of the Icelandic Research Council's Outstanding Young Researcher Award in 1997. He received the IEEE Third Millennium Medal in 2000. He was a co-recipient of the University of Iceland's Technology Innovation Award in 2004. He received the yearly research award from the Engineering Research Institute, University of Iceland, in 2006, and the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society in 2007. He was a co-recipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award. He received the 2013 IEEE/VFI Electrical Engineer of the Year Award, and in 2013, he was a co-recipient of the IEEE Geoscience and Remote Sensing Society (GRSS) Highest Impact Paper Award. He was the President of the IEEE GRSS from 2011 to 2012. He has been on GRSS AdCom since 2000. He was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2007 to 2010. He was an Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008. He has been serving as an Associate Editor of TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and the IEEE ACCESS since 2013. He is on the International Editorial Board of the *International Journal of Image and Data Fusion*.