Robust Matrix Discriminative Analysis for Feature Extraction From Hyperspectral Images

Renlong Hang, Qingshan Liu, *Senior Member, IEEE*, Yubao Sun, Xiaotong Yuan, Hucheng Pei, Javier Plaza, *Senior Member, IEEE*, and Antonio Plaza, *Fellow, IEEE*

***Abstract*—Linear discriminative analysis (LDA) is an effective feature extraction method for hyperspectral image (HSI) classification. Most of the existing LDA-related methods are based on spectral features, ignoring spatial information. Recently, a matrix discriminative analysis (MDA) model has been proposed to incorporate the spatial information into the LDA. However, due to sensor interferers, calibration errors, and other issues, HSIs can be noisy. These corrupted data easily degrade the performance of the MDA. In this paper, a robust MDA (RMDA) model is proposed to address this important issue. Specifically, based on the prior knowledge that the pixels in a small spatial neighborhood of the HSI lie in a low-rank subspace, a denoising model is first employed to recover the intrinsic components from the noisy HSI. Then, the MDA model is used to extract discriminative spatial–spectral features from the recovered components. Besides, different HSIs exhibit different spatial contextual structures, and even a single HSI may contain both large and small homogeneous regions simultaneously. To sufficiently describe these multiscale spatial structures, a multiscale RMDA model is further proposed. Experiments have been conducted using three widely used HSIs, and the obtained results show that the proposed method allows for a significant improvement in the classification performance when compared to other LDA-based methods.**

***Index Terms*—Hyperspectral image (HSI) classification, multiscale fusion, robust matrix discriminative analysis (RMDA), spatial-spectral feature extraction (FE).**

I. Introduction

HYPERSPECTRAL sensors can fully portray the surface of the Earth using hundreds of contiguous and narrow spectral bands, ranging from the visible to the near infrared part of the electromagnetic spectrum. Different materials exhibit different absorptions or reflections at a certain spectral bands. Such rich spectral information provides an avenue for accurate classifications of hyperspectral images (HSIs). The goal of such classification techniques is to assign one given class label to each pixel. However, the high-dimensional nature of HSIs, together with the (generally) limited number of training samples available often leads to the Hughes phenomenon [1]. Meanwhile, the processing of such high-dimensional datasets also requires advanced computation resources and storage capacities. Besides, the spectral bands are often correlated, and not all of them are useful for classification purposes.

To address these issues, feature extraction (FE) has been a widely used method in the literature [2], [3]. It aims at reducing the dimensionality of HSIs while retaining as much intrinsic information as possible. Popular FE methods include unsupervised and supervised approaches. A typical unsupervised method is principal component analysis (PCA) [4], which aims to find an orthogonal set of vectors that maximizes the variance of the projected vectors. However, without using the label information provided by training data, the extracted features by unsupervised methods may be not the optimal ones for the subsequent classification. Different from them, supervised methods focus on learning discriminative features from labeled data. Linear discriminative analysis (LDA) [5] is a classical supervised learning method, which generates the best projection by maximizing the between-class scatter matrix while minimizing the within-class scatter matrix. In the past decades, LDA and its variants have been widely explored in the field of remote sensing. Nonparametric weighted feature extraction (NWFE) is a creative extension of LDA [6]. It introduced a new criteria to underline the separability between the class distribution boundaries. To cope with a particular ill-posed problem, where the ratio between the number of training samples and the number of spectral features is small, a regularized LDA was proposed in [7]. In [8], motivated by [9], Li *et al.* employed a local Fisher's discriminant analysis (LFDA) model to characterize HSIs, which effectively combines the ideas of LDA and locality preserving projection [10]. Similarly, Liao *et al.* proposed to preserve the local neighborhood information inferred from unlabeled samples [11]. Recently, a regularized local discriminant embedding (RLDE) model has been proposed for HSI classifications [12].

Most of the aforementioned LDA-related FE methods are based on spectral features, while ignoring spatial information.

Manuscript received October 11, 2016; revised December 30, 2016; accepted January 20, 2017. Date of publication February 14, 2017; date of current version April 10, 2017. This work was supported in part by the Natural Science Foundation of China under Grant 61532009, Grant 61672292, and Grant 61402232, in part by the Natural Science Foundation of Jiangsu Province, China, under Grant 15KJA520001 and Grant BK20141003, in part by the Six Talent Peaks Project of Jiangsu Province, China, under Grant DZXX-037, and in part by the Innovation Project for Graduate Student of Jiangsu Province, China, under Grant KYZZ16-0350. *(Corresponding author: Qingshan Liu.)*

R. Hang, Q. Liu, Y. Sun, and X. Yuan are with the Jiangsu Key Laboratory of Big Data Analysis Technology, Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: renlong_hang@163.com; qsliu@nuist.edu.cn; sunyb@nuist.edu.cn; xtyuan1980@gmail.com).

H. Pei is with Beijing Electro-Mechanical Engineering Institute, Beijing 100083, China (e-mail: Peihch@163.com).

J. Plaza and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Computer Technology and Communications, University of Extremadura, Caceres E-10071, Spain (e-mail: jplaza@unex.es; aplaza@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/JSTARS.2017.2658948

With the development of imaging spectroscopy technologies, current sensors can acquire HSIs with very high spatial resolutions. For instance, the spatial resolution of reflective optics system imaging spectrometer (ROSIS) sensor can be up to 1.3 m. Therefore, the pixels in a small spatial neighborhood belong to the same class with a high probability. For a large homogeneous region, the pixels may have different spectral responses. If we only use the spectral features, the pixels will be classified into different subregions. On the contrary, for multiple neighboring regions, if we only use the spatial information, these regions will be classified as the same one. Thus, for accurate classifications, it is essential to take into account the spatial and spectral information simultaneously.

Spatial–spectral techniques have attracted increasing attention in recent years [13]–[15]. A wide variety of methods have been proposed, and they can be roughly categorized into three classes: feature-level fusion, decision-level fusion, and regularization-based fusion. Feature-level fusion techniques often extract the spectral features and the spatial features independently, and then concatenate them into a vector [16]–[20] or construct multiple kernel functions on their corresponding features [21]–[23] followed by a classifier. For decision-level fusion, multiple prediction results are first derived using the spatial and the spectral information, respectively, and then they are combined according to some strategies such as majority voting in [24]–[26]. For regularization-based fusion, a regularizer representing the spatial information is incorporated into the original objective function. For instance, in [27], [28], Markov random fields were employed to model the joint prior probabilities of each pixel and its spatial neighbors.

Recently, to fuse the spatial information into LDA directly, a matrix discriminative analysis (MDA) model has been proposed in [29]. Specifically, a matrix-based representation is designed for each pixel to capture the local spatial-contextual and the spectral information of all the spectral bands. Then, MDA is employed to learn the discriminative spatial–spectral features from it. Despite the effectiveness of this method, there still exist some issues. The first one is that the HSIs in real applications often suffer from various degradations, e.g., noise contamination, stripe corruption, and missing data, due to the sensor acquisition and calibration errors [30]. These corrupted data easily degrade the performance of MDA. Another issue is that different HSIs exhibit different spatial contextual structures, and even a single HSI may contain large and small homogeneous regions simultaneously. Thus, it is difficult to accurately describe these multiscale spatial structures with a predefined scale in MDA.

In this paper, a new multiscale robust MDA (MRMDA) model is proposed to address the aforementioned issues. The flowchart of the proposed method is shown in Fig. 1. It consists of three steps. First of all, for each pixel in a given HSI, multiscale cubes centered at it are extracted. Second, RMDA models are employed to learn the discriminative spatial–spectral features from the corrupted cubes. Finally, the learned representations are fed into a support vector machine (SVM) classifier to obtain a classification result at each scale, and a majority voting strategy is used to combine the complementary results from multiple scales. In the following section, we will introduce the core algorithm RMDA in detail.

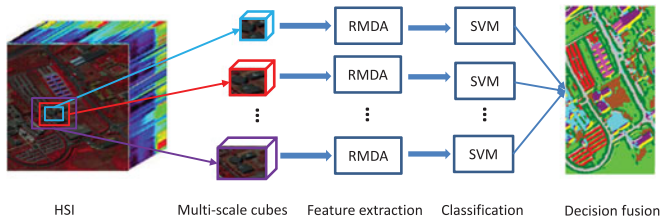


Fig. 1. Flowchart of the proposed MRMDA model.

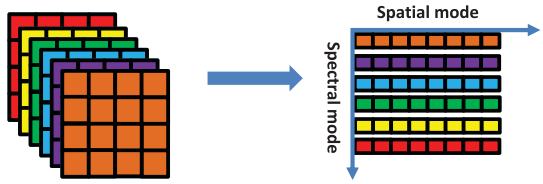


Fig. 2. Schematic diagram of the matrix-based spatial–spectral representation adopted in this work.

II. Robust Matrix Discriminative Analysis (RMDA)

Let us assume that an HSI is denoted as a three-dimensional (3-D) matrix $\mathbf{X} \in \mathbf{R}^{m\times n\times \ell}$ with $m \times n$ pixels and ℓ spectral bands. As discussed above, the pixels in a small spatial neighbourhood belong to the same class with a high probability, we thus exploit the spatial neighborhood to combine the spectral and spatial-contextual information. We select a subcube $\mathbf{X}_{ij}$ with size $\omega \times \omega \times \ell$ centered at the spatial position (i, j), where $1 \le i \le m$ and $1 \le j \le n$. Then, for each spectral band k, $1 \le k \le \ell$, we reshape the matrix $\mathbf{X}_{ijk} \in \mathbf{R}^{\omega\times\omega}$ to a row vector $\mathbf{r}_{ijk}$. Thereafter, we can obtain a matrix-based spatial–spectral representation of $\mathbf{X}_{ij}$: $\mathbf{X}_{ij}^{\text{new}} = [\mathbf{r}_{ij1}^\top, \mathbf{r}_{ij2}^\top, \ldots, \mathbf{r}_{ij\ell}^\top]^\top$, where $\mathbf{X}_{ij}^{\text{new}} \in \mathbf{R}^{\ell\times\omega^2}$. Fig. 2 demonstrates a schematic diagram of the matrix-based spatial–spectral representation.

Since HSIs often suffer from various degradations, $\mathbf{X}_{ij}^{\text{new}}$ can be decomposed as $\mathbf{X}_{ij}^{\text{new}} = \mathbf{Y}_{ij} + \mathbf{E}_{ij}$, where $\mathbf{Y}_{ij}$ represents the clean HSI data and $\mathbf{E}_{ij}$ denotes the noise. From a linear hyperspectral unmixing point of view, each pixel can be represented as a linear combination of a few pure substances or endmembers [31]. Let us suppose that the number of endmembers is ψ, then $\mathbf{Y}_{ij}$ can be decomposed as $\mathbf{Y}_{ij} = \mathbf{DA}$, where $\mathbf{D} \in \mathbf{R}^{\ell\times\psi}$ is a matrix whose columns denote the pure spectral endmembers, and $\mathbf{A} \in \mathbf{R}^{\psi\times\omega^2}$ is a matrix whose columns represent the percentage of each endmember. Since ψ is relatively small and $\text{rank}(\mathbf{Y}_{ij}) \le \psi$, $\mathbf{Y}_{ij}$ is a low rank matrix. Besides, the common degradations (including the impulse noise, dead pixels or lines, and stripes) only corrupt some parts or some bands of the HSI [30]. Thus, $\mathbf{E}_{ij}$ is a sparse matrix. Based on these motivations, we employ a new denoising model which can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{Y}_{ij},\mathbf{E}_{ij}} \quad & \|\mathbf{Y}_{ij}\|_* + \lambda\|\mathbf{E}_{ij}\|_1, \\ \text{s.t.} \quad & \mathbf{X}_{ij}^{\text{new}} = \mathbf{Y}_{ij} + \mathbf{E}_{ij} \end{aligned} \tag{1}$$

Fig. 3. Example of the proposed denoising model. (a) An image patch selected from the PUS dataset. (b) The recovered data from (a). (c) The noise component from (a).

where $\|\cdot\|_*$ denotes the *nuclear norm* of a matrix, i.e., the sum of singular values of the matrix, $\mathbf{E}_{ij}$ represents the sparse noise, which is measured by ℓ_1 norm, and $\lambda > 0$ is a regularization parameter. Inspired by [32], (1) can be solved by using an inexact augmented Lagrange multiplier method. Fig. 3 shows an example of how to recover the intrinsic component from a corrupted image patch.

After obtaining the new representation $\mathbf{Y}_{ij}$, the MDA model [33] is adopted to extract the most discriminative spatial–spectral features from it. Let $\mathbf{P} \in \mathbf{R}^{\ell \times r}$ and $\mathbf{Q} \in \mathbf{R}^{\omega^2 \times c}$ define two mapping matrices, then the projection of $\mathbf{Y}_{ij}$ onto the $(r \times c)$-dimensional space can be formulated as $\mathbf{Z}_{ij} = \mathbf{P}^\top \mathbf{Y}_{ij} \mathbf{Q}$. Let us suppose that there exist N training pixels labeled as Ω different classes, and $N_\iota, \iota \in \{1, 2, \ldots, \Omega\}$ is the number of training pixels in class ι. The mean values of all pixels and the pixels in the ιth class are $\mathbf{M}$ and $\mathbf{M}_\iota$, respectively. Similar to LDA, MDA aims to map the $\mathbf{Y}_{ij}$ to a subspace, where the between-class scatter matrix of the projected data is maximized, and the within-class scatter matrix of the projected data is minimized. Thus, the objective function of MDA can be described as

$$J(\mathbf{P}, \mathbf{Q}) = \max \frac{\mathbf{S}_B}{\mathbf{S}_W} \tag{2}$$

where $\mathbf{S}_B$ and $\mathbf{S}_W$ are defined as follows:

$$\begin{aligned} \mathbf{S}_B &= tr\left(\sum_{\iota=1}^{\Omega} N_\iota (\tilde{\mathbf{M}}_\iota - \tilde{\mathbf{M}})(\tilde{\mathbf{M}}_\iota - \tilde{\mathbf{M}})^\top\right), \\ \mathbf{S}_W &= tr\left(\sum_{\iota=1}^{\Omega} \sum_{\mathbf{Z}_{ij} \in \iota} (\mathbf{Z}_{ij} - \tilde{\mathbf{M}}_\iota)(\mathbf{Z}_{ij} - \tilde{\mathbf{M}}_\iota)^\top\right). \end{aligned} \tag{3}$$

Since $\tilde{\mathbf{M}}_\iota = \frac{1}{N_\iota} \sum_{\mathbf{Y}_{ij} \in \iota} \mathbf{P}^\top \mathbf{Y}_{ij} \mathbf{Q} = \mathbf{P}^\top \mathbf{M}_\iota \mathbf{Q}$, (3) can be rewritten as

$$\begin{aligned} \mathbf{S}_B &= tr\left(\sum_{\iota=1}^{\Omega} N_\iota \mathbf{P}^\top (\mathbf{M}_\iota - \mathbf{M}) \mathbf{Q}\mathbf{Q}^\top (\mathbf{M}_\iota - \mathbf{M})^\top \mathbf{P}\right), \\ \mathbf{S}_W &= tr\left(\sum_{\iota=1}^{\Omega} \sum_{\mathbf{Y}_{ij} \in \iota} \mathbf{P}^\top (\mathbf{Y}_{ij} - \mathbf{M}_\iota) \mathbf{Q}\mathbf{Q}^\top (\mathbf{Y}_{ij} - \mathbf{M}_\iota)^\top \mathbf{P}\right). \end{aligned}$$

The same as in [33], we adopt an iterative algorithm to optimize $\mathbf{P}$ and $\mathbf{Q}$. Specifically, for a fixed $\mathbf{Q}$, (2) can be rewritten as

$$J(\mathbf{P}) = \max \frac{\mathbf{P}^\top \mathbf{S}_B^{\mathbf{Q}} \mathbf{P}}{\mathbf{P}^\top \mathbf{S}_W^{\mathbf{Q}} \mathbf{P}}$$

Algorithm 1: Algorithm for the RMDA model.

Input: Training data $\mathbf{X}_{ij}$, the reduced dimensions r and c, the number of classes Ω
Output: Two mapping matrices $\mathbf{P}$ and $\mathbf{Q}$, the extracted representation $\mathbf{Z}_{ij}$

Reshape the three-dimensional data $\mathbf{X}_{ij}$ to a matrix-based representation $\mathbf{X}_{ij}^{new}$;
Recover the intrinsic component $\mathbf{Y}_{ij}$ from $\mathbf{X}_{ij}^{new}$ by using an inexact augmented Lagrange multiplier method [32].
Compute the mean values $\mathbf{M}_\iota$ for each class ι : $\mathbf{M}_\iota = \frac{1}{N_\iota} \sum_{\mathbf{Y}_{ij} \in \iota} \mathbf{Y}_{ij}$;
Compute the mean value of all pixels : $\mathbf{M} = \frac{1}{N} \sum_{\iota=1}^{\Omega} N_\iota \mathbf{M}_\iota$;
$\mathbf{Q} \leftarrow (\mathbf{I}^{c \times c}, \mathbf{0})^\top$;
for *each iteration* **do**
  $\mathbf{S}_B^{\mathbf{Q}} \leftarrow \sum_{\iota=1}^{\Omega} N_\iota (\mathbf{M}_\iota - \mathbf{M}) \mathbf{Q}\mathbf{Q}^\top (\mathbf{M}_\iota - \mathbf{M})^\top$;
  $\mathbf{S}_W^{\mathbf{Q}} \leftarrow \sum_{\iota=1}^{\Omega} \sum_{\mathbf{Y}_{ij} \in \iota} (\mathbf{Y}_{ij} - \mathbf{M}_\iota) \mathbf{Q}\mathbf{Q}^\top (\mathbf{Y}_{ij} - \mathbf{M}_\iota)^\top$;
  Compute the first r eigenvectors $[\mathbf{p}_1, \cdots, \mathbf{p}_r]$ of $(\mathbf{S}_W^{\mathbf{Q}})^{-1} \mathbf{S}_B^{\mathbf{Q}}$;
  $\mathbf{P} \leftarrow [\mathbf{p}_1, \cdots, \mathbf{p}_r]$;
  $\mathbf{S}_B^{\mathbf{P}} \leftarrow \sum_{\iota=1}^{\Omega} N_\iota (\mathbf{M}_\iota - \mathbf{M})^\top \mathbf{P}\mathbf{P}^\top (\mathbf{M}_\iota - \mathbf{M})$;
  $\mathbf{S}_W^{\mathbf{P}} \leftarrow \sum_{\iota=1}^{\Omega} \sum_{\mathbf{Y}_{ij} \in \iota} (\mathbf{Y}_{ij} - \mathbf{M}_\iota)^\top \mathbf{P}\mathbf{P}^\top (\mathbf{Y}_{ij} - \mathbf{M}_\iota)$;
  Compute the first c eigenvectors $[\mathbf{q}_1, \cdots, \mathbf{q}_c]$ of $(\mathbf{S}_W^{\mathbf{P}})^{-1} \mathbf{S}_B^{\mathbf{P}}$;
  $\mathbf{Q} \leftarrow [\mathbf{q}_1, \cdots, \mathbf{q}_c]$;
End **for**;
$\mathbf{Z}_{ij} \leftarrow \mathbf{P}^\top \mathbf{Y}_{ij} \mathbf{Q}$;

where $\mathbf{S}_B^{\mathbf{Q}} = \sum_{\iota=1}^{\Omega} N_\iota (\mathbf{M}_\iota - \mathbf{M}) \mathbf{Q}\mathbf{Q}^\top (\mathbf{M}_\iota - \mathbf{M})^\top$ and $\mathbf{S}_W^{\mathbf{Q}} = \sum_{\iota=1}^{\Omega} \sum_{\mathbf{Y}_{ij} \in \iota} (\mathbf{Y}_{ij} - \mathbf{M}_\iota) \mathbf{Q}\mathbf{Q}^\top (\mathbf{Y}_{ij} - \mathbf{M}_\iota)^\top$. Hence, the optimal solution of $\mathbf{P}$ consists of r eigenvectors corresponding to r maximal eigenvalues by computing an eigen-decomposition on $(\mathbf{S}_W^{\mathbf{Q}})^{-1} \mathbf{S}_B^{\mathbf{Q}}$. Subsequently, the optimal solution of $\mathbf{Q}$ can be obtained when $\mathbf{P}$ is fixed. The whole process will iterate until a predefined convergence condition is arrived. The pseudocode for the RMDA model is given in Algorithm 1. For the final classification stage, the learned representation $\mathbf{Z}_{ij}$ is reshaped to a vector and fed into the SVM classifier.

III. Experiments

A. Dataset

We test the proposed MRMDA model on three HSIs, which are widely used to evaluate classification algorithms.

1) Pavia University Scene (PUS) dataset: The first dataset was acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy, on July 8, 2002. The original image was recorded with 115 spectral channels ranging from 0.43 to 0.86 μm. After removing noisy bands, 103 bands are used. The image size is 610×340 pixels with a spatial resolution of 1.3 m. A three band false-color composite image and the ground-truth map are shown in Fig. 4(a)–(b). In the ground-truth map, there are nine classes of land covers with more than 1000 labeled pixels for each class.

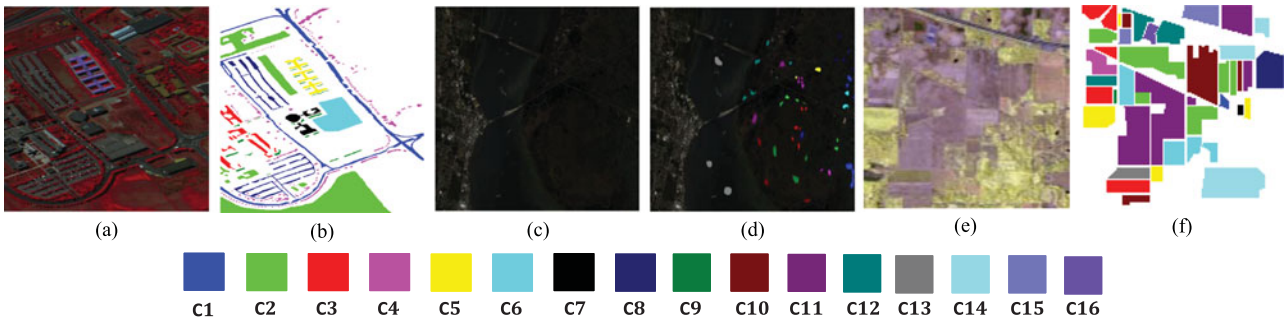


Fig. 4. RGB composite images and ground-truth maps on the three datasets. (a), (b) The PUS dataset. (c), (d) The KSC dataset. (e), (f) The IP dataset.

2) Kennedy Space Center (KSC) dataset: The second dataset was acquired by the AVIRIS sensor over KSC, Florida, on March 23, 1996. It contains 224 spectral bands. We utilize 176 bands of them after removing bands with water absorption and low signal noise ratio. The spatial size of the image is 512×614 pixels, and the spatial resolution is 18 m. Discriminating different land covers in this dataset is difficult due to the similarity of spectral signatures among certain vegetation types. For classification purposes, 13 classes representing the various land-cover types that occur in this environment are defined. Fig. 4(c)–(d) demonstrates a three band false-color composite image and the ground-truth map.
3) Indian Pines (IP) dataset: The third dataset was acquired by the AVIRIS sensor over the Indian Pine test site in northwestern Indiana, USA, on June 12, 1992 and it contains 224 spectral bands. We utilize 200 bands after removing four bands containing zero values and 20 noisy bands affected by water absorption. The spatial size of the image is 145×145 pixels, and the spatial resolution is 20 m. The false-color composite image and the ground-truth map are shown in Fig. 4(e)–(f). The available number of samples is 10 249 ranging from 20 to 2455 in each class.

B. Experimental Setup

We compare the proposed MRMDA model with several FE methods, including PCA, LDA, LFDA [8], NWFE [6], LDE [34], RLDE [12] and MDA [29]. Additionally, we also directly use the raw pixels as a benchmark. For LDA, the within-class scatter matrix $\mathbf{S}_W$ is replaced by $\mathbf{S}_W + \varepsilon\mathbf{I}$, where $\varepsilon = 10^{-3}$, to alleviate the singular problem. For LDE, a PCA preprocessing method is employed to overcome the singularity of the local preserving scatter matrix. Besides MDA, the optimal reduced dimensions for the other FE methods are chosen from [2, 30]. For MDA and RMDA, the optimal window size ω is selected from a given set $\{3, 5, 7, 9, 11\}$, while MRMDA combines the classification results from all of these sizes. The parameter λ in (1) is empirically set as $1/\sqrt{\ell}$ [32].

For the three datasets, we randomly select 5 and 10 pixels from each class as the training set, and use the remaining pixels as the testing set. The training set is used to learn the mapping functions for all of the FE methods. For classification purposes, the extracted features are fed into SVMs with linear kernels. The regularization parameter C in SVMs is chosen from a given set $\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$ via a fivefold cross validation method. Since HSIs contain more than two land-cover classes, an one-against-one strategy is adopted.

In order to reduce the effects of random selection, all the algorithms are repeated ten times and the average results are reported. The classification performance is evaluated by the overall accuracy (OA), the average accuracy (AA), the per-class accuracy, and the Kappa coefficient κ. OA defines the ratio between the number of correctly classified pixels to the total number of pixels in the testing set, AA refers to the average of accuracies in all classes, and κ is the percentage of agreement corrected by the number of agreements that would be expected purely by chance.

C. Parameter Selection

There are three important parameters in the MDA and the RMDA models, including the window size ω of each subcube, the reduced dimensions r and c. To show the effects of the window size on the classification performance, we fix the reduced dimensions and change the ω from 3 to 11. Fig. 5 demonstrates the changes of the OAs as a function of the window sizes on the three datasets. It can be observed that as the window size increases, the OAs of both methods first increase and then decrease when ω exceeds 5 for MDA and 11 for RMDA. Therefore, the optimal window sizes for MDA and RMDA are 5 and 11, respectively. Besides, it is interesting to compare the OAs of the MDA and the RMDA models. For the KSC dataset, RMDA achieves significantly superior performance as compared to MDA at any given window sizes. This validates the effectiveness of the RMDA model. Different from the KSC dataset, RMDA is inferior to MDA when $\omega \leq 7$ on the PUS and the IP datasets. This can be explained by the fact that the small patches in the PUS and the IP datasets are relatively clean, thus the eliminated noise $\mathbf{E}$ contains effective discriminant information for the classifications.

Similarly, to explore the effects of the reduced dimensions r and c, we fix the window size ω and select the r from 1 to ℓ with a step size 10, the c from 1 to ω^2 with a step size $\omega - 1$. Six 3-D diagrams of the OAs against different reduced dimensions for MDA and RMDA on the three datasets are shown in Fig. 6, where the x-axis denotes the reduced dimensions r, the y-axis the reduced dimensions c, and the z-axis the OAs.

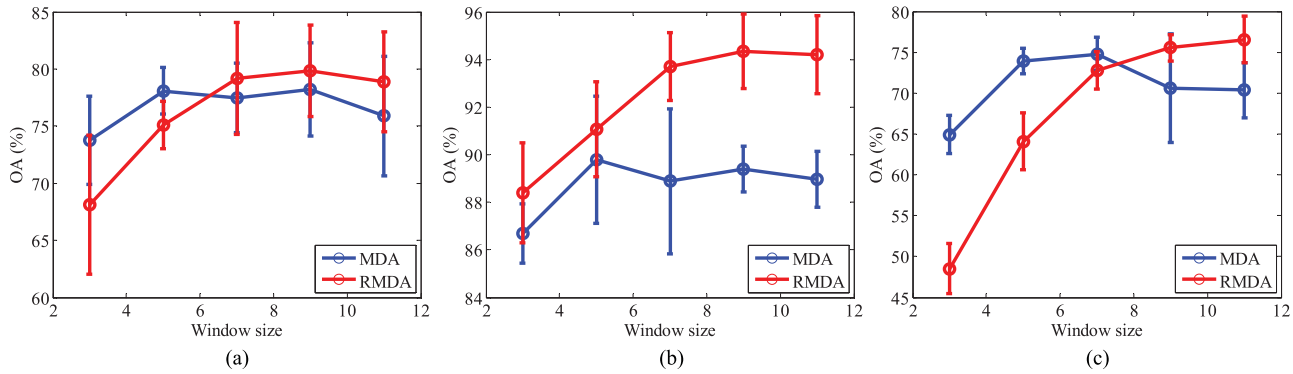


Fig. 5. OAs of the MDA and the RMDA models with different window sizes using 5 pixels from each class as the training set on the three datasets: (a) the PUS dataset, (b) the KSC dataset, and (c) the IP dataset.

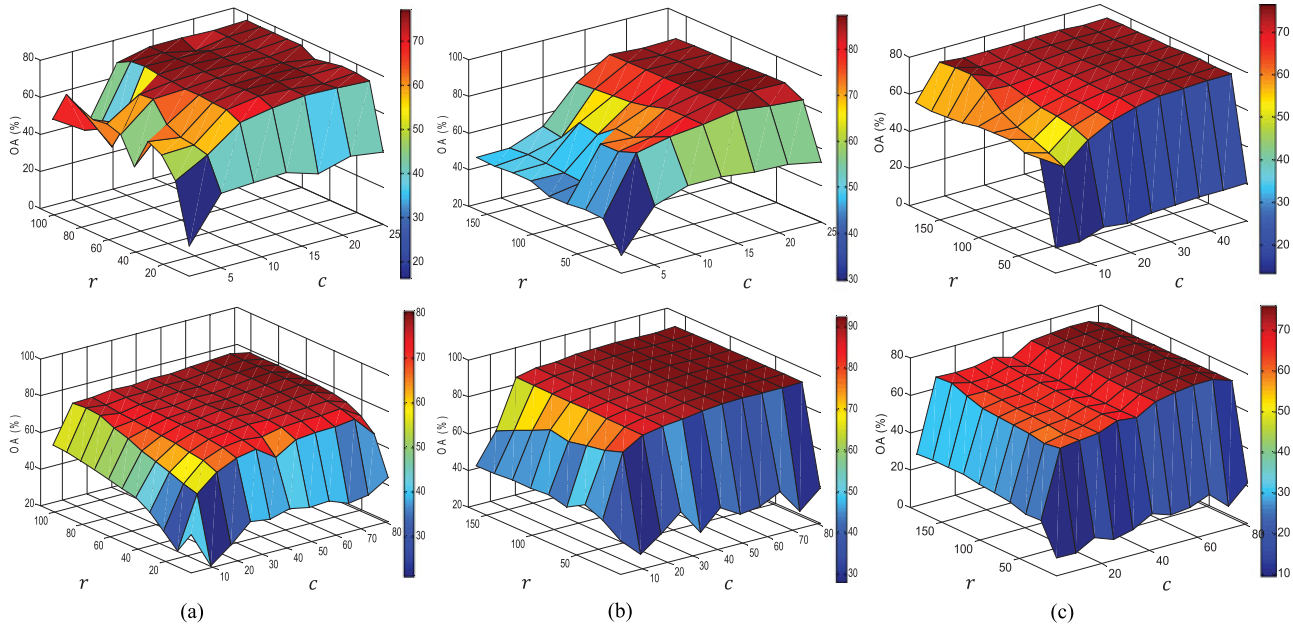


Fig. 6. OAs of the MDA (the first row) and the RMDA (the second row) models with different reduced dimensions r and c using 5 pixels from each class as the training set on the three different datasets: (a) the PUS dataset, (b) the KSC dataset, and (c) the IP dataset.

For the PUS dataset, the optimal dimensions are $r = 81, c = 9$ and $r = 71, c = 73$ for MDA and RMDA, respectively. For the KSC dataset, the optimal dimensions are $r = 41, c = 21$ and $r = 61, c = 37$ for MDA and RMDA, respectively. For the IP dataset, the optimal dimensions are $r = 141, c = 7$ and $r = 21, c = 73$ for MDA and RMDA, respectively.

D. Performance Comparison

To demonstrate the effectiveness of the proposed MRMDA model, we quantitatively and qualitatively compare it with the aforementioned FE methods. Tables I and II report the quantitative results achieved by ten different methods on the PUS dataset using 5 and 10 pixels from each class as the training set respectively, and bold values indicate the best results. From these tables, we can observe that PCA achieves higher performance than RAW in terms of OA, because it is able to extract the intrinsic information from the redundant spectral features. Besides, most of the LDA-related FE methods are better than PCA due to the use of label information in the training set. It is worth noting that LDE is inferior to RAW. This is caused by that the PCA preprocessing in LDE leads to the information losses for the subsequent FE. Instead of using the PCA preprocessing to solve the singular problem, RLDE adds a regularization term into LDE without discarding the discriminative information, thus it significantly improves the performance of LDE. Additionally, MDA achieves better performance than the other LDA-related methods which consider spectral information only, because it can extract spatial and spectral features simultaneously. This certifies the importance of spatial information in the FE and classifications of HSIs. Compared to MDA, RMDA obtains superior performance, since the HSI in real applications often suffers from various noises, and the corrupted data degrade the performance of MDA. However, due to the choice of the large window size, RMDA is difficult to accurately differentiate the small regions. MRMDA can address this issue by combining

TABLE I
OA, AA, PER-CLASS ACCURACY (%), κ AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE PUS DATASET USING 5 PIXELS FROM EACH CLASS AS THE TRAINING SET

Label	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	MDA	RMDA	MRMDA
C1	62.63 ± 10.68	63.61 ± 9.39	53.03 ± 13.00	59.67 ± 13.48	65.26 ± 9.24	61.77 ± 8.05	60.09 ± 10.84	71.24 ± 5.53	72.53 ± 13.25	**80.13 ± 8.87**
C2	63.78 ± 12.31	68.50 ± 12.67	75.35 ± 7.19	70.91 ± 15.16	73.19 ± 8.61	62.18 ± 15.47	79.47 ± 9.78	81.53 ± 7.66	81.88 ± 8.58	**81.96 ± 9.63**
C3	50.97 ± 14.64	50.17 ± 14.59	59.91 ± 12.55	56.47 ± 9.99	63.80 ± 20.37	37.72 ± 16.56	65.24 ± 5.25	82.79 ± 11.74	82.61 ± 11.21	**83.67 ± 13.61**
C4	83.17 ± 11.78	80.42 ± 14.76	76.48 ± 13.76	78.57 ± 8.17	83.19 ± 11.23	81.22 ± 17.30	79.32 ± 11.14	80.16 ± 2.52	79.37 ± 6.92	**83.22 ± 3.85**
C5	99.19 ± 0.36	99.16 ± 0.32	99.01 ± 0.96	99.55 ± 0.20	99.46 ± 0.39	99.34 ± 0.51	**99.73 ± 0.12**	95.49 ± 5.19	97.64 ± 3.27	99.18 ± 0.70
C6	62.30 ± 13.54	54.38 ± 14.81	54.43 ± 12.79	63.53 ± 10.57	57.13 ± 11.10	51.16 ± 14.00	56.78 ± 7.43	**91.21 ± 6.95**	91.01 ± 7.40	90.05 ± 7.31
C7	78.04 ± 9.88	78.11 ± 9.14	62.58 ± 11.62	69.57 ± 15.99	72.30 ± 17.77	68.05 ± 13.98	78.75 ± 11.16	78.94 ± 13.20	96.35 ± 2.17	**97.78 ± 1.20**
C8	51.77 ± 18.84	50.77 ± 20.44	49.04 ± 9.56	**60.38 ± 13.44**	51.28 ± 12.56	53.39 ± 20.29	56.08 ± 12.34	40.55 ± 14.38	49.09 ± 9.55	51.91 ± 16.69
C9	99.87 ± 0.12	99.87 ± 0.12	97.96 ± 2.51	99.26 ± 1.17	**99.96 ± 0.06**	99.83 ± 0.16	99.60 ± 0.53	97.49 ± 1.67	98.13 ± 1.04	99.17 ± 0.54
OA	65.52 ± 5.22	66.46 ± 4.58	67.33 ± 4.90	68.72 ± 4.99	69.88 ± 3.83	62.41 ± 4.46	72.13 ± 4.98	78.22 ± 4.10	79.84 ± 4.03	**81.61 ± 4.40**
AA	72.43 ± 3.13	71.67 ± 4.32	69.76 ± 5.65	73.10 ± 4.41	74.03 ± 4.86	68.29 ± 2.64	75.01 ± 3.85	79.94 ± 2.20	83.18 ± 1.92	**85.21 ± 1.87**
κ	56.75 ± 5.63	57.31 ± 5.05	58.21 ± 6.02	60.29 ± 5.54	61.40 ± 4.52	52.64 ± 4.17	63.93 ± 6.07	72.07 ± 5.01	74.19 ± 4.79	**76.42 ± 5.17**

TABLE II
OA, AA, PER-CLASS ACCURACY (%), κ AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE PUS DATASET USING 10 PIXELS FROM EACH CLASS AS THE TRAINING SET

Label	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	MDA	RMDA	MRMDA
C1	67.79 ± 2.59	67.35 ± 2.08	61.24 ± 15.51	66.63 ± 4.07	69.14 ± 6.66	57.83 ± 12.69	70.94 ± 7.82	80.55 ± 2.80	80.65 ± 2.00	**87.09 ± 4.53**
C2	61.91 ± 11.25	62.03 ± 10.96	67.21 ± 8.42	69.03 ± 9.82	65.83 ± 9.05	51.27 ± 11.91	72.19 ± 6.91	78.65 ± 8.42	**82.37 ± 5.71**	81.85 ± 4.05
C3	65.95 ± 9.52	65.39 ± 8.88	66.44 ± 4.12	48.75 ± 11.76	66.89 ± 7.53	64.38 ± 13.18	66.78 ± 3.28	74.13 ± 12.55	77.41 ± 13.41	**78.44 ± 10.42**
C4	92.38 ± 3.20	92.38 ± 3.16	94.68 ± 0.98	93.28 ± 2.13	93.55 ± 3.30	91.97 ± 4.21	**95.54 ± 1.51**	92.39 ± 4.39	88.04 ± 4.76	92.35 ± 2.87
C5	99.28 ± 0.09	99.25 ± 0.07	99.60 ± 0.11	**99.75 ± 0.09**	99.60 ± 0.19	99.55 ± 0.13	99.68 ± 0.09	99.60 ± 0.11	98.88 ± 0.64	99.50 ± 0.34
C6	68.46 ± 13.05	68.47 ± 12.69	71.92 ± 11.42	48.06 ± 10.38	65.45 ± 11.58	56.09 ± 6.06	64.33 ± 12.15	94.32 ± 2.78	**96.15 ± 2.53**	95.04 ± 1.52
C7	80.13 ± 9.23	80.56 ± 9.45	65.88 ± 23.21	72.32 ± 9.37	82.20 ± 10.25	75.83 ± 5.73	81.21 ± 12.86	96.49 ± 3.02	96.94 ± 2.22	**98.13 ± 2.07**
C8	72.29 ± 3.25	72.65 ± 3.01	64.59 ± 7.92	64.05 ± 13.29	74.39 ± 4.45	55.58 ± 4.69	**76.03 ± 8.59**	70.63 ± 14.34	59.89 ± 32.57	65.91 ± 26.51
C9	99.79 ± 0.11	99.79 ± 0.11	99.04 ± 1.57	99.79 ± 0.11	**99.96 ± 0.06**	99.89 ± 0	99.89 ± 0	98.86 ± 0.89	97.08 ± 2.30	98.90 ± 1.36
OA	69.42 ± 3.54	69.43 ± 3.45	70.21 ± 4.05	68.24 ± 2.31	71.38 ± 3.25	60.12 ± 2.64	74.56 ± 1.62	82.51 ± 4.61	83.24 ± 4.96	**84.85 ± 3.91**
AA	78.66 ± 1.92	78.65 ± 2.06	76.74 ± 1.30	73.52 ± 0.33	79.67 ± 1.39	72.49 ± 2.68	80.73 ± 2.24	87.29 ± 2.06	86.38 ± 3.31	**88.58 ± 2.79**
κ	61.97 ± 3.46	61.96 ± 3.41	62.86 ± 4.15	59.96 ± 2.19	64.14 ± 3.25	51.26 ± 2.31	67.70 ± 1.48	77.80 ± 5.44	78.53 ± 6.08	**80.61 ± 4.79**

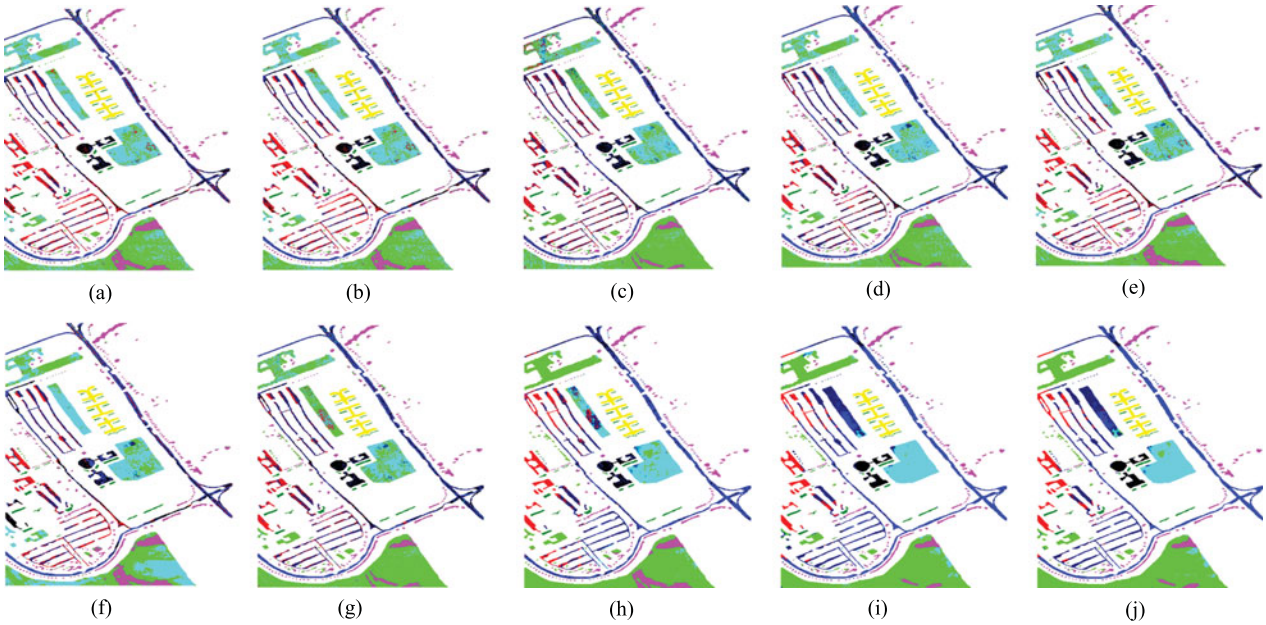


Fig. 7. Classification maps using ten different methods on the PUS dataset with 5 training pixels from each class. (a) RAW. (b) PCA. (c) LDA. (d) LFDA. (e) NWFE. (f) LDE. (g) RLDE. (h) MDA. (i) RMDA. (j) MRMDA.

TABLE III
OA, AA, PER-CLASS ACCURACY (%), κ AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE KSC DATASET USING 5 PIXELS FROM EACH CLASS AS THE TRAINING SET

Label	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	MDA	RMDA	MRMDA
C1	76.11 ± 12.43	79.26 ± 7.58	81.38 ± 4.88	82.01 ± 3.58	88.57 ± 3.75	70.82 ± 13.17	81.46 ± 4.33	94.70 ± 1.67	**98.20 ± 1.76**	97.01 ± 1.93
C2	79.75 ± 4.92	80.67 ± 5.26	74.03 ± 7.47	77.48 ± 9.63	80.08 ± 8.16	**82.10 ± 6.59**	81.09 ± 4.91	71.26 ± 9.62	77.23 ± 15.56	77.82 ± 15.91
C3	86.14 ± 5.73	83.59 ± 5.92	71.71 ± 10.20	79.04 ± 3.91	55.94 ± 45.25	87.89 ± 7.33	78.41 ± 5.21	92.75 ± 5.69	98.96 ± 1.25	**99.36 ± 0.60**
C4	49.31 ± 21.09	47.53 ± 19.27	43.64 ± 12.10	46.96 ± 14.99	32.87 ± 27.19	43.48 ± 15.96	40.49 ± 18.47	58.38 ± 8.99	76.36 ± 14.51	**78.54 ± 15.54**
C5	59.74 ± 11.39	58.59 ± 11.18	65.51 ± 11.14	63.85 ± 12.19	57.82 ± 25.03	51.28 ± 14.50	63.59 ± 7.87	82.82 ± 9.90	87.56 ± 9.31	**89.87 ± 10.19**
C6	46.96 ± 11.20	44.46 ± 11.38	60.27 ± 7.11	60.71 ± 5.75	30.18 ± 15.55	37.95 ± 7.93	55.71 ± 8.71	92.68 ± 3.86	**99.73 ± 0.40**	99.64 ± 0.37
C7	89.60 ± 8.91	89.40 ± 8.79	85.40 ± 9.93	88.80 ± 9.44	83.40 ± 18.96	90.20 ± 11.39	90.40 ± 8.38	**100 ± 0**	**100 ± 0**	**100 ± 0**
C8	69.01 ± 12.93	67.89 ± 13.12	78.50 ± 7.91	77.98 ± 10.38	45.68 ± 20.26	67.32 ± 14.15	84.69 ± 5.15	81.74 ± 13.52	90.33 ± 10.98	**90.94 ± 9.38**
C9	76.85 ± 17.44	76.43 ± 17.61	84.62 ± 11.22	89.01 ± 5.79	87.30 ± 10.58	75.84 ± 11.92	88.23 ± 10.73	**94.33 ± 6.61**	88.62 ± 10.31	90.68 ± 9.04
C10	79.45 ± 4.22	76.14 ± 9.32	94.79 ± 3.69	93.38 ± 4.52	70.58 ± 8.75	80.30 ± 7.50	89.62 ± 3.58	90.83 ± 9.65	**100 ± 0**	**100 ± 0**
C11	86.33 ± 4.84	86.28 ± 4.79	85.99 ± 5.89	85.27 ± 6.92	83.33 ± 6.99	91.50 ± 4.99	90.87 ± 5.11	92.08 ± 2.50	90.87 ± 7.46	**92.17 ± 6.66**
C12	78.88 ± 6.89	82.33 ± 3.22	88.03 ± 4.29	83.98 ± 8.89	75.54 ± 13.02	73.90 ± 13.10	83.33 ± 4.55	85.74 ± 7.22	98.99 ± 1.39	**99.16 ± 1.36**
C13	**100 ± 0**	**100 ± 0**	**100 ± 0**	**100 ± 0**	99.13 ± 0.59	99.83 ± 0.25	98.78 ± 0.62	99.54 ± 0.54	**100 ± 0**	**100 ± 0**
OA	79.09 ± 2.67	79.18 ± 2.32	82.91 ± 1.31	83.55 ± 2.12	75.48 ± 3.42	77.32 ± 2.57	83.46 ± 0.66	89.78 ± 2.68	94.33 ± 1.56	**94.85 ± 1.53**
AA	75.24 ± 1.53	74.81 ± 1.92	77.99 ± 1.56	79.11 ± 2.25	68.49 ± 4.51	73.26 ± 1.43	78.98 ± 0.87	87.43 ± 3.07	92.84 ± 2.17	**93.48 ± 2.35**
κ	76.79 ± 2.91	76.88 ± 2.54	81.01 ± 1.45	81.72 ± 2.35	72.67 ± 3.79	74.86 ± 2.77	81.63 ± 0.73	88.61 ± 2.99	93.70 ± 1.74	**94.27 ± 1.70**

TABLE IV
OA, AA, PER-CLASS ACCURACY (%), κ AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE KSC DATASET USING 10 PIXELS FROM EACH CLASS AS THE TRAINING SET

Label	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	MDA	RMDA	MRMDA
C1	86.29 ± 7.75	86.15 ± 7.54	78.19 ± 7.35	73.87 ± 5.64	85.33 ± 12.36	84.42 ± 8.87	88.26 ± 3.77	92.78 ± 1.67	95.34 ± 1.65	**97.63 ± 3.15**
C2	83.61 ± 3.07	82.66 ± 2.86	75.54 ± 7.65	76.22 ± 2.94	79.49 ± 10.33	81.46 ± 9.72	80.86 ± 4.29	88.67 ± 5.50	90.64 ± 9.71	**96.74 ± 4.22**
C3	91.71 ± 2.77	90.89 ± 3.05	70.33 ± 3.82	68.46 ± 4.25	70.65 ± 33.81	89.76 ± 5.35	83.25 ± 6.95	92.93 ± 5.84	97.64 ± 1.92	**99.02 ± 1.30**
C4	63.80 ± 10.47	66.69 ± 10.71	55.79 ± 8.41	50.25 ± 5.11	39.92 ± 35.97	52.89 ± 9.76	53.80 ± 7.86	87.02 ± 2.89	88.51 ± 3.04	**89.67 ± 4.10**
C5	67.68 ± 9.24	66.36 ± 9.44	67.95 ± 5.27	60.93 ± 5.36	48.34 ± 28.86	54.44 ± 8.16	74.70 ± 7.28	87.95 ± 5.41	**96.16 ± 4.28**	95.76 ± 5.03
C6	63.47 ± 5.87	61.37 ± 8.68	67.03 ± 10.44	57.08 ± 8.34	35.89 ± 13.46	43.11 ± 10.10	70.14 ± 4.35	88.13 ± 4.96	99.36 ± 0.52	**99.82 ± 0.41**
C7	90.11 ± 7.47	89.26 ± 8.40	90.11 ± 6.85	91.79 ± 4.30	89.05 ± 14.71	86.11 ± 9.49	92.63 ± 4.82	99.79 ± 0.47	**100 ± 0**	**100 ± 0**
C8	81.24 ± 10.78	80.81 ± 11.21	78.15 ± 10.55	75.01 ± 9.79	31.12 ± 11.95	76.86 ± 15.52	87.03 ± 2.32	93.40 ± 3.48	94.63 ± 4.66	**97.20 ± 3.60**
C9	88.71 ± 5.69	89.65 ± 6.12	89.02 ± 11.73	78.63 ± 20.26	91.92 ± 8.16	88.86 ± 6.07	93.65 ± 7.31	96.71 ± 5.40	**97.57 ± 4.25**	97.06 ± 5.74
C10	83.76 ± 10.04	83.96 ± 9.73	94.72 ± 2.63	92.28 ± 2.57	67.46 ± 10.85	89.19 ± 5.01	89.75 ± 4.50	95.18 ± 2.61	**100 ± 0**	**100 ± 0**
C11	89.49 ± 3.58	89.63 ± 3.42	88.26 ± 6.25	87.04 ± 2.58	81.32 ± 5.02	90.81 ± 5.01	89.78 ± 5.28	93.59 ± 5.51	93.50 ± 7.26	**93.69 ± 7.04**
C12	84.54 ± 7.06	83.61 ± 7.21	88.76 ± 2.91	88.76 ± 3.63	81.01 ± 4.18	80.41 ± 7.42	85.15 ± 4.70	85.07 ± 3.03	98.62 ± 1.56	**99.35 ± 1.01**
C13	**100 ± 0**	**100 ± 0**	**100 ± 0**	**100 ± 0**	98.89 ± 0.80	**100 ± 0**	98.52 ± 0.54	99.32 ± 0.09	**100 ± 0**	**100 ± 0**
OA	86.08 ± 2.20	85.97 ± 2.07	84.20 ± 1.63	81.04 ± 1.76	75.52 ± 2.55	83.53 ± 1.72	87.16 ± 1.70	93.24 ± 1.04	96.73 ± 1.47	**97.73 ± 1.07**
AA	82.64 ± 2.53	82.39 ± 2.59	80.30 ± 1.16	76.95 ± 1.23	69.26 ± 2.00	78.33 ± 1.85	83.65 ± 1.45	92.35 ± 1.20	96.31 ± 1.85	**97.38 ± 1.45**
κ	84.50 ± 2.42	84.37 ± 2.28	82.43 ± 1.80	78.94 ± 1.93	72.71 ± 2.76	81.66 ± 1.91	85.71 ± 1.88	92.47 ± 1.16	96.36 ± 1.63	**97.47 ± 1.20**

TABLE V
OA, AA, PER-CLASS ACCURACY (%), κ AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE IP DATASET USING 5 PIXELS FROM EACH CLASS AS THE TRAINING SET

Label	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	MDA	RMDA	MRMDA
C1	84.39 ± 5.62	80.49 ± 6.90	73.17 ± 14.74	74.15 ± 9.99	80.98 ± 6.77	74.15 ± 6.12	81.46 ± 7.83	92.19 ± 8.69	96.59 ± 5.06	**97.10 ± 3.70**
C2	38.24 ± 11.01	49.83 ± 9.40	33.62 ± 19.53	35.63 ± 17.99	44.22 ± 11.78	40.15 ± 20.29	40.38 ± 20.32	**68.57 ± 14.31**	59.89 ± 10.84	63.58 ± 12.93
C3	37.24 ± 6.30	39.59 ± 6.85	37.62 ± 13.94	37.84 ± 11.84	47.37 ± 9.11	42.13 ± 5.62	46.91 ± 13.30	66.64 ± 6.70	77.24 ± 17.61	**77.59 ± 19.34**
C4	46.72 ± 10.47	47.84 ± 10.93	36.98 ± 6.23	36.64 ± 7.80	46.21 ± 11.79	37.76 ± 11.44	47.67 ± 12.16	85.34 ± 11.27	93.62 ± 8.84	**94.40 ± 7.84**
C5	66.90 ± 9.47	62.43 ± 9.73	62.59 ± 9.85	67.57 ± 9.46	72.76 ± 13.21	71.38 ± 9.94	74.52 ± 9.93	**86.90 ± 12.15**	82.68 ± 10.28	83.01 ± 10.23
C6	74.07 ± 14.08	72.08 ± 12.03	75.53 ± 11.82	79.83 ± 12.60	84.36 ± 10.80	73.49 ± 12.21	87.17 ± 9.25	**97.24 ± 1.71**	87.64 ± 5.36	88.50 ± 5.80
C7	87.82 ± 3.64	84.35 ± 5.83	86.09 ± 7.78	86.96 ± 5.33	91.30 ± 0	86.09 ± 7.14	86.09 ± 5.67	**100 ± 0**	**100 ± 0**	**100 ± 0**
C8	66.51 ± 16.71	67.99 ± 17.15	60.21 ± 8.72	60.38 ± 8.34	76.45 ± 8.81	63.34 ± 18.16	63.55 ± 15.78	**96.49 ± 6.71**	94.08 ± 9.32	93.36 ± 9.98
C9	86.67 ± 15.63	80.00 ± 21.08	81.33 ± 17.26	81.33 ± 14.45	86.67 ± 12.47	84.00 ± 5.96	88.00 ± 16.60	**100 ± 0**	**100 ± 0**	**100 ± 0**
C10	41.96 ± 11.59	47.24 ± 17.56	38.26 ± 12.83	41.68 ± 11.93	47.11 ± 11.29	46.18 ± 17.36	52.41 ± 14.64	66.02 ± 15.37	69.87 ± 18.58	**71.02 ± 17.36**
C11	36.87 ± 14.01	33.11 ± 14.30	36.33 ± 10.28	34.68 ± 9.08	41.20 ± 18.91	31.32 ± 9.00	37.90 ± 11.63	58.48 ± 9.92	**70.14 ± 10.25**	68.13 ± 11.41
C12	39.08 ± 8.34	38.95 ± 7.38	37.52 ± 9.61	40.48 ± 6.85	47.14 ± 5.95	37.82 ± 4.33	49.22 ± 8.36	**75.58 ± 15.67**	70.44 ± 5.42	70.00 ± 8.23
C13	92.20 ± 4.18	90.60 ± 6.07	91.90 ± 3.47	92.40 ± 3.66	94.60 ± 2.27	94.10 ± 2.99	96.50 ± 2.24	**98.90 ± 0.89**	98.10 ± 1.39	98.30 ± 1.82
C14	65.70 ± 7.77	61.95 ± 8.40	61.70 ± 13.99	64.14 ± 14.00	67.84 ± 14.21	56.78 ± 12.36	70.78 ± 15.86	**90.65 ± 4.90**	90.11 ± 7.09	90.17 ± 6.73
C15	29.08 ± 7.14	25.04 ± 11.77	35.01 ± 12.23	33.60 ± 12.70	38.32 ± 13.15	33.18 ± 8.45	36.59 ± 10.25	75.75 ± 9.49	**78.58 ± 9.39**	77.85 ± 12.33
C16	88.86 ± 4.57	88.18 ± 3.65	86.82 ± 2.96	86.14 ± 4.71	88.41 ± 3.97	87.73 ± 5.47	88.18 ± 6.15	96.82 ± 5.29	98.18 ± 2.36	**98.19 ± 2.66**
OA	48.57 ± 4.78	48.05 ± 3.76	46.41 ± 3.45	47.61 ± 2.56	54.34 ± 3.73	47.07 ± 2.32	53.63 ± 4.02	74.79 ± 2.00	76.56 ± 2.87	**77.10 ± 2.36**
AA	61.40 ± 2.29	60.17 ± 3.52	58.42 ± 2.03	59.59 ± 1.15	65.93 ± 2.40	59.98 ± 1.55	65.46 ± 1.74	84.72 ± 1.12	85.45 ± 2.10	**85.70 ± 2.01**
κ	42.62 ± 4.98	42.34 ± 3.81	40.24 ± 3.68	41.64 ± 2.66	48.97 ± 3.74	41.26 ± 2.37	48.33 ± 4.27	71.54 ± 2.19	73.54 ± 3.16	**74.21 ± 2.58**

TABLE VI
OA, AA, PER-CLASS ACCURACY (%), κ AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE IP DATASET USING 10 PIXELS FROM EACH CLASS AS THE TRAINING SET

Label	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	MDA	RMDA	MRMDA
C1	82.22 ± 8.91	76.67 ± 13.41	78.89 ± 9.13	81.67 ± 8.91	88.89 ± 3.40	75.00 ± 9.21	86.67 ± 5.34	97.78 ± 2.32	**98.89 ± 1.52**	**98.89 ± 1.52**
C2	46.15 ± 7.52	40.24 ± 10.51	30.34 ± 5.10	31.20 ± 8.97	58.42 ± 9.58	50.97 ± 10.48	43.15 ± 9.19	72.28 ± 7.94	69.79 ± 8.17	**72.71 ± 7.97**
C3	45.41 ± 6.74	43.44 ± 14.20	33.68 ± 5.92	32.17 ± 4.44	53.32 ± 12.37	53.80 ± 5.07	57.44 ± 7.16	79.49 ± 2.77	90.32 ± 4.37	**90.93 ± 3.42**
C4	58.94 ± 9.64	59.56 ± 8.97	32.42 ± 4.96	38.94 ± 7.98	63.35 ± 9.56	49.16 ± 4.52	58.33 ± 12.07	91.37 ± 6.36	98.15 ± 2.05	**98.41 ± 2.58**
C5	79.87 ± 4.36	70.36 ± 9.00	66.43 ± 5.85	71.59 ± 5.25	80.89 ± 4.12	77.76 ± 7.45	83.13 ± 4.97	**94.67 ± 4.09**	93.57 ± 2.90	94.50 ± 1.92
C6	79.61 ± 6.59	64.47 ± 11.96	80.81 ± 3.67	80.75 ± 5.02	80.22 ± 9.17	79.31 ± 5.05	89.00 ± 4.64	**95.94 ± 4.23**	93.81 ± 4.99	94.36 ± 4.53
C7	90.00 ± 2.48	92.22 ± 3.04	86.67 ± 8.43	85.56 ± 4.97	93.33 ± 4.65	87.78 ± 2.48	92.22 ± 3.04	**100 ± 0**	**100 ± 0**	**100 ± 0**
C8	70.85 ± 11.98	70.94 ± 11.94	71.79 ± 7.25	69.62 ± 7.17	82.22 ± 8.45	71.41 ± 15.58	79.23 ± 9.50	**100 ± 0**	99.57 ± 0.54	99.49 ± 0.54
C9	94.00 ± 8.94	80.00 ± 13.04	78.00 ± 10.95	74.00 ± 20.74	100 ± 0	96.00 ± 5.48	96.00 ± 8.94	**100 ± 0**	**100 ± 0**	**100 ± 0**
C10	54.03 ± 5.39	57.36 ± 7.11	40.98 ± 7.40	42.20 ± 8.10	63.80 ± 6.85	52.97 ± 5.44	54.28 ± 4.46	**75.47 ± 5.62**	74.24 ± 8.13	74.51 ± 8.59
C11	33.01 ± 6.91	49.87 ± 5.12	32.53 ± 9.74	33.67 ± 8.74	47.20 ± 6.44	34.32 ± 7.67	37.03 ± 6.30	65.36 ± 11.42	**75.68 ± 8.89**	74.82 ± 8.74
C12	41.92 ± 10.79	35.44 ± 4.94	46.18 ± 6.41	43.95 ± 6.39	46.11 ± 14.81	43.05 ± 10.70	53.52 ± 7.88	85.66 ± 6.97	**86.66 ± 5.60**	86.48 ± 8.88
C13	94.05 ± 3.95	92.92 ± 3.19	93.33 ± 5.87	93.23 ± 6.78	94.77 ± 7.02	94.77 ± 5.55	97.03 ± 4.11	**99.18 ± 0.93**	98.56 ± 1.75	98.56 ± 1.75
C14	76.43 ± 9.95	78.53 ± 9.47	63.19 ± 16.32	69.07 ± 14.93	81.61 ± 9.98	72.92 ± 12.02	76.81 ± 18.88	87.24 ± 5.29	87.25 ± 5.43	**87.70 ± 5.66**
C15	34.57 ± 5.09	27.77 ± 8.15	49.57 ± 10.75	44.79 ± 5.22	43.62 ± 13.79	41.38 ± 5.76	52.34 ± 7.01	91.91 ± 9.07	**92.93 ± 10.80**	91.44 ± 11.69
C16	89.64 ± 6.18	89.16 ± 7.38	85.54 ± 5.96	83.13 ± 5.52	90.60 ± 5.80	87.23 ± 3.77	88.92 ± 5.28	98.55 ± 1.98	**99.76 ± 0.54**	**99.76 ± 0.54**
OA	53.69 ± 2.88	55.18 ± 1.44	47.13 ± 2.74	48.21 ± 3.27	62.42 ± 3.37	54.80 ± 4.04	57.89 ± 3.45	80.22 ± 3.42	83.16 ± 2.36	**83.51 ± 2.54**
AA	66.92 ± 1.02	64.81 ± 0.75	60.65 ± 1.78	60.97 ± 3.04	73.02 ± 2.69	66.74 ± 2.24	71.57 ± 1.60	89.68 ± 1.81	91.20 ± 0.85	**91.41 ± 1.22**
κ	48.40 ± 3.11	49.78 ± 1.57	41.17 ± 2.92	42.31 ± 3.25	57.77 ± 3.82	49.67 ± 4.36	53.06 ± 3.70	77.70 ± 3.77	80.95 ± 2.61	**81.35 ± 2.83**

the classification results from different window sizes. In particular, it improves the OA from 79.84% to 81.61% and 83.24% to 84.85%, respectively, when compared to RMDA. For per-class accuracies, MRMDA achieves the highest accuracies in five classes in Table I. Fig. 7 qualitatively shows the classification maps of different methods with 5 training pixels from each class. In these figures, different colors refer to different land covers. It is demonstrated that MDA, RMDA, and MRMDA remove the outliers in the homogeneous regions, leading to much smoother classification maps than the other methods. Compared to MDA and RMDA, MRMDA obtains the best results in *Asphalt* and *Meadows* classes from the visual perspective.

Tables III–VI list the performance of ten different methods on the KSC and the IP datasets in terms of OA, AA, κ, and per-class accuracy. Similar to the PUS dataset, MDA, RMDA, and MRMDA achieve higher performance than the other FE methods. In comparison with MDA, RMDA can remarkably improve the OA. Besides, the performance improvement from RMDA to MRMDA is not significant as that on the PUS dataset, because RMDA has already obtained a high performance and a further increase is relatively difficult. Nevertheless, in terms of per-class accuracy, MRMDA achieves higher OAs than RMDA in most classes, which also indicates the superiority of the MRMDA.

IV. CONCLUSION

In this paper, we have proposed a new model for spatial–spectral FE from HSIs. Based on the prior knowledge that the pixels in a small spatial neighborhood of the image often lie in a low-rank subspace, a denoising model was employed to recover the intrinsic component from the corrupted data. The recovered data were then fed into MDA to learn the spatial–spectral discriminative features. Due to the existence of homogeneous regions with different sizes in the image, it is difficult to choose an optimal scale in MDA. To address this issue, a new so-called MRMDA model was proposed to combine the complementary information from different scales. Specifically, for each pixel, multiscale cubes centered at it were extracted, and a majority voting strategy was used to fuse the classification results of these cubes. To validate the effectiveness of the proposed model, we compared it with several FE methods using three widely used HSIs. The obtained results show that RMDA achieves higher performance than MDA, and MRMDA can further improve the performance as compared to RMDA. Besides, we also thoroughly evaluated the effects of different parameters, including the window size ω, the reduced dimensions r and c, on the classification performance.

REFERENCES

[1] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
[2] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.
[3] Z. Wu, Y. Li, A. Plaza, and J. Li, "Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 6, pp. 2270–2278, Jun. 2016.
[4] H. Abdi and L. J. Williams, *Principal Component Analysis*, vol. 2. Hoboken, NJ, USA: Wiley, 2010.
[5] B. Scholkopft and K.-R. Mullert, "Fisher discriminant analysis with kernels," in *Proc. IEEE Signal Process. Soc. Workshop*, 1999, pp. 41–48.
[6] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.
[7] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.
[8] 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.
[9] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, no. 1, pp. 1027–1061, 2007.
[10] X. He, "Locality preserving projections," *Adv. Neural Inf. Process. Syst.*, vol. 45, no. 1, pp. 186–197, 2005.

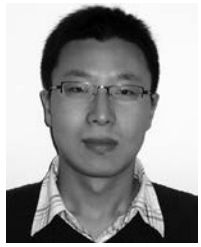
[11] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.
[12] Y. Zhou, J. Peng, and C. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, pp. 1082–1095, Feb. 2015.
[13] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 9, pp. S110–S122, 2009.
[14] M. Fauvel, 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.
[15] Z. Wu, Q. Wang, A. Plaza, and J. Li, "Parallel spatialcspectral hyperspectral image classification with sparse representation and Markov random fields on gpus," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 1–13, Jun. 2015.
[16] 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.
[17] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
[18] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosc. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[19] M. Dalla 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.
[20] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
[21] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial–spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, 2012.
[22] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
[23] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.
[24] L. O. Jimenez, J. L. Rivera-Medina, E. Rodríguez-Díaz, E. Arzuaga-Cruz, and M. Ramírez-Vélez, "Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 844–851, Apr. 2005.
[25] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
[26] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.
[27] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.
[28] X. Jia and J. Richards, "Managing the spectral-spatial mix in context classification using Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 311–314, Apr. 2008.
[29] R. Hang, Q. Liu, H. Song, and Y. Sun, "Matrix-based discriminant subspace ensemble for hyperspectral image spatialcspectral feature fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 783–794, Feb. 2016.
[30] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
[31] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
[32] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–73, Dec. 2009.
[33] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," *Adv. Neural Inf. Process. Syst.*, vol. 5, no. 6, pp. 1431–1441, 2004.
[34] H. T. Chen, H. W. Chang, and T. L. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 846–853.

Renlong Hang received the B.S. and M.S. degrees from the Nanjing University of Information Science and Technology, Nanjing, China, in 2011 and 2014, respectively. He is currently working toward the Ph.D. degree in the School of Atmospheric Science, Nanjing University of Information Science and Technology.

His current research interests include machine learning and pattern recognition.

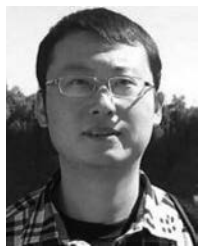
Qingshan Liu (M'05–SM'07) received the M.S. degree in automation from Southeast University, Nanjing, China, in 2000, and the Ph.D. degree in pattern recognition from the Chinese Academy of Sciences, Beijing, China, in 2003.

From 2010 to 2011, he was an Assistant Research Professor in the Department of Computer Science, Computational Biomedicine Imaging and Modeling Center, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. Before he joined Rutgers University, he was an Associate Professor with the National Laboratory of Pattern Recognition, Chinese Academy of Sciences. During June 2004 and April 2005, he was an Associate Researcher with the Multimedia Laboratory, The Chinese University of Hong Kong, Hong Kong. He is currently a Professor in the School of Information and Control Engineering, Nanjing University of Information Science and Technology, Nanjing. His research interests include image and vision analysis.



Yubao Sun received the Ph.D. degree in computer science and technology from the Nanjing University of Science and Technology, Nanjing, China, in 2010.

He is currently an Associate Professor in the School of Information and Control Engineering, Nanjing University of Information Science and Technology. His current research interests include sparse representation theory and application, low-rank representation, video analysis and understanding, and graph and hypergraph learning.

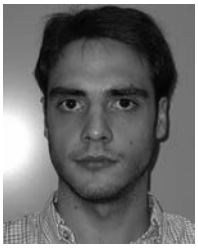


Xiaotong Yuan received the B.A. degree in computer science from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2002, the M.E. degree in electrical engineering from Shanghai Jiao-Tong University, Minhang Qu, China, in 2005, and the Ph.D. degree in pattern recognition from the Chinese Academy of Sciences, Beijing, China, in 2009.

After graduation, he held various appointments as a Postdoctoral Research Associate working in the Department of Electrical and Computer Engineering at the National University of Singapore, the Department of Statistics and Biostatistics at Rutgers University, and the Department of Statistical Science at Cornell University. In 2013, he joined the Nanjing University of Information Science and Technology, Nanjing, where he is currently a Professor of computer science. His main research interests include machine learning, data mining, and computer vision.

Hucheng Pei received the B.S. and M.S. degrees in automation from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively.

He is currently a Professor at the Beijing Electro-Mechanical Engineering Institute, Beijing, China. His research interests include signal processing and automatic target recognition.



Javier Plaza (M'09–SM'15) received the M.Sc. and the Ph.D. degrees 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 received the Outstanding Ph.D. Dissertation Award at the University of Extremadura in 2008. His main research interests include hyperspectral data processing and parallel computing of remote sensing data. He has authored more than 120 publications, including 36 JCR journal papers, ten book chapters, and 80 peer-reviewed conference proceeding papers. He has guest edited two special issues on hyperspectral remote sensing for different journals.

Dr. Plaza is an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and an Associate Editor of the IEEE REMOTE SENSING CODE LIBRARY. He received the Best Column Award of the IEEE Signal Processing Magazine in 2015 and the most highly cited paper (2005–2010) in the *Journal of Parallel and Distributed Computing*, and also best paper awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology.

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

He is the Head of the Hyperspectral Computing Laboratory in the Department of Technology of Computers and Communications, University of ExtremaduraHis main research interests include hyperspectral data processing and parallel computing of remote sensing data. He has authored more than 500 publications, 23 book chapters, and 285 peer-reviewed conference proceeding papers. He has guest edited 10 special issues on hyperspectral remote sensing for different journals.

Dr. Plaza is a Fellow of IEEE for contributions to hyperspectral data processing and parallel computing of Earth observation data. He received the recognition of Best Reviewers of the IEEE Geoscience and Remote Sensing Letters (in 2009) and 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 is also an Associate Editor for IEEE Access, and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE(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 received 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 (2005–2010) in the *Journal of Parallel and Distributed Computing*, and also best paper awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and is currently serving as the President of the Spanish Chapter of IEEE GRSS. He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal.