This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A New Spatial–Spectral Feature Extraction Method for Hyperspectral Images Using Local Covariance Matrix Representation

Leyuan Fang, *Senior Member, IEEE*, Nanjun He, *Student Member, IEEE*, Shutao Li, *Senior Member, IEEE*, Antonio J. Plaza, *Fellow, IEEE*, and Javier Plaza, *Senior Member, IEEE*

***Abstract*— In this paper, a novel local covariance matrix (CM) representation method is proposed to fully characterize the correlation among different spectral bands and the spatial–contextual information in the scene when conducting feature extraction (FE) from hyperspectral images (HSIs). Specifically, our method first projects the HSI into a subspace, using the maximum noise fraction method. Then, for each test pixel in the subspace, its most similar neighboring pixels (within a local spatial window) are clustered using the cosine distance measurement. The test pixel and its neighbors are used to calculate a local CM for FE purposes. Each nondiagonal entry in the matrix characterizes the correlation between different spectral bands. Finally, these matrices are used as spatial–spectral features and fed to a support vector machine for classification purposes. The proposed method offers a new strategy to characterize the spatial–spectral information in the HSI prior to classification. Experimental results have been conducted using three publicly available hyperspectral data sets for classification, indicating that the proposed method can outperform several state-of-the-art techniques, especially when the training samples available are limited.**

***Index Terms*— Covariance matrix representation (CMR), feature extraction (FE), hyperspectral image (HSI) classification, manifold space (MS).**

I. Introduction

DUE to the advancement in the hyperspectral imaging technology, hyperspectral images (HSIs) comprise hundreds of narrow spectral bands. Each pixel in the HSI is a high-dimensional vector (or spectral signature) that provides discriminative spectral information that can be used to distinguish between materials of interest [1]–[3]. With such rich spectral information, HSIs have been widely used in many different applications, such as target detection [4], change detection [5], and classification [6]–[8]. Among these problems, the classification of HSIs has been the focus of many recent research efforts, since it plays an important role in precision agriculture [9], urban mapping [10], and environmental monitoring [11]. Given a set of training samples, the classification of HSIs aims to assign a unique label for each test pixel in the scene. To achieve this, various classifiers have been adopted by the hyperspectral imaging community, including the support vector machine (SVM) [12], neural networks [13], random forests [14], and sparse representation techniques [15]. However, when the training samples available *a priori* are limited, the classification accuracy achieved by these methods [12]–[14] can significantly degrade due to the so-called Hughes effect or the curse of dimensionality.

To tackle this problem, many feature extraction (FE) methods have been developed in recent years. They aim at finding a new subspace in which the separability among the transformed samples can be enlarged as much as possible. Typical FE methods include unsupervised approaches (e.g., the principal component analysis (PCA) [16], independent component analysis (ICA) [17], the maximum noise fraction (MNF) [18]), and supervised approaches (e.g., linear discriminant analysis [19]). The PCA maximizes the variance of the projected samples with an orthogonal set of vectors. However, since only a small number of PCs are preserved, the PCA cannot ensure that the most informative spectral signatures are retained. Different from the PCA, the MNF aims at maximizing the signal-to-noise ratio (SNR) instead of the variance. This can remove the noise of the HSI more effectively in the process of dimensionality reduction.

In order to incorporate the spatial information in the classification of HSIs, many spatial–spectral FE techniques have been proposed in recent decades. Benediktsson *et al.* [20] introduced the extended morphological profile (EMP) to adaptively extract the spatial information based on the structure of HSIs using morphological opening and closing operations. In [21], based on the concept of morphological attribute filters, the spatial information of HSIs was characterized with a multilevel multiattribute approach termed as extended

Manuscript received November 25, 2017; revised January 23, 2018; accepted January 29, 2018. This work was supported by 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, in part by the National Natural Science Foundation for Young Scientist of China under Grant 61501180, in part by the Fund of Hunan Province for Science and Technology Plan Project under Grant 2017RS3024, and in part by the China Scholarship Council. *(Corresponding author: Shutao Li.)*

L. Fang and S. Li are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: fangleyuan@gmail.com; shutao_li@hnu.edu.cn).

N. He is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, E-10003 Cáceres, Spain (henanjun@hnu.edu.cn).

A. J. Plaza and J. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, E-10003 Cáceres, Spain (e-mail: aplaza@unex.es; jplaza@unex.es).

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.2018.2801387

0196-2892 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

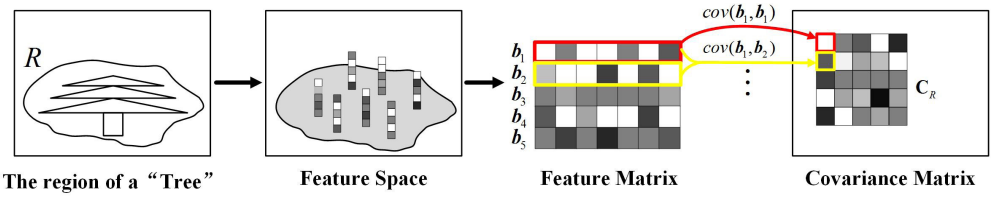


Fig. 1. Illustration of the construction of the CMR for a "Tree" region. $\boldsymbol{b}_1 \ldots \boldsymbol{b}_5$ are five row feature bands. *cov* denotes the covariance operator and $cov(\boldsymbol{x},\boldsymbol{y})(1/(n-1))\sum_{i=1}^{n} z_i - (1/(n-1)\sum_{i=1}^{n} x_i) \times (1/(n-1)\sum_{i=1}^{n} y_i)$ for two given vectors $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]$, $\boldsymbol{y} = [y_1, y_2, \ldots, y_n] \in \mathbb{R}^n$ and $z_i = x_i y_i$. The $\mathbf{C}_R$ is the extracted CM.

morphological attribute profiles (EMAPs). By using the extinction filter [22], a more effective and automatic extinction profile method has been proposed. Subsequently, a series of related works have been introduced in [23] and [24]. Meanwhile, Kang *et al.* [25], [26] use edge-preserving filtering (EPF) to combine the spatial and the spectral information in the HSI based on the assumption that there is usually a strong correlation among neighboring pixels. Li *et al.* [27] exploit the contextual information using conventional feature descriptors, i.e., local binary patterns (LBPs), to enhance classification performance. To further accelerate the speed of processing and reduce the storage space for the HSI classification, a series of parallel and distributed-based approaches are proposed in [28] and [29].

In addition, deep learning techniques have also been used for HSI classification [30]–[33] motivated by their success in the computer vision community. In [30], the pixels of the HSI are regarded as sequential data, and a recurrent neural network is used to exploit the discriminative information provided by the spectral domain for the HSI classification. In [31]–[33], convolutional neural networks (CNNs) are adopted to exploit both spatial and spectral information to enhance the HSI classification performance. Specifically, Zhao and Du [31] used a local discriminant embedding algorithm and a CNN model to combine the spatial as well as the spectral information in HSI data exploitation. In [32], a pixel-pair model is developed to increase the amount of training samples which can be used to train the CNN model more effectively. In [33], the CNN is used to extract high-level spatial–spectral features followed by a logistic regression technique for the HSI classification. However, it is worthwhile to note that, in order to build the EMP, EMAP, or LBP features, a dimensionality reduction method (e.g., the PCA and ICA) is first applied to the HSI. The resulting bands after such dimensionality reduction are processed in completely separate fashion by the corresponding filters (e.g., morphological) or feature descriptors (e.g., LBP). As a result, the spectral correlation (which can offer important discriminative information) is not fully exploited by these methods.

Recently, the covariance matrix representation (CMR) method has been successfully applied on a variety of computer vision tasks, including fields such as image segmentation [34], texture classification [35], image set classification [36], and face recognition [37]. With the CMR, the region of interest is characterized by a CM and each nondiagonal entry in the CM stands for the correlation between two different features, which introduces a natural way of fusing multiple features that might be correlated. Fig. 1 shows an illustration of the CMR concept for the characterization of a tree object in a color image. As can be observed, the correlation among multiple features can be fully characterized by the $\mathbf{C}_R$ extracted from the tree region, which is the main advantage of the CMR [38]. Note that the feature space in Fig. 1 can be built by an FE method (e.g., Gabor features) or simply using the features in the RGB color space.

In this paper, in order to fully exploit the spectral correlation among different bands in addition to the spatial correlation in the original HSI, a novel local CMR (LCMR) method is proposed inspired by the advantages of the CMR in other fields. Using the concept of CMR, the newly proposed method can simultaneously exploit the correlation among different spectral bands and the spatial information in the HSI, allowing for a more discriminative FE process. In our method, the MNF-based dimensionality reduction method is first applied to the original HSI to reduce the dimensionality and discard noise. Then, for each test pixel of the dimensionally reduced HSI, the cosine distance measurement is used to find its $K-1$ most similar neighboring pixels within the window. Next, using the CMR method, the CM is calculated among the set of K pixels including the test pixel and its neighbors. Each nondiagonal entry in the CM characterizes the correlation of two different spectral bands. Thus, the correlation among different spectral bands and the spatial information are fully integrated in the obtained CM. Finally, the extracted covariance matrices are used as spatial–spectral features and fed to an SVM with the Log-Euclidean-based kernel for label assignment, since the covariance matrices do not lie on the Euclidean space but on the manifold space (MS).

The remainder of this paper is organized as follows. In Section II, the MNF-based dimensionality reduction and CMR are briefly reviewed. Section III describes the proposed LCMR method for the HSI classification. Section IV provides experimental results using three real HSIs and presents an exhaustive comparison with the other state-of-the-art FE methods prior to classification. Section V concludes this paper with some remarks at plausible future research.

II. Related Works

A. Maximum Noise Fraction-Based Dimensionality Reduction

Given a data matrix $\mathbf{X} \in \mathbb{R}^{M\times N}$, where M is the number of variables and N is the number of the observations. The MNF aims to find a linear transformation matrix $\mathbf{W}$ to maximize the SNR of transformed data. Assuming $\mathbf{X} = \mathbf{S} + \mathbf{N}$, where $\mathbf{S}$ and $\mathbf{N}$ are the uncorrelated signal and noise matrix. Thus, $cov(\mathbf{X}) = \mathbf{\Sigma}_X = \mathbf{\Sigma}_S + \mathbf{\Sigma}_N$, where $\mathbf{\Sigma}_S$ and $\mathbf{\Sigma}_N$ are CM

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

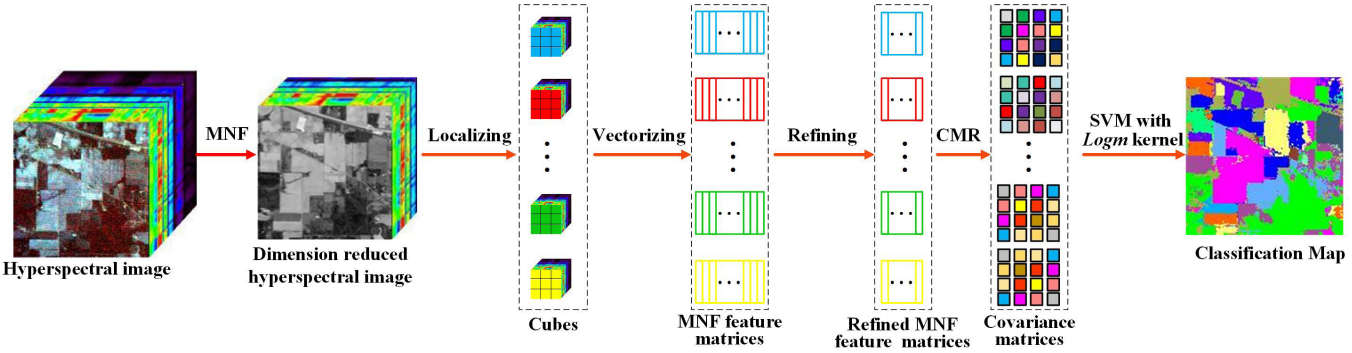


Fig. 2. Flowchart of the proposed LCMR classification method. Generally, the proposed LCMR method consists of three main steps, i.e., maximum noise fraction-based dimensionality reduction, KNN-based neighboring pixels refining, and LED kernel-based SVM for label assignment.

of $\mathbf{S}$ and $\mathbf{N}$, respectively. $\mathbf{W}$ can be obtained by solving the following problem:

$$\underset{\mathbf{W}}{\operatorname{argmax}} \frac{\mathbf{W}^T \boldsymbol{\Sigma}_S \mathbf{W}}{\mathbf{W}^T \boldsymbol{\Sigma}_N \mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} \frac{\mathbf{W}^T \boldsymbol{\Sigma}_X \mathbf{W}}{\mathbf{W}^T \boldsymbol{\Sigma}_N \mathbf{W}} - 1. \tag{1}$$

$\mathbf{W}$ is the eigenvectors associated with the L largest eigenvalues of $\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\Sigma}_X$. L is the number of MNF PCs, and the dimensionality reduced data $\mathbf{Y}$ is obtained as

$$\mathbf{Y} = \mathbf{W}^T \mathbf{X}. \tag{2}$$

Note that $\boldsymbol{\Sigma}_X$ is estimated using the sample CM of $\mathbf{X}$, and $\boldsymbol{\Sigma}_N$ is estimated by the minimum/maximum autocorrelation factors method [18].

B. Covariance Matrix Representation

The CMR was first introduced as an object descriptor by Tuzel *et al.* [38] and [39]. Let us assume that $\mathbf{I}$ denotes a color image. For a given region $\boldsymbol{R} \subseteq \mathbf{I}$ with N pixels, let $\boldsymbol{x}_i$ denote the d-dimensional feature (e.g., the color feature or the Gabor feature) extracted from the ith pixel within $\boldsymbol{R}$, and let $\boldsymbol{\mu}$ denote the mean vector of the set of feature vectors $\boldsymbol{x}_i, i = 1 \ldots N$. The CMR of region $\boldsymbol{R}$ can be then calculated as follows:

$$\mathbf{C}_R = \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T. \tag{3}$$

The diagonal entries of the CM represent the variance of each feature, and the nondiagonal entries represent the covariance among different features. Note that, to make the CM strictly positive definite, regularization is often applied to the original CM as: $\mathbf{C}_R^* = \mathbf{C}_R + \lambda \mathbf{E}$, where $\mathbf{E}$ is the identity matrix and λ is set to $10^{-3} \times trace(\mathbf{C}_R)$ [36], [40], [41]. In the proposed method, regularization is also adopted.

As a common symmetric positive definite (SPD) matrix, the covariance matrices do not lie on a Euclidean space but on a Riemannian manifold [36] space, and the common Euclidean-based distance metric is not suitable for them. In [42], a manifold-based distance metric, termed Log-Euclidean distance (LED) metric, is proposed to address this problem. Given two SPD matrices $\mathbf{C}_1$ and $\mathbf{C}_2$, the LED between the two SPD matrices is defined as follows:

$$d_{\text{LED}}(\mathbf{C}_1, \mathbf{C}_2) = \|\text{logm}(\mathbf{C}_1) - \text{logm}(\mathbf{C}_2)\|_F \tag{4}$$

where logm denotes the ordinary matrix logarithm operator and $\|\cdot\|_F$ denotes the Frobenius norm. Let $\mathbf{C} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T$ be the eigen-decomposition of the SPD matrix $\mathbf{C}$, then its logm operator can be computed by

$$\text{logm}(\mathbf{C}) = \mathbf{U}\log(\boldsymbol{\Sigma})\mathbf{U}^T. \tag{5}$$

C. Log-Euclidean-Based Kernel Function

Since the CM lies on an MS, it cannot serve as a direct input for classical learning algorithms, such as the SVM. Fortunately, the logm operator can project a point $\mathbf{C}$ on the Riemannian MS to the Euclidean space. In [36], by computing the inner product on the Euclidean space with the logm operator, the corresponding Log-Euclidean-based kernel function on the MS can also be implicitly derived as follows:

$$k_{\text{logm}}(\mathbf{C}_1, \mathbf{C}_2) = trace[\text{logm}(\mathbf{C}_1) \cdot \text{logm}(\mathbf{C}_2)]. \tag{6}$$

It is easy to notice that (6) satisfies the conditions of Mercer's theorem, which allows us to utilize any standard vector space learning algorithm [36].

III. LCMR Method for Classification of HSIs

Fig. 2 shows the flowchart of the proposed method for the classification of HSIs, which consists of the following three steps. First, for each pixel, the local neighboring pixels on the subspace (e.g., the MNF-based dimensionality reduced subspace) are obtained using the cosine distance. Then, the CMR is applied on each pixel and its local neighboring pixels on the subspace. Finally, the obtained covariance matrices are used as spatial–spectral features and fed into an SVM with Log-Euclidean-based kernel for label assignment.

A. Spectral Dimensionality Reduction

In order to reduce the computational complexity and discard noise, the MNF [18] is first applied to the original HSI. Specifically, given an HSI $\mathbf{Z} \in \mathbb{R}^{I \times J \times K}$, the dimensionally reduced HSI $\mathbf{F} \in \mathbb{R}^{I \times J \times L}$ can be obtained with (2), where I and J denote the size of two spatial dimensions, K is the size of the spectral dimension, and L represents the number of MNF components. Note that, $\mathbf{Z}$ is vectorized to a matrix along the spectral dimension at first and the inverse procedure of vectorization is conducted to obtain $\mathbf{F}$. Compared

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I

NORMALIZED DISTANCE AMONG FOUR REPRESENTATIVE PIXELS (BELONGING TO DIFFERENT CLASSES OF THE AVIRIS INDIAN PINES SCENE) SHOWN IN FIG. 3 ON BOTH THE OSS AND THE MS OBTAINED BY CMR

Normalized distance on the original spectral space: 0.0096					Normalized distance on CMR manifold space: 0.2742				
Location	(a)	(b)	(c)	(d)	Location	(a)	(b)	(c)	(d)
(a)	0	0.0013	0.0010	0.0009	(a)	0	0.0298	0.285	0.0322
(b)	0.0013	0	0.0003	0.0007	(b)	0.0298	0	0.0199	0.0117
(c)	0.0010	0.0003	0	0.0006	(c)	0.2850	0.0199	0	0.0150
(d)	0.0009	0.0007	0.0006	0	(d)	0.0322	0.0117	0.0150	0

with other transform-based dimensionality reduction methods (e.g., PCA [16] and ICA [17]), the main advantage of the MNF is that noise can be effectively removed on the MNF transformed space, since the MNF aims at maximizing the SNR.

B. Identification of Local Neighboring Pixels

Using a square window of fixed size is a common way to explore the spatial information [43], [44] for the HSI classification. However, it should be noted that there may still be some pixels with low spectral correlation within the window, especially around the object edges. To address this issue, we introduce a local neighboring pixels construction method to discard the dissimilar neighboring pixels in the reduced subspace.

First, a large window with size $T \times T$ is used to extract the neighbors with regard to a central pixel denoted by $\boldsymbol{f}_1^j$, where the neighboring pixels are denoted by $\{\boldsymbol{f}_i^j\}_{i=2,\ldots,T^2}$, $j = 1, \ldots, N$. N is the total number of pixels in the image. Then, the K-NN method with the cosine distance is applied on these pixels, i.e., the $\boldsymbol{f}_1^j$ and its $T^2 - 1$ neighboring pixels, i.e., $\{\boldsymbol{f}_i^j\}_{i=2,\ldots,T^2}$. The cosine distance between the central pixel and its neighboring pixels is calculated as follows:

$$\cos(\boldsymbol{f}_1^j, \boldsymbol{f}_i^j) = \frac{\langle \boldsymbol{f}_1^j, \boldsymbol{f}_i^j \rangle}{\|\boldsymbol{f}_1^j\|_2 \cdot \|\boldsymbol{f}_i^j\|_2} \tag{7}$$

where $\langle \cdot \rangle$ and $\| \cdot \|$ denote the inner product and the Frobenius norm, respectively. Only the $K - 1$ most similar pixels with regard to the central pixel within the window are taken into consideration, whereas the rest of them are discarded. In this way, we can construct a local neighborhood for each test pixel, i.e., $\{\boldsymbol{f}_k^j\}_{k=1,\ldots,K}$. Obviously, the selected $K - 1$ pixels are not only spatially close to the central pixel but also share relatively similar spectral information which can be used to extract more discriminative features in the construction of CMRs. At this point, it is important to emphasize that the cosine distance-based K-NN strategy is used in our method because of the following two reasons. First, it is quite common to use the cosine distance-based similarity measurement for HSI data processing [45], [46] and it has been reported to specifically provide a good performance in the HSI classification [45]. Second, although there are some shape adaptive-based methods which can automatically search for similar pixels that show a good performance on the HSI classification and denoising, they need careful parameter setting. In addition, these methods exhibit heavier computational burden [47], [48]. Overall, to make the proposed method simple yet effective, the cosine distance-based K-NN is adopted in our method.

C. Construction of CMRs for Each Pixel and Its Local Neighbors

After obtaining $\{\boldsymbol{f}_k^j\}_{k=1,\ldots,K}$, it is easy to construct the CMR for the central pixel $\boldsymbol{f}_1^j$ with (3), which can be rewritten as follows:

$$\mathbf{C}_j = \frac{1}{K-1} \sum_{k=1}^{K} (\boldsymbol{f}_k^j - \boldsymbol{\mu})(\boldsymbol{f}_k^j - \boldsymbol{\mu})^T \tag{8}$$

where $\boldsymbol{\mu}$ denotes the mean vector of the set of feature vectors $\{\boldsymbol{f}_k^j\}_{k=1,\ldots,K}$. By conducting this process, a set of covariance matrices are extracted for each of the pixels in the image, i.e., $\{\mathbf{C}_j\}_{j=1,\ldots,N}$. The covariance matrices sets are then used as spatial–spectral features and fed into an SVM with Log-Euclidean-based kernel for final classification.

The basic motivation behind this step is to fully exploit the correlation among different spectral bands using the CMR for more discriminative FE. Fig. 3 provides an example using four representative pixels from different classes on the original spectral space (OSS) and the MS obtained by the CMR. The well-known Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines scene is used for illustrative purposes. As can be seen from the first row in Fig. 3, the four pixels exhibit very similar spectral signatures, although they belong to different classes (i.e., grass-pasture-mowed, corn, soybean-mintill, and soybean-clean). As a result, it is very challenging to discriminate these pixels. The normalized distance among the four pixels both on the OSS and the MS is reported in Table I. As can be seen, the total normalized distance among the four pixels in the OSS is 0.0097, whereas the total distance among the four pixels in the MS is 0.2742. Obviously, the CMR can enlarge the distance among the four pixels (with very similar spectral signatures but actually belonging to different classes). In turn, let us now consider four pixels from the same class (i.e., Corn-mintill) both on the OSS and the MS, as shown in Fig. 4. As can be observed from the first row, the four representative pixels now show very different spectral signatures, although they actually belong to the same class. The normalized distance among the four pixels on the OSS and the MS is also reported in the third row of Table II. Here, we can observe that the total normalized distance among the four pixels is reduced from 0.1056 to 0.0226. Overall, this indicates that the CMR can not only enlarge the distance for pixels in different classes but also reduce the distance for pixels within the same class, which offers valuable discriminative information for the classification task. The normalized distance is calculated as follows:

$$d_{\text{norm}}(\boldsymbol{y}_1, \boldsymbol{y}_2) = \frac{0.5 \times var(\boldsymbol{y}_1 - \boldsymbol{y}_2)}{var(\boldsymbol{y}_1) + var(\boldsymbol{y}_2)} \tag{9}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

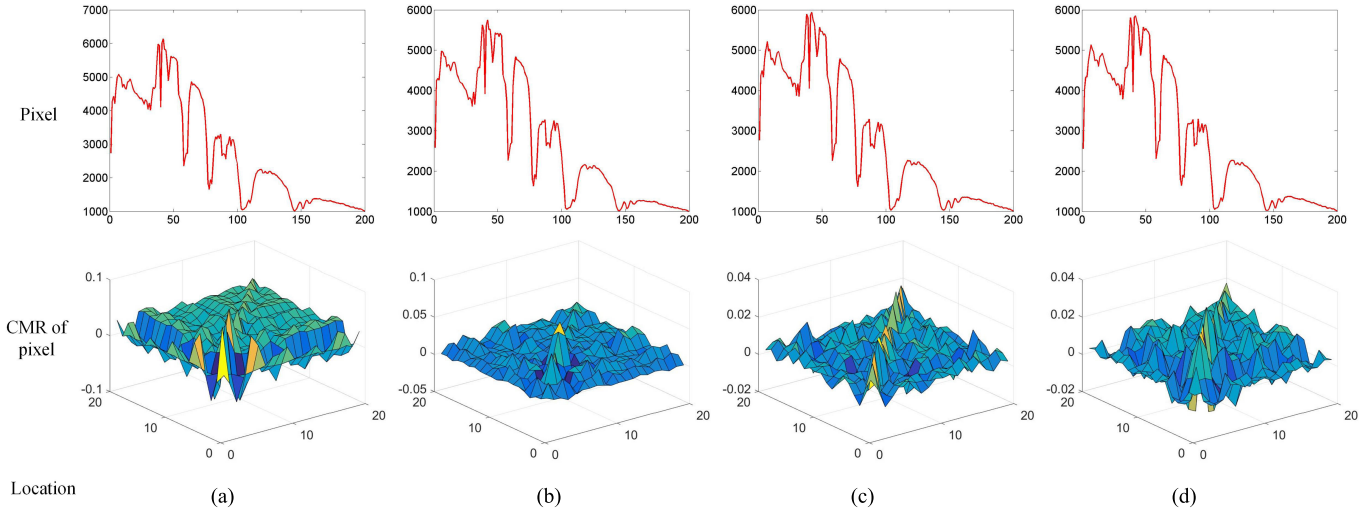


Fig. 3. Comparison between the OSS and the MS obtained by CMR for pixels belonging to different classes in the AVIRIS Indian Pines scene. The first row shows the spectral signatures of four representative pixels from four different classes in the OSS. The second row shows the corresponding four representative pixels on the MS obtained by the CMR. As can be observed, the four pixels, belonging to four different classes, share very similar signatures, which are hard to distinguish on the OSS. The normalized distance among the four pixels on both the OSS and the MS is reported in Table I. (a) 77,110. (b) 34,4. (c) 58,114. (d) 18,60.

TABLE II

NORMALIZED DISTANCE AMONG FOUR REPRESENTATIVE PIXELS (BELONGING TO THE SAME CLASS OF THE AVIRIS INDIAN PINES SCENE) SHOWN IN FIG. 4 ON BOTH THE OSS AND THE MS OBTAINED BY CMR

Normalized distance on the original spectral space: 0.1056					Normalized distance on CMR manifold space: 0.0226				
Location	(a)	(b)	(c)	(d)	Location	(a)	(b)	(c)	(d)
(a)	0	0.0150	0.0126	0.0111	(a)	0	0.0022	0.0011	0.0020
(b)	0.0150	0	0.0027	0.0063	(b)	0.0022	0	0.0026	0.0005
(c)	0.0126	0.0027	0	0.0051	(c)	0.0011	0.0026	0	0.0029
(d)	0.0111	0.0063	0.0051	0	(d)	0.0020	0.0005	0.0029	0

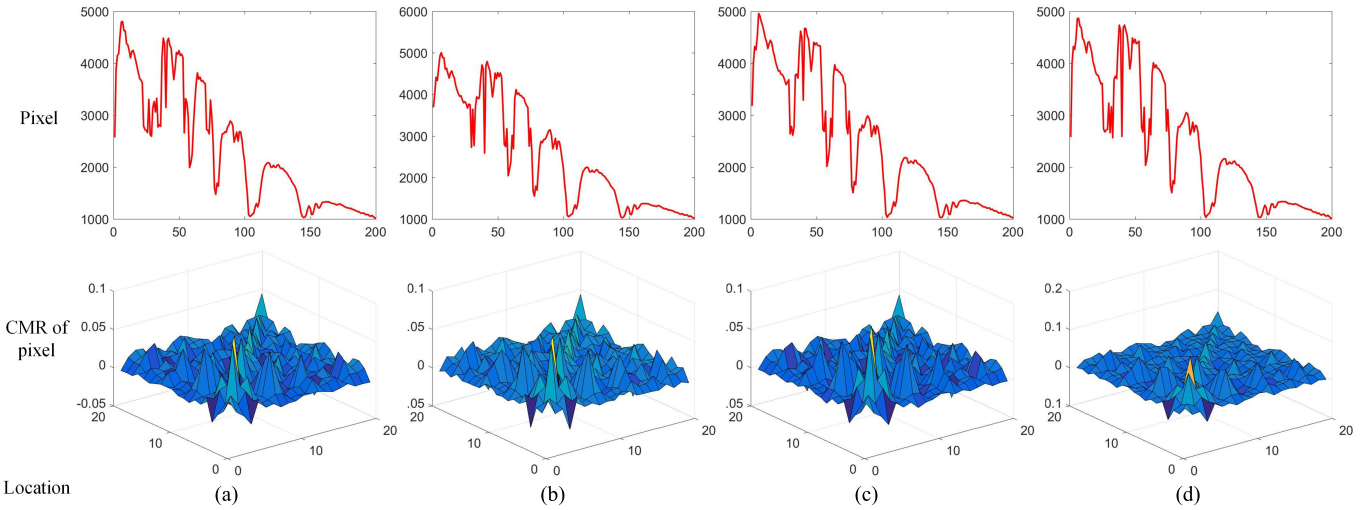


Fig. 4. Comparison between the OSS and the MS obtained by CMR for pixels belonging to the same class in the AVIRIS Indian Pines scene. The first row shows the spectral signatures of four representative pixels from the same class on the OSS. The second row shows the corresponding four representative pixels on the MS obtained by CMR. As can be observed, although the four pixels belong to the same class, they exhibit very different spectral curves. The normalized distance among the four pixels on both the OSS and the MS is reported in Table II. (a) 34,48. (b)36,57. (c)36,46. (d) 35,56.

where $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are two vectors on the Euclidean space. *var* denotes the variance operator and $var(\mathbf{y}) = (1/n)\sum_{i=1}^{n}(y_i - (1/n)\sum_{i=1}^{n} y_i)^2$ for a vector $\boldsymbol{y} = [y_1, y_2, \cdots, y_n]^T \in \mathbb{R}^n$. Note that, to calculate the normalized distance between two matrices with the same size, the two matrices are first reshaped to vectors.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Data Sets

In our experiments, three real hyperspectral data sets, i.e., the "Reflective Optics System Imaging Spectrometer (ROSIS-03) University of Pavia," "AVIRIS Indian Pines," and "AVIRIS Salinas" images are used.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
NUMBERS OF TRAINING AND TEST SAMPLES ON THE THREE TEST DATA SETS

Indian Pines				University of Pavia				Salinas			
Class	Name	Train	Test	Class	Name	Train	Test	Class	Name	Train	Test
1	Alfalfa	5	41	1	Asphalt	10	6621	1	Brocoli green weeds1	5	2004
2	Corn-notill	5	1423	2	Meadows	10	18639	2	Brocoli green weeds	5	3721
3	Corn-mintill	5	825	3	Gravel	10	2089	3	Fallow	5	1971
4	Corn	5	232	4	Trees	10	3054	4	Fallow rough plow	5	1386
5	Grass-pasture	5	478	5	Metal sheet	10	1335	5	Fallow smooth	5	2673
6	Grass-trees	5	725	6	Bare Soil	10	5019	6	Stubble	5	3954
7	Grass-pasture-mowed	5	24	7	Bitumen	10	1320	7	Celery	5	3574
8	Hay-windrowed	5	473	8	Bricks	10	3672	8	Grapes untrained	5	11266
9	Oats	5	15	9	Shadows	10	937	9	Soil vinyard develop	5	6198
10	Soybean-notill	5	967		Total	90	42686	10	Corn senesced green weeds	5	3273
11	Soybean-mintill	5	2450					11	Lettuce romaine 4wk	5	1063
12	Soybean-clean	5	588					12	Lettuce romaine 5wk	5	1922
13	Wheat	5	200					13	Lettuce romaine 6wk	5	911
14	Woods	5	1260					14	Lettuce romaine 7wk	5	1065
15	Buildings-Grass-Trees-Drives	5	381					15	Vinyard untrained	5	7263
16	Stone-Steel-Towers	5	88					16	Vinyard vertical trellis	5	1802
	Total	80	10169						Total	80	54049

1) University of Pavia: The University of Pavia image was acquired by the ROSIS-03 sensor over the campus at the University of Pavia, Italy. This data set contains 103 spectral bands after the noise-corrupted bands are discarded, and each band is of size 610×340. The spatial resolution of this data set is 1.3 m, and the spectral coverage ranges from 0.43 to 0.86 μm.

2) Indian Pines: The Indian Pines image covers the agricultural Indian Pines test site in Northwestern Indiana and was collected by the AVIRIS sensor. The data set is of size $145 \times 145 \times 220$, with a spatial resolution of 20 m and a spectral range from 0.2 to 2.4 μm. Before the classification, 20 spectral bands (i.e., 104th–108th, 150th–163rd, and 220th) are discarded due to low SNR. This image contains 16 classes.

3) Salinas: The Salinas image was also collected by the AVIRIS sensor over the Salinas Valley, CA, USA. The data set is of size $512 \times 217 \times 224$, and it has spatial resolution of 3.7 m/pixel. Before classification, 20 water absorption bands were removed (i.e., 108th–112th, 154th–167th, and 224th).

B. Experimental Setup

Here, several state-of-the-art HSI classification methods, i.e., the SVM on the original data set [12], the SVM on the dimensionality-reduced HSI obtained by the MNF method [18], the EPF-based method [26], the multiple feature learning (MFL)-based method [49], the LBP-based method [27], the weighted Markov random field (WMRF)-based method [50], the SVM on EMAP method [21], and the superpixel-based classification via multiple kernels (SC-MK) [6], are also used for comparison purposes. The SVM method has been implemented using the LIBSVM library [51] and a Gaussian kernel with fivefold cross validation. For the MNF, the number of MNF PCs is set to be the same with the LCMR. For the EMAP method, the attributes are extracted using threshold values in the range from 2.5% to 10% with respect to the mean of the individual features, with a step of 2.5% for the standard deviation attribute and thresholds of 200, 500, and 1000 for the area attribute. For the EPF, LBP, SC-MK, and WMRF, the parameters are set using the default values in [6], [26], [27], and [50]. The code of the proposed LCMR method is available online (https://sites.google.com/site/leyuanfang/home).

To demonstrate the performance of our proposed method, we consider very small training sets in our experiments. For the University of Pavia data set, only ten training samples are randomly selected per class, and the remaining samples are then used for evaluation purposes. For the Indian Pines and Salinas data sets, only five training samples are randomly selected per class, and the remaining samples are then used for evaluation. The detailed numbers of training and test samples used are given in Table III. All experiments are repeated ten times, and the average accuracies (AAs) and standard deviations are reported. Moreover, three widely used quantitative metrics, i.e., overall accuracy (OA), AA, and Kappa coefficient are used to evaluate the classification performance.

C. Parameter Analysis

First, the effect of the number of MNF components (i.e., L) and the number of local neighboring pixels (i.e., K) are investigated. In our experiments, K ranges from 100 to 500 with step 20 and L ranges from 10 to 50 with step 10. The window $(T \times T)$ is fixed with the size of 25×25. As can observed from Fig. 5, with the increase of L, there is a slight improvement of the OA on the three data sets. However, it should also be noted that the computing time increase rapidly as well. This is due to the fact that the computational complexity of the LCMR is mainly determined by the size of the extracted covariance matrices. The larger the L value, the larger the size of the extracted covariance matrices and this increases the computational cost. Therefore, L is set to 20 and remains unchanged in the following experiments. For parameter K, it is used to determine the number of local neighboring pixels. In general, a small K implies that less neighboring pixels are used in the procedure of the CMR calculation and vice versa. From Fig. 5, it is easy to observe that the OA is increased when the K grows from 100 to 320 on both the University of Pavia data set and the Salinas data set. A further increase of K will decrease the OA. This is because larger K value

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

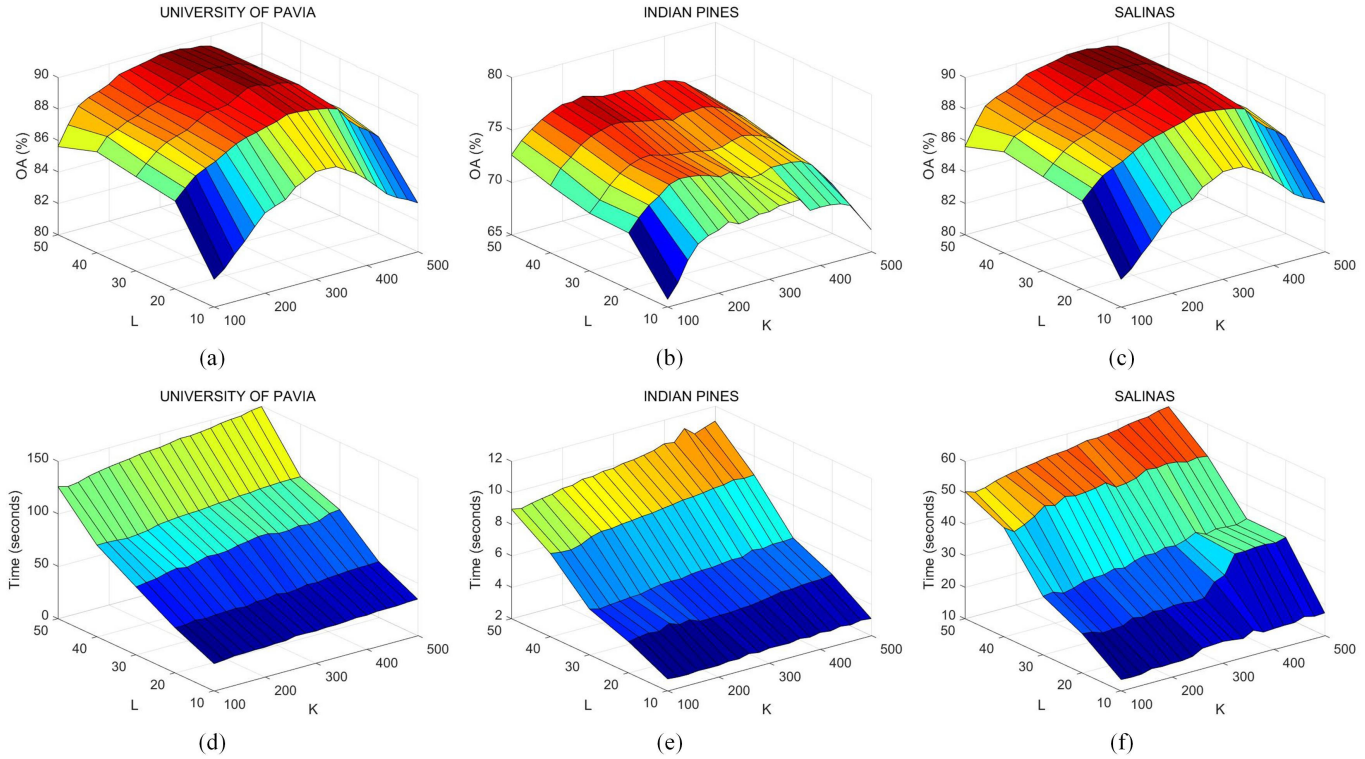


Fig. 5. Sensitivity analysis of parameters L and K in the proposed LCMR method. As can be seen, although the OAs obtained by the proposed method on the three test data sets can improve slightly with the growing of L (i.e., from 10 to 50), the corresponding computation times increase rapidly as well. As a result, the parameter L is set to 20. On the other hand, when K ranges from 200 to 220, relative higher classification accuracies can be obtained on the three data sets. Therefore, K is set to 220 for the remaining experiments.

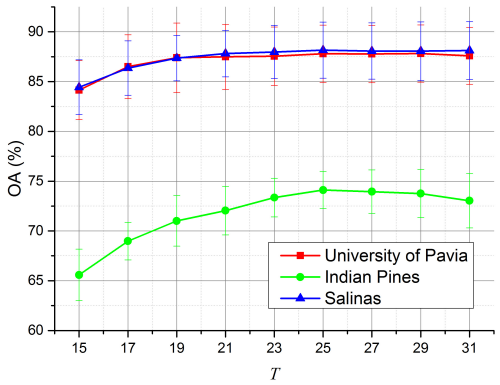


Fig. 6. Analysis of the sensitivity of parameter T in the proposed LCMR method. As can be seen, when T grows from 15 to 25, the classification performance improve stably on the three data sets and when T is larger than 25, the OAs will decrease. Therefore, T is set to 25.

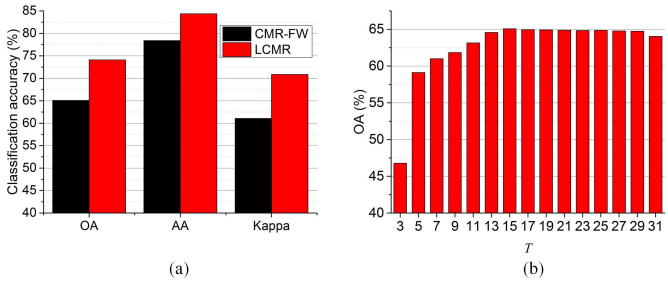


Fig. 7. Impact of our local neighboring pixels construction strategy. (a) Classification accuracy comparison between the LCMR and the CMR-FW. (b) OAs of CMR-FW with various T values, and the best T value for the CMR-FW is 15.

lead to the incorporation of relatively dissimilar pixels with regard to the central pixel in the CMR calculation. A similar situation can also be observed for the Indian Pines data set but with a different optimal value of K. In this case, the optimal K for the Indian Pines scene is 220. This is reasonable, since the Indian Pines image has a smaller size and a larger K will include relatively more dissimilar pixels. When the K ranges from 200 to 220, high classification accuracies can be obtained on the three data sets. As a result, K is set to 220 for the remaining experiments.

In addition, the sensitivity of parameter T is also investigated with parameters L and K fixed to 20 and 220, respectively. T is tuned from 15 to 31 with a step of 2. As can be seen from Fig. 6, the improvement of OA on the three test images is obvious when T grows from 15 to 25. When T is larger than 25, the OA on both the University of Pavia and Salinas data sets remain stable whereas there is a slight decrease in the case of the Indian Pines data set. This is mainly due to the fact that the Indian Pines data set has a smaller size, and a large window size contains relatively more dissimilar pixels, thus degrading the classification performance. As a result, T is set to be 25. Table IV summarizes the parameter setup for the proposed LCMR method in our experiments.

TABLE IV
PARAMETERS SETUP FOR THE PROPOSED LCMR METHOD

Parameter	Explanation	Value
L	Number of MNF principal components	20
K	Number of local neighboring pixels	220
T	Window size ($T \times T$)	25

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

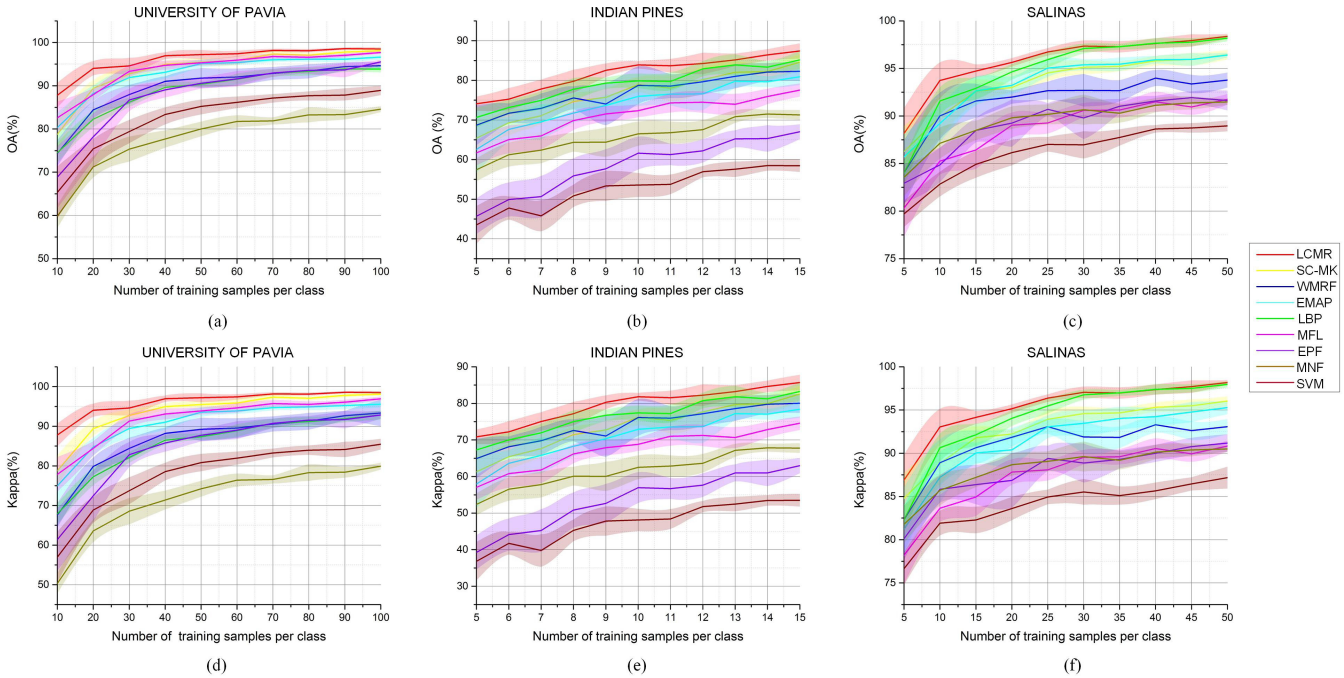


Fig. 8. Effect of using different numbers of training samples in the classification of different data sets using SVM [12], MNF [18], EPF [26], MFL [49], LBP [27], EMAP [21], WMRF [50], SC-MK [6], and the proposed LCMR method.

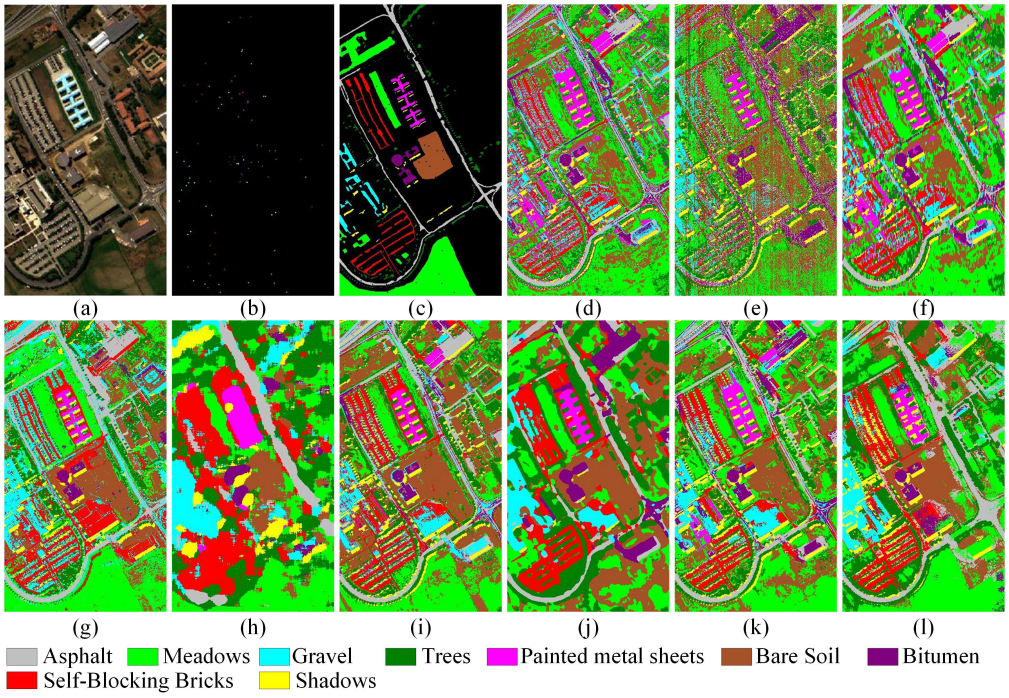


Fig. 9. University of Pavia data set. (a) Three-band color composite image. (b) Training set (with ten labeled samples per class). (c) Test set. Classification maps obtained by (d) SVM [12], (e) MNF [18], (f) EPF [26], (g) MFL [49], (h) LBP [27], (i) EMAP [21], (j) WMRF [50], (k) SC-MK [6], and (l) proposed LCMR.

D. Contribution of Local Neighboring Pixels Identification

To demonstrate the necessity of a local neighboring pixel identification strategy, a simplified version of LCMR, the CMR with fixed size window (CMR-FW) is used for comparison. As can be observed from Fig. 7, by using our local neighboring pixel construction strategy to refine the set of neighboring pixels, there is an obvious improvement in terms of the classification performance with regard to the CMR-FW. For example, the OA is improved from 65% to almost 75%, and the Kappa is improved from about 62.5% to 70%. Overall, the local neighboring pixel identification strategy can discard relatively dissimilar pixels and thus enhance the classification performance. Note that, to make a fair comparison, L for the CMR-FW is set to 20, which is the same value used with the LCMR. In addition, the optimal T value for the CMR-FW is obtained experimentally, as shown in Fig. 7(b). The optimal T value for CMR-FW is 15. This experiment is again performed on the AVIRIS Indian Pines data set.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V

AA (%) AND STANDARD DEVIATION (IN THE PARENTHESES) OF TEN REPEATED EXPERIMENTS IN THE UNIVERSITY OF PAVIA IMAGE OBTAINED BY DIFFERENT METHODS WITH TEN TRAINING SAMPLES PER CLASS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Class	SVM [12]	MNF [18]	EPF [26]	MFL [49]	LBP [27]	EMAP [21]	WMRF [50]	SC-MK [6]	LCMR
1	61.11(6.20)	44.63(8.33)	72.23(7.87)	**81.83**(5.00)	60.24(5.45)	77.88(9.83)	68.75(11.12)	73.32(6.72)	79.32(7.50)
2	59.43(12.96)	59.81(8.28)	60.09(13.28)	81.00(7.49)	74.30(5.10)	76.70(7.51)	71.18(12.39)	72.62(7.85)	**85.61**(7.12)
3	52.97(17.07)	67.36(3.71)	57.95(19.21)	81.18(6.07)	81.22(11.20)	75.11(11.71)	84.83(8.42)	79.60(13.32)	**92.41**(5.35)
4	90.00(7.85)	82.75(4.25)	89.37(6.78)	77.54(8.22)	61.67(8.07)	94.30(3.16)	91.22(4.40)	93.87(4.41)	**95.64**(5.41)
5	99.17(0.30)	99.57(0.31)	99.61(0.11)	98.49(0.47)	91.29(5.39)	99.63(0.06)	**99.93**(0.12)	99.69(0.16)	98.13(5.50)
6	53.65(12.80)	58.67(8.65)	60.18(11.96)	85.34(8.03)	86.45(9.29)	79.58(11.75)	63.80(20.37)	84.43(6.54)	**95.03**(3.99)
7	88.80(4.33)	71.18(8.07)	93.19(3.25)	97.06(2.14)	88.67(4.48)	91.72(3.90)	**99.27**(0.37)	86.52(7.52)	97.48(1.03)
8	75.54(6.11)	37.53(6.69)	80.58(6.04)	79.11(6.77)	83.61(7.47)	79.50(7.15)	85.06(7.00)	83.10(6.72)	**86.12**(6.41)
9	99.75(0.27)	95.80(2.24)	99.65(0.27)	94.77(6.65)	62.63(6.39)	99.93(0.10)	41.75(20.07)	**99.97**(0.07)	94.70(3.15)
OA	65.30(6.05)	59.80(2.68)	68.86(6.87)	82.58(3.74)	74.50(2.39)	80.33(3.47)	74.36(3.02)	78.76(3.86)	**87.79**(2.88)
AA	75.60(3.12)	68.59(1.01)	79.20(3.90)	86.26(1.75	76.67(1.32)	86.04(1.53)	78.42(2.82)	85.90(1.70)	**91.60**(0.95)
Kappa	56.96(6.41)	50.34(2.58)	61.39(7.56)	77.85(4.37)	67.88(2.67)	74.97(4.13)	67.56(3.12)	73.25(4.47)	**84.33**(3.48)

TABLE VI

AA (%) AND STANDARD DEVIATION (IN THE PARENTHESES) OF TEN REPEATED EXPERIMENTS ON THE INDIAN PINES IMAGE OBTAINED BY DIFFERENT METHODS WITH FIVE TRAINING SAMPLES PER CLASS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Class	SVM [12]	MNF [18]	EPF [26]	MFL [49]	LBP [27]	EMAP [21]	WMRF [50]	SC-MK [6]	LCMR
1	81.95(7.44)	88.05(4.94)	84.39(28.68)	93.90(5.12)	**99.51**(1.40)	96.10(3.31)	97.56(1.89)	97.56(0.01)	**99.51**(0.93)
2	30.67(10.31)	47.66(7.94)	31.05(11.46)	46.49(9.45)	52.04(7.34)	46.62(8.33)	54.81(11.47)	56.12(9.47)	**65.51**(6.46)
3	33.92(13.46)	41.53(14.88)	36.36(15.89)	48.55(14.94)	50.3(10.65)	56.40(14.16)	**61.80**(17.75)	53.58(11.50)	60.82(13.11)
4	31.55(11.17)	42.63(16.64)	36.51(16.33)	39.09(12.97)	**90.73**(9.44)	49.74(15.19)	71.77(25.58)	72.03(18.74)	74.44(17.16)
5	44.44(20.58)	81.86(12.27)	55.10(20.14)	61.97(14.73)	77.87(7.32)	65.46(24.26)	68.91(14.39)	70.04(13.56)	**90.54**(8.57)
6	61.74(11.26)	87.68(6.87)	62.92(11.09)	73.20(15.50)	79.78(8.38)	82.55(14.04)	**94.14**(8.02)	78.29(14.02)	88.62(7.07)
7	86.09(6.63)	93.48(4.23)	89.57(29.88)	96.09(2.90)	**100**(0)	96.09(2.34)	97.83(2.92)	99.13(1.74)	**100**(0)
8	38.86(13.71)	80.04(13.53)	43.36(20.57)	92.77(5.26)	98.82(3.13)	93.85(7.87)	99.60(0.83)	97.17(5.60)	**100**(0)
9	90.00(12.47)	99.33(2.11)	78.67(27.29)	96.67(5.12)	98.67(3.81)	99.33(2.00)	95.33(6.00)	100(0)	**100**(0)
10	33.55(16.68)	52.31(12.21)	34.98(19.03)	54.24(11.74)	**73.80**(8.54)	57.40(15.56)	66.62(8.34)	57.98(10.12)	67.91(13.40)
11	39.51(7.31)	37.20(14.43)	41.55(13.02)	59.52(10.28)	**61.72**(4.74)	51.23(14.58)	54.87(18.63)	51.09(13.14)	60.44(5.82)
12	31.97(5.65)	55.68(11.98)	32.60(8.92)	62.43(10.67)	61.96(8.17)	51.41(10.68)	60.00(16.47)	66.00(11.99)	**67.43**(10.28)
13	93.65(3.36)	98.85(0.58)	99.15(0.32)	95.80(2.18)	97.45(3.54)	97.45(1.98)	**99.95**(0.15)	99.90(0.20)	**99.95**(0.14)
14	71.67(10.77)	80.08(10.78)	72.75(13.79)	75.77(6.37)	89.97(3.90)	82.29(8.98)	**92.60**(6.85)	83.20(8.69)	91.56(4.10)
15	22.57(9.62)	56.48(11.67)	25.41(12.27)	55.88(14.72)	83.78(21.69)	66.77(12.04)	46.40(21.38)	71.50(21.60)	**86.51**(20.28)
16	77.73(15.29)	85.57(6.16)	77.95(17.77)	82.27(6.50)	97.50(3.42)	**98.41**(1.54)	93.41(16.14)	97.73(0.01)	97.16(2.34)
OA	43.52(4.76)	57.42(3.07)	45.69(4.49)	61.71(3.18)	70.67(2.53)	62.56(5.96)	68.67(5.07)	65.44(2.97)	**74.11**(1.78)
AA	54.37(4.50)	70.53(2.86)	56.40(3.21)	70.92(2.35)	82.12(2.22)	74.45(4.81)	78.48(2.44)	78.21(2.21)	**84.40**(1.93)
Kappa	36.84(5.26)	52.40(3.15)	39.27(4.66)	57.04(3.42)	67.28(2.80)	57.97(6.63)	64.92(5.45)	61.30(3.22)	**70.86**(2.01)

E. Comparison With Different Methods

The first experiment is conducted on the University of Pavia data set and ten training samples are randomly selected per class. The corresponding classification results, i.e., OA, AA, Kappa, and classification accuracy of each class obtained by different test methods are reported in Table V. As can be observed from Table V, the proposed method delivers the best results in terms of OA, AA, and Kappa. Moreover, Fig. 9 shows a false color composite image, training set, the test set, and the final classification maps for the nine considered methods in a single experiment. As can be observed, since the SVM and MNF only take the spectral information into account, the classification map is quite noisy, especially on large homogenous regions (e.g., the Bare Soil class which lies on the middle of the image). By optimizing the probability map obtained from the SVM with an edge-preserving filter, the EPF method can yield more smooth classification maps. By taking advantage of the complementary information provided by multiple features, the MFL can obtain a more smooth classification map when compared with the SVM, MNF and EPF, but there is still some noticeable noise on the bottom of the map. In addition, the LBP and the WMRF lead to oversmoothed classification maps. For the LBP method, the classical LBP descriptor with a slack variable is used to extract the spatial information separately on each of the PCs. This step could greatly decrease the impact of noise. However, when the number of training samples is limited, this step may result in oversmoothed result, since the correlation among different spectral bands is not taken into account. For the WMRF, a spatially adaptive total variation regularization is used to enforce a more spatially smooth classification, which is helpful while it may also yield an oversmoothed classification map when training samples are limited. In contrast, the proposed LCMR can not only preserve the edges and boundaries in accordance with the false color composite image but also deliver smoother results. The reason is twofold. First, by using the local neighboring pixel construction strategy, the spatial information is accurately taken into consideration. Second, instead of separately processing the spectral bands of the HSI, the CMR can fully exploit the correlation among different spectral bands, thus allowing for the extraction of more discriminative features.

We have also investigated the influence of the number of training samples on the performance of the different tested methods. Fig. 8(a) and (d), respectively, shows the OA and Kappa coefficient achieved by the different tested methods as a function of the number of training samples (ranging from 10 to 100, with a step of 10 per class). As shown in Fig. 8,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VII

AA (%) AND STANDARD DEVIATION (IN THE PARENTHESES) OF TEN REPEATED EXPERIMENTS ON THE SALINAS IMAGE OBTAINED BY DIFFERENT METHODS WITH FIVE TRAINING SAMPLES PER CLASS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Class	SVM [12]	MNF [18]	EPF [26]	MFL [49]	LBP [27]	EMAP [21]	WMRF [50]	SC-MK [6]	LCMR
1	97.93(1.88)	96.13(3.60)	**99.40**(0.99)	95.92(4.39)	98.20(1.33)	97.25(7.48)	99.15(1.07)	98.00(2.50)	96.08(4.22)
2	95.26(5.06)	98.41(1.03)	**99.04**(1.27)	98.86(0.29)	91.14(4.46)	98.51(0.98)	99.12(2.52)	99.13(0.98)	97.20(3.75)
3	73.97(12.26)	97.29(2.78)	78.20(13.56)	**94.16**(10.69)	90.86(9.31)	88.40(15.16)	81.18(14.16)	89.04(11.08)	93.42(9.02)
4	96.60(1.98)	91.70(5.82)	97.73(2.38)	94.95(4.32)	97.07(1.90)	98.93(0.65)	99.37(0.14)	95.65(3.20)	**99.65**(0.87)
5	94.55(3.79)	97.79(0.84)	96.33(3.50)	**98.71**(0.22)	85.81(6.35)	96.30(2.35)	96.61(4.62)	94.49(4.86)	96.28(1.22)
6	95.39(5.56)	99.03(0.85)	95.91(4.79)	85.23(9.77)	86.16(9.46)	96.75(1.78)	**100**(0)	99.88(0.04)	99.21(0.40)
7	98.89(0.54)	98.82(0.46)	99.38(0.32)	88.23(8.36)	87.89(5.21)	99.23(0.28)	**99.94**(0.06)	96.58(4.30)	98.42(0.69)
8	51.49(15.63)	53.31(12.04)	54.03(12.89)	59.19(12.46)	**75.44**(7.62)	50.98(16.83)	50.33(18.21)	62.90(14.11)	58.72(13.95)
9	97.39(1.80)	97.48(2.45)	98.30(1.01)	97.92(0.52)	83.98(5.03)	96.83(0.97)	98.25(4.89)	99.00(0.63)	**99.30**(0.21)
10	78.33(9.33)	88.87(9.61)	87.62(8.66)	86.24(8.60)	81.69(8.86)	87.33(7.24)	87.08(8.41)	75.80(12.87)	**93.70**(7.92)
11	85.34(4.32)	94.34(3.34)	90.45(3.85)	92.80(2.67)	93.81(5.21)	94.58(1.43)	**98.19**(1.73)	92.85(6.97)	93.10(5.42)
12	97.91(1.85)	92.57(7.60)	**99.97**(0.07)	99.82(0.11)	79.43(12.25)	**99.97**(0.06)	99.69(0.85)	97.65(3.40)	97.77(1.88)
13	97.74(0.95)	93.10(7.18)	98.18(0.74)	97.28(3.87)	79.67(9.18)	**98.38**(0.74)	98.72(1.45)	98.13(0.01)	97.33(0.58)
14	88.30(3.26)	86.17(6.87)	93.09(3.88)	82.82(7.46)	79.61(9.69)	93.26(2.42)	99.36(0.51)	97.01(0.97)	**97.18**(0.78)
15	54.85(19.32)	70.76(9.43)	61.60(17.71)	55.67(11.15)	81.92(9.40)	70.69(20.35)	75.49(22.93)	80.13(8.09)	**91.16**(5.91)
16	83.94(9.46)	**92.85**(9.29)	86.87(9.96)	66.75(15.11)	90.55(6.74)	91.95(5.51)	90.92(10.23)	91.91(5.27)	89.63(8.46)
OA	78.92(1.79)	83.55(2.00)	82.06(2.23)	80.33(3.05)	83.99(1.32)	83.05(2.54)	84.02(3.42)	86.28(2.69)	**88.15**(2.68)
AA	86.74(1.64)	90.54(1.32)	89.76(1.82)	87.16(2.81)	86.45(1.21)	91.21(2.02)	92.09(2.29)	91.76(1.60)	**93.63**(1.41)
Kappa	76.65(1.91)	81.79(2.17)	80.11(2.47)	78.21(3.35)	82.28(1.46)	81.24(2.81)	82.32(3.79)	84.80(2.95)	**86.90**(2.94)

TABLE VIII

COMPARISON OF COMPUTING TIME (SECONDS) WITH THE SAME NUMBER OF TRAINING SAMPLES IN TABLES V–VII

Data sets	SVM [12]	MNF [18]	EPF [26]	MFL [49]	LBP [27]	EMAP [21]	WMRF [50]	SC-MK [6]	LCMR
University of Pavia	3.61	1.76	7.46	13.97	799.56	11.68	66.86	3.34	36.58
Indian Pines	1.94	0.82	3.34	1.64	81.15	1.81	11.43	0.64	3.89
Salinas	4.68	1.43	8.39	6.26	425.61	6.82	41.16	3.04	19.23

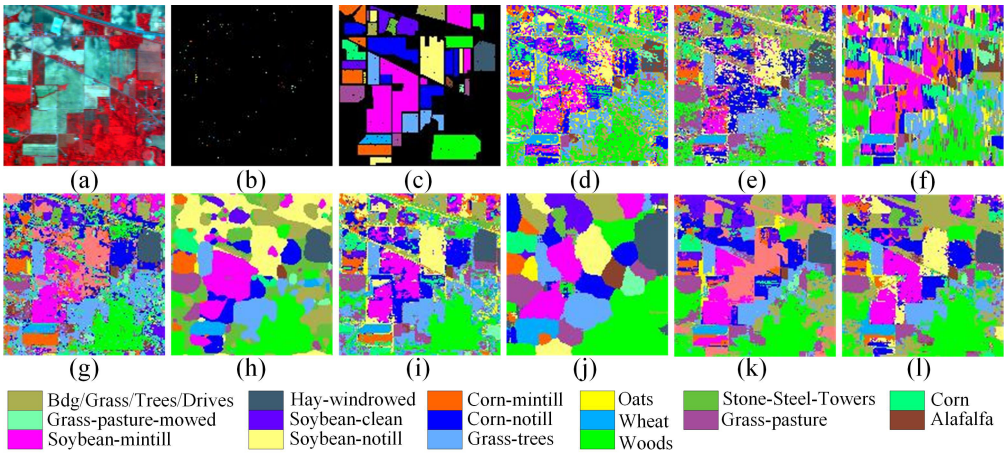


Fig. 10. Indian Pines data set. (a) Three-band color composite image. (b) Training set (with five labeled samples per class). (c) Test set. Classification maps obtained by (d) SVM [12], (e) MNF [18], (f) EPF [26], (g) MFL [49], (h) LBP [27], (i) EMAP [21], (j) WMRF [50], (k) SC-MK [6], and (l) proposed LCMR.

with the increase of the training set size, there is a stable improvement in classification performance for all considered classification methods. Most importantly, the proposed LCMR method consistently provides the best results, which suggests the effectiveness of the proposed approach (especially, when the training set size is very small).

The second experiment is performed on the Indian Pines data set. Fig. 8(b) and (e), respectively, illustrates the OA and Kappa coefficient using various numbers of training samples (ranging from 5 to 15, with a step of 1 per class). As can also be observed, the classification accuracies improve as the number of training samples increases, and the classification performance of the proposed LCMR method is the best among the considered methods. In addition, Table VI reports the detailed classification results when the number of training samples is 5 per class. As can be seen, the proposed method has the highest OA, AA, and Kappa coefficient. The corresponding classification maps are shown in Fig. 10.

The third experiment is performed on the Salinas data set. Fig. 8(c) and (f), respectively, illustrates the OA and Kappa coefficient using various numbers of training samples (ranging from 5 to 50, with a step of 5 per class). Moreover, the detailed classification results with five training samples selected per class are reported in Table VII. Furthermore, Fig. 11 shows the classification maps obtained by different methods in a single experiment. As can also be seen from Fig. 11, the proposed LCMR method can deliver competitive classification results in terms of accuracy.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

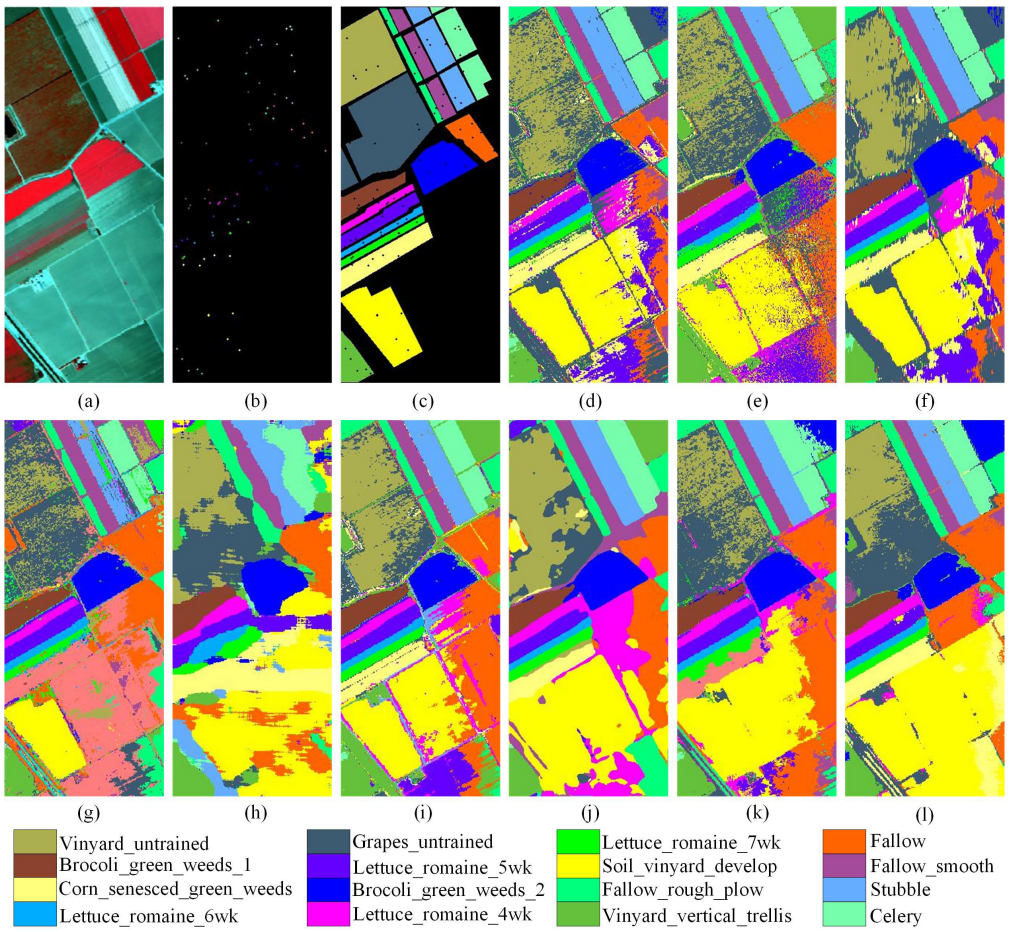


Fig. 11. Salinas data set. (a) Three-band color composite image. (b) Training set (with five labeled samples per class). (c) Test set. Classification maps obtained by (d) SVM [12], (e) MNF [18], (f) EPF [26], (g) MFL [49], (h) LBP [27], (i) EMAP [21], (j) WMRF [50], (k) SC-MK [6], and (l) proposed LCMR.

Finally, the computing time of the nine considered methods when applied to the three tested HSIs is reported in Table VIII. As can be observed, the LBP method is the slowest one, since it needs to extract a very high dimensional feature for each single pixel. In contrast, the proposed method is much faster. Moreover, the LCMR method is faster than the WMRF on all data sets. All the experiments are conducted using MATLAB 8.0 with a 2.6-GHz CPU (four cores) and 16 GB of RAM.

V. Conclusion and Future Work

In this paper, we have developed a novel LCMR FE method for spatial–spectral FE prior to the HSI classification. The proposed method consists of two main stages, i.e., local neighboring pixel construction and the exploration of correlation among different bands. With the first stage, the rich spatial information present in the HSI is exploited. With the second stage, the spectral correlation among different bands is also fully exploited, providing a competitive advantage with regard to other FE methods. Our experiments with three real HSIs demonstrate that the proposed method can outperform other state-of-the-art methods in terms of qualitative and quantitative performance, especially when the number of training samples is very small.

In our future work, we plan to combine superpixel-based segmentation methods with the CMR to exploit the spatial information in a more efficient way. Specifically, the following two directions will be considered. First, the HSI classification at the superpixel level (rather than at the pixel level) will be performed using CMRs. Superpixel segmentation methods can partition the HSI into many homogeneous regions based on the spatial-contextual information. However, the superpixels have arbitrary sizes and it might be hard to process them in a uniform way, whereas the CMR can adapt them into structures with the same size, which makes their processing more manageable. Second, superpixel methods can also provide a postprocessing framework to optimize the classification map obtained by the LCMR.

Acknowledgment

The authors would like to thank Prof. D. Landgrebe from Purdue University and the NASA Jet Propulsion Laboratory for providing the hyperspectral data sets. They would also like to thank Dr. J. Li, Dr. X. Kang, Dr. W. Li, and Dr. L. Sun for providing the multiple feature learning, edge-preserving filtering, local binary pattern, and weighted Markov random field implementations used in this paper for comparative purposes. Last but not least, they would also like to thank the editors and the anonymous reviewers for their detailed comments and suggestions, which greatly helped them to improve the technical quality and presentation of this paper.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

REFERENCES

[1] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.

[2] Y. Xu, Z. Wu, J. Chanussot, M. D. Mura, A. L. Bertozzi, and Z. Wei, "Low-rank decomposition and total variation regularization of hyperspectral video sequences," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1680–1694, Mar. 2018.

[3] L. Fang, H. Zhuo, and S. Li, "Super-resolution of hyperspectral image via superpixel-based sparse representation," *Neurocomputing*, vol. 273, no. 17, pp. 171–177, Jan. 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] C. Wu, L. Zhang, and B. Du, "Kernel slow feature analysis for scene change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2367–2384, Apr. 2017.

[6] 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.

[7] L. Y. Fang, S. T. Li, X. D. 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.

[8] Y. Liu, J. Li, P. Du, A. Plaza, X. Jia, and X. Zhang, "Class-oriented spectral partitioning for remotely sensed hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 691–711, Feb. 2017.

[9] Y. Guan, S. Guo, Y. Xue, J. Liu, and X. Zhang, "Application of airborne hyperspectral data for precise agriculture," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Sep. 2004, pp. 4195–4198.

[10] J. Rosentreter, R. Hagensieker, A. Okujeni, R. Roscher, P. D. Wagner, and B. Waske, "Subpixel mapping of urban areas using EnMap data and multioutput support vector regression," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1938–1948, May 2017.

[11] H. C. Chang and A. Burke, "Integration of hyperspectral and polarimetric radar remote sensing techniques for monitoring invasive weeds," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2014, pp. 2950–2952.

[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] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2271–2282, May 2010.

[14] 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.

[15] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017.

[16] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 625–629, Oct. 2008.

[17] 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.

[18] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-26, no. 1, pp. 65–74, Jan. 1988.

[19] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[20] 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.

[21] M. Dalla 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.

[22] 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.

[23] Y. Gu, T. Liu, X. Jia, and J. A. Benediktsson, "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.

[24] L. Fang, N. He, S. Li, P. Ghamisi, and J. A. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1803–1815, Mar. 2018.

[25] 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.

[26] 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.

[27] 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.

[28] Z. Wu *et al.*, "GPU parallel implementation of spatially adaptive hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published, doi: 10.1109/JSTARS.2017.2755639.

[29] Z. Wu, Y. Li, A. Plaza, J. Li, F. Xiao, and Z. Wei, "Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2270–2278, Jun. 2016.

[30] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[31] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[32] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.

[33] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[34] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu, "Free-form region description with second-order pooling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1177–1189, Jun. 2015.

[35] P. Li, Q. Wang, W. Zuo, and L. Zhang, "Log-Euclidean kernels for sparse representation and dictionary learning," in *Proc. ICCV*, Dec. 2013, pp. 1601–1608.

[36] R. Wang, H. Guo, L. S. Davis, and Q. Dai, "Covariance discriminative learning: A natural and efficient approach to image set classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2496–2503.

[37] Y. Pang, Y. Yuan, and X. Li, "Gabor-based region covariance matrices for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 989–993, Jul. 2008.

[38] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.

[39] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. Eur. Conf. Comput. Vis.*, Jun. 2006, pp. 589–600.

[40] Z. Huang, R. Wang, S. Shan, L. Van Gool, and X. Chen, "Cross Euclidean-to-Riemannian metric learning with application to face recognition from video," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2017.2776154.

[41] W. Wang, R. Wang, Z. Huang, S. Shan, and X. Chen, "Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 151–163, Jan. 2018.

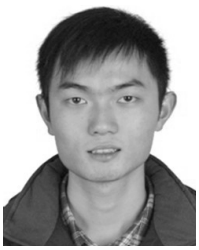
[42] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 29, no. 1, pp. 328–347, 2006.

[43] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[44] 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.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[45] 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.
[46] C.-I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.
[47] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 556–567, Feb. 2016.
[48] T. Lu, S. Li, L. Fang, Y. Ma, and J. A. Benediktsson, "Spectral–spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, Jan. 2016.
[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] 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.
[51] 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.



Leyuan Fang (S'10–M'14–SM'17) received the Ph.D. degree from Hunan University, Changsha, China, in 2015.

From 2011 to 2012, he was a Visiting Ph.D. Student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. From 2016 to 2017, he was a Post-Doctoral Research Fellow with the Department of Biomedical Engineering, Duke University. Since 2017, he has been an Associate Professor with the College of Electrical and Information Engineering, Hunan University. His research interests include sparse representation and multiresolution analysis in remote sensing and medical image processing.

Dr. Fang has received the Scholarship Award for Excellent Doctoral Student granted by the Chinese Ministry of Education in 2011.

Nanjun He (S'17) received the B.S. degree from the Central South University of Forestry and Technology, Changsha, China, in 2013. He is currently pursuing the Ph.D. degree with the Laboratory of Vision and Image Processing, Hunan University, Changsha.

Since 2017, he has been a Visiting Ph.D. Student with the Hyperspectral Computing Laboratory, University of Extremadura, Cáceres, Spain, supported by the China Scholarship Council. His research interests include hyperspectral imagery classification, hyperspectral imagery restoration, and computer vision in remote sensing images.

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

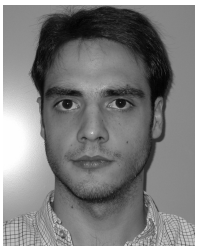
He was a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong, in 2011. From 2002 to 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, London, U.K., with Prof. J. Shawe-Taylor. In 2005, he visited the Department of Computer Science, Hong Kong University of Science and Technology, as a Visiting Professor. He joined the College of Electrical and Information Engineering, Hunan University, in 2001, where he is currently a Full Professor. He has authored or co-authored over 160 refereed papers. His research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

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

Antonio J. Plaza (M'05–SM'07–F'15) received the M.Sc. degree in computer engineering and the Ph.D. degree in computer engineering from the University of Extremadura, Badajoz, Spain, in 1999 and 2002, respectively.

He is currently the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored over 600 publications, including over 200 *Journal of Controlled Release* journal papers (over 150 in IEEE journals), 24 book chapters, and over 300 peer-reviewed conference proceeding papers. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Prof. Plaza is a fellow of the IEEE for contributions to hyperspectral data processing and parallel computing of Earth observation data. He was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER from 2011 to 2012 and the *IEEE Geoscience and Remote Sensing Magazine* in 2013. He was also a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He was a recipient of the recognition of best reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2009 and a recipient of the recognition of best reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2010, for which he served as an Associate Editor in 2007–2012. He was a recipient of the Best Column Award of the *IEEE Signal Processing Magazine* in 2015, the 2013 Best Paper Award of the JSTARS journal, and the most highly cited paper in 2005–2010 in the *Journal of Parallel and Distributed Computing*. He received best paper awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He is also an Associate Editor of the IEEE ACCESS. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012, and as a President of the Spanish Chapter of the IEEE GRSS from 2012 to 2016. He has reviewed over 500 manuscripts for over 50 different journals. He served as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal from 2013 to 2017. He has guest edited ten special issues on hyperspectral remote sensing for different journals.



Javier Plaza (M'09–SM'15) received the M.Sc. degree in computer engineering and the Ph.D. degree in computer engineering from the University of Extremadura, Badajoz, Spain, in 2004 and 2008, respectively.

He is currently a member of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored over 150 publications, including over 50 *Journal of Controlled Release* journal papers, ten book chapters, and 90 peer-reviewed conference proceeding papers. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was a recipient of the Outstanding Ph.D. Dissertation Award at the University of Extremadura in 2008. He was a recipient of the Best Column Award of the *IEEE Signal Processing Magazine* in 2015 and the most highly cited paper in 2005–2010 in the *Journal of Parallel and Distributed Computing*. He has guest edited three special issues on hyperspectral remote sensing for different journals. He is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and the IEEE Remote Sensing Code Library. He received best paper awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology.