Spectral–Spatial Classification of Hyperspectral Data via Morphological Component Analysis-Based Image Separation

Zhaohui Xue, Jun Li, *Member, IEEE*, Liang Cheng, and Peijun Du, *Senior Member, IEEE*

Abstract—This paper presents a new spectral–spatial classification method for hyperspectral images via morphological component analysis-based image separation rationale in sparse representation. The method consists of three main steps. First, the high-dimensional spectral domain of hyperspectral images is reduced into a low-dimensional feature domain by using minimum noise fraction (MNF). Second, the proposed separation method is acted on each features to generate the morphological components (MCs), i.e., the content and texture components. To this end, the dictionaries for these two components are built by using local *curvelet* and *Gabor* wavelet transforms within the randomly chosen image partitions. Then, sparse coding of one of the MCs and update of the associated dictionary are sequentially performed with the other one fixed. To better direct the separation process, an undecimated *Haar* wavelet with soft threshold is performed for the content component to make it smooth. This process is repeated until some stopping criterion is met. Finally, a support vector machine is adopted to obtain the classification maps based on the MCs. The experimental results with hyperspectral images collected by the National Aeronautics and Space Administration Jet Propulsion Laboratory's Airborne Visible/Infrared Imaging Spectrometer and the Reflective Optics Spectrographic Imaging System indicate that the proposed scheme provides better performance when compared with other widely used methods.

Index Terms—Hyperspectral imaging, image separation, morphological component analysis (MCA), sparse representation, spectral–spatial classification, support vector machine (SVM).

I. INTRODUCTION

HYPERSPECTRAL remote sensing sensors provide hundreds of contiguous bands for the same area on the surface of the Earth [1], [2], allowing for the acquisition of plenty of useful information that increases the accurate discrimination of spectrally similar materials of interest. Hyperspectral images have been extensively and increasingly exploited in fusion, unmixing, classification, target detection, land physical and chemical parameter estimation, and fast computing [3]. Among many processing tasks, classification has been attracting plenty of attention from researchers in the last few decades, which consists of assigning each pixel with one thematic class for an object in a scene [2].

The curse of dimensionality and the generally limited availability of labeled training samples pose critical challenge to supervised classification of hyperspectral images since theoretical and practical problems arise with the increase in dimensionality [3]–[5]. More precisely, the high dimensionality and the limited available training samples are likely to make the accurate classification an ill-posed problem [6]. Intuitively, we can resort to dimensionality reduction to address such ill-posed problems [7]–[9]. Another challenge for pixelwise classification is considering the spatial–contextual information [2], [10], which will further improve the classification performance [11]. Extensive works have been done in the remote sensing community during the last decade to build accurate classifiers for spectral–spatial classification, such as segmentation-based [12]–[17], Markov random field-based [18]–[21], graph-cut-based [22]–[24], morphological profile filter-based [25]–[27], and many other approaches. These approaches have revealed that the classification performance can be greatly improved by integrating the contextual information with the spectral information.

Recently, integrating image separation [28], [29] with classification methods has been a new trend to improve the classification performance, in which images are exhaustively split into several independent components representing constructive information for classification, such as piecewise-smooth (content) and texture components. Image separation has been widely studied in many applications such as image inpainting, denoising, edge detection, object recognition, and image content editing [30]–[34]. More recently, the authors in [33] proposed a morphological component analysis (MCA)-based image separation approach, where the separated components are represented by the transformed coefficients. Then, in [34], sparsity and morphological diversity have emerged as effective features for blind source separation. The basic idea for MCA and sparse representation based image separation is to choose two dictionaries for the representation of morphological components (MCs) (i.e., content and texture) and compute the sparse coefficients over the images they are serving. Total

variation (TV) regularization [35] is generally employed to better direct the separation process.

A perfect image separation has been achieved in these works. However, limitations appear when these approaches being applied to hyperspectral images. First and foremost, these traditional methods are not completely sparse representation based; some are wavelet transformation based problems, where the dictionaries are implicit and cannot be updated and the obtained wavelet transformation coefficients are not completely sparse for promoting the separation process. Furthermore, from a computational viewpoint, the conventional image separation strategy was often acted on the whole image, which is generally time-consuming and leads to large memory burden when applied to hyperspectral data.

Inspired by the idea to design a sparse representation based approach for MCA, in this paper, we propose a new spectral–spatial classification scheme for hyperspectral images, which consists of three steps. First, a hyperspectral image is reduced to several major features preserving spectral information. Since hyperspectral data generally live in a much lower subspace than the original spectral space, we use the minimum noise fraction (MNF) [8] to reduce the dimensionality of the input data. MNF has been widely used in various hyperspectral image studies for dimensionality reduction. Second, the obtained features are split into MCs containing spatial information by performing image separation on each feature. In this step, the *curvelet* transform and *Gabor* wavelet transform (GWT) are used to build the content and texture dictionaries, respectively. Then, sparse coding is promoted by using a spectral unmixing by variable splitting and augmented Lagrangian (SUnSAL) algorithm [36]. In order to iteratively update the dictionaries, some TV and hard threshold regularization strategies are adopted. It is worth noting that, in order to address some of the limitations that the conventional approaches have, we focus on local image partitions to randomly generate the dictionaries and conduct sparse representation to achieve the whole image separation. Moreover, the separated components can capture more detailed information due to the adopted local and sparse representation based approach. Finally, all the extracted MCs are stacked for classification by employing a support vector machine (SVM) [37], which is a well-established classifier in hyperspectral imaging.

The main novelty and contribution of this paper is to propose a new sparse representation based method for image separation via MCA, with the goal of extracting relevant components jointly using the spatial and spectral information from the data, and then use these components for classification purpose.

The remainder of this paper is organized as follows. Section II formulates the considered problem and introduces the proposed MCA-based image separation and classification scheme. Section III describes an experimental evaluation of the proposed scheme. The experiment is conducted using two real hyperspectral data sets respectively collected by the Airborne Visible/infrared Imaging Spectrometer (AVIRIS) over the Indian Pines region in Indiana, and by the Reflective Optics Spectrographic Imaging System (ROSIS) over the city of Pavia, Italy. A comparison with some advanced spectral–spatial techniques for hyperspectral image classification is also reported in
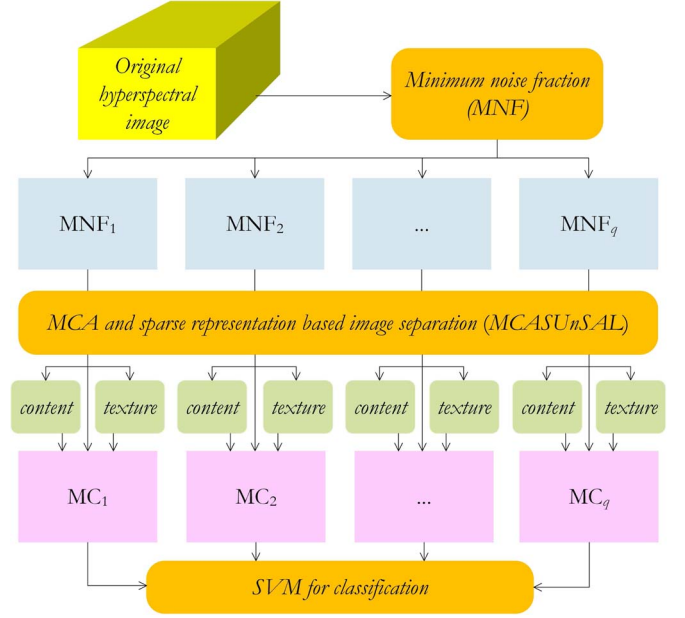

Fig. 1. Proposed image separation and classification scheme. Here, q denotes the number of input MNF features, and MC denotes the MC that can be content, texture, stacked content and texture, or the direct summation of the content and texture.

this section. Finally, Section IV concludes with some remarks and hints at some plausible future research lines.

II. MCA-BASED HYPERSPECTRAL IMAGE SEPARATION AND CLASSIFICATION SCHEME

First of all, we introduce the notations that will be adopted throughout this paper. Let $\mathbf{L} = \{\mathbf{l}_i \in \mathbb{R}^N, i = 1, 2, \ldots, B\}$, denote the observed hyperspectral data, where N is the number of pixels and B is the number of spectral bands in $\mathbf{L}$. In the first step, the high-dimensional spectral domain of the hyperspectral images is reduced into a low-dimensional feature domain by using MNF, which preserves the spectral information simultaneously. In the second step, the MCs that represent the spatial information are extracted from the obtained features by using the proposed sparse representation based MCA technique. In the final step, the method is evaluated by using SVM based on the extracted MCs. In the remainder of this section, we will review the details of each step, particularly for the second step, which contains the main contribution of this paper. The proposed MCA-based image separation and classification scheme is graphically illustrated in Fig. 1.

A. Dimensionality Reduction

It should be noted that dimensionality reduction is not mandatory for the presented scheme. We could perform MCA-based image separation for each band of the original data. Nevertheless, here, we use a dimensionality reduction step as hyperspectral data generally live in a much lower subspace than the original spectral space. The purpose of this step is twofold. One reason is that, by using only the most relevant features, we can work in the subspace where the hyperspectral data live,

which can significantly reduce the computational complexity without sacrificing accuracy.

In this paper, we use MNF for dimensionality reduction as it is widely used and very effective for various hyperspectral imaging problems [38]. Another reason is that MNF is robust to heterogeneous distributed noise, image linear transformation, and pixel value fluctuation [39]. Nevertheless, other dimensionality reduction approaches could be also used since MNF is just an effective dimensionality reduction method that is not directly connected with the classification step. Principal component analysis (PCA), which is a routine dimensionality reduction method for hyperspectral images, has significant role in data compressing, decorrelation, denoising, and feature extraction [7]. As a competitive dimensionality reduction method, we also adopt it for comparison in this paper. Therefore, after dimensionality reduction, the observed hyperspectral data $\mathbf{L}$ are transformed to $\mathbf{Y} = \{\mathbf{y}_i \in \mathbb{R}^N, \ i = 1, 2, \ldots, q\}$, where q is the number of features with $q \ll B$.

B. Image Separation

In this step, we perform image separation for each MNF feature to extract the MCs. Since our proposed approach is sparse representation based, an important issue is to build the dictionaries for the content and texture components. First of all, we randomly choose several image partitions. Then, for each component, we build the dictionaries by certain transformations on the chosen image partitions and then stack the dictionaries from each partition to obtain the specific dictionary. Precisely, for the content component, a local *curvelet* transform is performed to generate the dictionary from the randomly chosen image partitions. As for the texture component, a local GWT, which is considered one of the best representative transformations for texture extraction among orthogonal, biorthogonal, and tree-structured wavelet transforms [40], is adopted to build the dictionary from the same image partitions.

Let $\mathbf{A}_c \in \mathbb{R}^{1 \times M_c}$ be the dictionary for the content component computed from the *curvelet* transform, where M_c is the number of atoms.[1] In $\mathbf{A}_c$, let $\mathbf{A}_t \in \mathbb{R}^{1 \times M_t}$ be the dictionary for the texture component computed from the GWT, where M_t is the number of atoms in $\mathbf{A}_t$; let $\mathbf{x}_c \in \mathbb{R}^{M_c \times N}$ and $\mathbf{x}_t \in \mathbb{R}^{M_t \times N}$ be the sparse coefficients corresponding to the content $\mathbf{y}_c$ and texture $\mathbf{y}_t$ components, respectively. With these definitions in place, an image feature $\mathbf{y} \in \mathbb{R}^{1 \times N}$ can be considered a linear superposition of $\mathbf{y}_c$ and $\mathbf{y}_t$ as follows:

$$\mathbf{y} = \mathbf{y}_c + \mathbf{y}_t + \mathbf{v}$$
$$= \mathbf{A}_c \mathbf{x}_c + \mathbf{A}_t \mathbf{x}_t + \mathbf{v} \qquad (1)$$

where $\mathbf{v}$ is the noise term. In order to recover $\mathbf{y}_c$ and $\mathbf{y}_t$, we need to solve the following optimization problem [31]:

$$\langle \widehat{\mathbf{x}}_c, \widehat{\mathbf{x}}_t \rangle = \underset{\mathbf{x}_c, \mathbf{x}_t}{\arg\min} \frac{1}{2} \|\mathbf{y} - \mathbf{A}_c \mathbf{x}_c - \mathbf{A}_t \mathbf{x}_t\|_2^2 + \lambda_1 \|\mathbf{x}_c\|_1 + \lambda_2 \|\mathbf{x}_t\|_1 \qquad (2)$$

where λ_1 and λ_2 are the regularization parameters.

[1]Each column element is an atom in a dictionary.

In this paper, instead of finding the sparse coefficient matrices $\langle \widehat{\mathbf{x}}_c, \widehat{\mathbf{x}}_t \rangle$, we adopt an alternative approach to recover those two MCs directly. Since $\mathbf{y}_c = \mathbf{A}_c \mathbf{x}_c$ and $\mathbf{y}_t = \mathbf{A}_t \mathbf{x}_t$, it is very easy to obtain $\mathbf{x}_c = (\mathbf{A}_c^T \mathbf{A}_c)^{-1} \mathbf{A}_c^T \mathbf{y}_c$ and $\mathbf{x}_t = (\mathbf{A}_t^T \mathbf{A}_t)^{-1} \mathbf{A}_t^T \mathbf{y}_t$. Let $\mathbf{T}_c = (\mathbf{A}_c^T \mathbf{A}_c)^{-1} \mathbf{A}_c^T$ and $\mathbf{T}_t = (\mathbf{A}_t^T \mathbf{A}_t)^{-1} \mathbf{A}_t^T$, problem (2) now changes to

$$\langle \widehat{\mathbf{y}}_c, \widehat{\mathbf{y}}_t \rangle = \underset{\mathbf{y}_c, \mathbf{y}_t}{\arg\min} \frac{1}{2} \|\mathbf{y} - \mathbf{y}_c - \mathbf{y}_t\|_2^2 + \lambda_1 \|\mathbf{T}_c \mathbf{y}_c\|_1 + \lambda_2 \|\mathbf{T}_t \mathbf{y}_t\|_1. \qquad (3)$$

Problem (3) is a variant of the $\ell_2 - \ell_1$ problem, which is very difficult to solve. However, several strategies have been discussed in the literature to address this optimization problem [41]. In this paper, we adopt the following iterative scheme to obtain the local minima at the $k + 1$th iteration, with respect to $\mathbf{y}_c$ and $\mathbf{y}_t$:

$$\widehat{\mathbf{y}}_c^{k+1} = \underset{\mathbf{y}_c}{\arg\min} \ \lambda_1 \|\mathbf{T}_c \mathbf{y}_c\|_1 + \frac{1}{2} \left\|\mathbf{y} - \mathbf{y}_c - \widehat{\mathbf{y}}_t^k\right\|_2^2 \qquad (4)$$

$$\widehat{\mathbf{y}}_t^{k+1} = \underset{\mathbf{y}_t}{\arg\min} \ \lambda_2 \|\mathbf{T}_t \mathbf{y}_t\|_1 + \frac{1}{2} \left\|\mathbf{y} - \widehat{\mathbf{y}}_c^k - \mathbf{y}_t\right\|_2^2. \qquad (5)$$

The numerical solution of problems (4) and (5) can be, respectively, obtained as follows:

$$\widehat{\mathbf{y}}_c^{k+1} = \mathbf{A}_c^k \left(\Phi \left(\lambda_1, \mathbf{A}_c^k, G_c \left(\mathbf{y} - \widehat{\mathbf{y}}_t^k \right) \right) \right)$$
$$\widehat{\mathbf{y}}_t^{k+1} = \mathbf{A}_t^k \left(\Phi \left(\lambda_2, \mathbf{A}_t^k, G_t \left(\mathbf{y} - \widehat{\mathbf{y}}_c^k \right) \right) \right) \qquad (6)$$

where Φ is a sparse coding operator, which could be any least absolute shrinkage and selection operator (LASSO)-type [42] solver. G_c and G_t are the methodologies (i.e., *Curvelet* wavelet transforms and GWTs) used for obtaining the content and texture components, respectively.

Let $U(a) = a \times a$ be a given window size and p be the number of partitions. Then, we can obtain the initial dictionary $\mathbf{A} \equiv \mathrm{DU}(U(a), p)$. Let $\{\theta_d^c, \theta_d^t\}$ be the hard thresholds for penalizing the atoms in the content and texture dictionaries, respectively. Let $\{\theta_c^c, \theta_c^t\}$ be the hard thresholds to regularize the sparse coefficients of the content and texture components, respectively. Let γ_c be the parameter controlling the degree of smoothness of the TV correction. With these definitions in place, we first introduce Algorithm 1 for dictionary update. As shown in line 4 of Algorithm 1, the coefficients derived from G are penalized via hard threshold. Then, in line 5, the dictionary is reconstructed by the penalized coefficients.

Algorithm 1 Dictionary update: DU.

1: **Input**: Image matrix $\mathbf{y}$, initial dictionary $\mathbf{A}$, transform function G, and threshold θ_d.
2: **Output**: The updated dictionary $\mathbf{A}'$
3: Compute the coefficients: Coef $\leftarrow G(\mathbf{y}, \mathbf{A})$
4: Update the coefficients via hard threshold:
    Coef $\leftarrow$ hard(Coef, θ_d)
5: Reconstruct the dictionary: $\mathbf{A}' \leftarrow G^{-1}(\text{Coef}, \mathbf{y})$

In the last part, the proposed MCA-based algorithm (see Algorithm 2) is introduced, which has been coined as

MCASUnSAL for short since we integrate MCA with the *SUnSAL* technique. As shown in Algorithm 2, lines 7–10 perform the reconstruction of the texture component, and lines 12–16 solve the problem with respect to the content component. More specifically, in lines 7 and 12, the dictionaries are updated by Algorithm 1. In lines 8 and 13, the sparse coefficients are estimated via a sparse coding operator, where *SUnSAL* is selected to be used due to its efficiency for solving ℓ_1 optimization problems.[2] Then, in lines 9 and 14, sparse coefficients are penalized by using hard threshold. Finally, in lines 10 and 15, the content and texture components are reconstructed. In order to improve the spatial smoothness, line 16 performs a TV regularization for the content component.

Unlike the traditional MCA method, which adopts the TV method on the whole image and has large computational burden [35], the proposed approach penalizes both the wavelet coefficients and the sparse coefficients derived from the image partitions to correct the output image in a very fast way, which allows more representative reconstruction of the content and texture components. In this paper, we only perform TV regularization for the content component as it contains the major information of the input image. However, this also can been done for the texture component if needed. The algorithm terminates when the iteration meets an end or the standard deviation $[\varepsilon \equiv \sigma(\mathbf{y} - \mathbf{y}_c - \mathbf{y}_t)]$ calculated from the remainder component remains stable with regard to a user-defined threshold ($\varepsilon_{\max}$). In the final step, we adopt SVM to obtain the final classification maps based on the extracted MCs.

Algorithm 2 *MCASUnSAL* for image separation

1: **Input**: $\mathbf{y}$, $\{\lambda_1, \lambda_2\}$, $U(a)$, p, $\{\theta_d^c, \theta_d^t\}$, $\{\theta_c^c, \theta_c^t\}$, γ_c, stopping threshold $\varepsilon_{\max}$, maximum iteration K, $\mathbf{A}_c^0$, $\mathbf{A}_t^0$
2: **Output**: $\mathbf{y}_c$, $\mathbf{y}_t$
3: **Initialization**: $\widehat{\mathbf{y}}_c^0 = \mathbf{0}$, $\widehat{\mathbf{y}}_t^0 = \mathbf{0}$, $k = 0$, $\varepsilon^0 = 0$, $\varepsilon^1 = \varepsilon_{\max}$
4: **while** $k < K$ **or** $\varepsilon^k - \varepsilon^{k-1} \geq \varepsilon_{\max}$ **do**
5:     $k \leftarrow k + 1$
6:     $-----$ For texture component $-----$
7:     Update dictionary:
      $\mathbf{A}_t^k \leftarrow \mathrm{DU}(G_t(\mathbf{y}, \mathbf{y}_t^{k-1}, \mathbf{A}_t^{k-1}), U(a), p, \theta_d^t)$
8:     Apply sparse coding: $\widehat{\mathbf{x}}_t^k \leftarrow \Phi(\mathbf{A}_t^k, \mathbf{y}_t^{k-1}, \mathbf{y}_c^{k-1}, G_t, \lambda_2)$
9:     Update the coefficients: $\widehat{\mathbf{x}}_t^k \leftarrow hard(\widehat{\mathbf{x}}_t^k, \theta_c^t)$
10:    Reconstruct the texture component: $\widehat{\mathbf{y}}_t^k = \mathbf{A}_t^k \cdot \widehat{\mathbf{x}}_t^k$
11:    $-----$ For content component $-----$
12:    Update dictionary:
      $\mathbf{A}_c^k \leftarrow \mathrm{DU}(G_c(\mathbf{y}, \mathbf{y}_c^{k-1}, \mathbf{A}_c^{k-1}), U(a), p, \theta_d^c)$
13:    Apply sparse coding: $\widehat{\mathbf{x}}_c^k \leftarrow \Phi(\mathbf{A}_c^k, \mathbf{y}_c^{k-1}, \mathbf{y}_t^{k-1}, G_c, \lambda_1)$
14:    Update the coefficients: $\widehat{\mathbf{x}}_c^k \leftarrow hard(\widehat{\mathbf{x}}_c^k, \theta_c^c)$
15:    Reconstruct the content component: $\widehat{\mathbf{y}}_c^k = \mathbf{A}_c^k \cdot \widehat{\mathbf{x}}_c^k$
16:    Apply TV correction: $\widehat{\mathbf{y}}_c^k \leftarrow TV(\widehat{\mathbf{y}}_c^k, \gamma_c)$
17:    Compute the error: $\varepsilon^k = \sigma(\mathbf{y} - \widehat{\mathbf{y}}_c^k - \widehat{\mathbf{y}}_t^k)$
18: **end while**

---

[2] *SUnSAL* solves the ℓ_1-norm optimization problem, which is more efficient and less complex with equal accurate solution as an alternate of LASSO, etc.
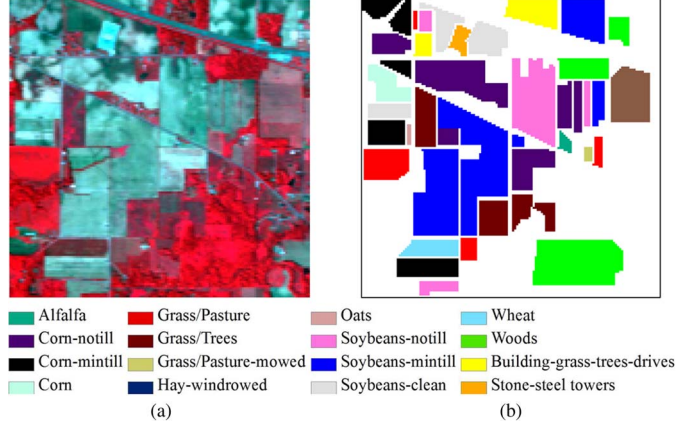

Fig. 2. (a) False color composition of the AVIRIS Indian Pines scene (R: 57, G: 27, B: 17). (b) Ground-truth map containing 16 mutually exclusive land-cover classes (bottom).

III. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we evaluate the proposed scheme by using two real airborne hyperspectral data sets, respectively, collected by the AVIRIS and ROSIS instruments. Before describing our experiments, we first introduce the parameter settings and notations adopted in our experiments.

- For the considered classifiers, superscripts "raw", "*c*", "(*c* + *t*)", and "(*c*, *t*)" denote the classifiers applied to the "original data or the feature obtained by using MNF (or PCA)", "content component", "the direct summation of content and texture components", and "the stacked content and texture components", respectively.

- For the dimensionality reduction, although we propose MNF as the priority, PCA is also performed to infer the possibility of using other method, as well as for comparison. We retain a number of features that contain 99.9% of the spectral information of the original data sets. As aforementioned, it is possible to apply MCA-based separation on the whole image. However, since hyperspectral data generally live in a much lower dimensional space than the original spectral space, the use of transformations such as MNF (or PCA) can greatly reduce the computational complexity.

- For the algorithmic parameters involved with the *curvelet* and *Gabor* transforms, we empirically select their values. For the *curvelet* transform, the coarsest scale is set as 1. For the *Gabor* transform, the frequency is set as 1/4, and we adopt three scale levels with four orientations to build the *Gabor* filter bank. Although this strategy might be suboptimal, it has been observed to produce good results in practice. Another important reason is that all methods share the same parameter settings, which ensures a fair comparison.

- For the dictionary construction, we set $U(a) = 8 \times 8$ and $p = 10$, which means that there are 640 atoms for each dictionary. We select a relatively large number to cover relatively more representative regions within the whole image without increasing the computational complexity too much.

TABLE I

OAs, AAs, AND INDIVIDUAL CLASS ACCURACY LEVELS (%), KAPPA STATISTICS κ ALONG WITH THE STANDARD DEVIATION OF
FIVE CONDUCTED MONTE CARLO RUNS, AND COMPUTATIONAL TIME IN SECONDS OBTAINED FOR BAGGING AND SVM BASED
ON DIFFERENT MCs OF INDIAN PINES DATA SET WITH A BALANCED TRAINING SET (10% OF THE LABELED SAMPLES
USED FOR TRAINING FOR A TOTAL OF 1036 SAMPLES AND THE REMAINING LABELED SAMPLES USED FOR TESTING)

Class	#Samples		Classification methods							
			MNF-based							
	Train	Test	bagging^{raw}	bagging^{c}	bagging$^{(c+t)}$	bagging$^{(c,t)}$	SVM^{raw}	SVM^{c}	SVM$^{(c+t)}$	SVM$^{(c,t)}$
Alfalfa	5	49	81.67±5.43	**90.78±4.69**	89.72±3.38	85.44±6.07	75.00±0.00	89.81±5.79	84.38±4.85	87.46±12.57
Corn-no till	143	1291	84.13±0.64	95.05±0.56	92.26±0.35	91.88±0.89	87.97±0.00	**95.16±0.79**	92.78±2.17	91.57±0.43
Corn-min till	83	751	72.44±1.59	88.94±1.43	85.32±0.96	84.23±1.75	79.52±0.00	**92.79±0.70**	89.61±2.36	89.26±1.32
Corn	23	211	61.69±5.04	87.39±5.72	77.12±2.88	81.70±6.13	71.60±0.00	**92.56±3.20**	85.52±2.04	85.91±1.72
Grass/Pasture	50	447	94.07±0.68	**97.31±0.73**	96.11±0.56	96.90±0.61	94.23±0.00	96.63±0.33	96.06±0.97	96.30±0.97
Grass/Trees	75	672	95.35±0.56	99.06±0.45	98.24±0.32	99.08±0.38	98.71±0.00	**99.89±0.05**	99.66±0.05	99.51±0.06
Grass/Pasture-mowed	3	23	79.33±9.45	86.21±10.68	**98.50±1.91**	91.17±6.57	88.00±0.00	98.00±0.00	90.00±11.31	89.00±12.73
Hay-windrowed	49	440	97.40±0.93	98.47±0.63	99.09±0.33	98.65±0.58	97.68±0.00	**99.15±0.55**	98.59±0.54	98.94±1.18
Oats	2	18	73.05±15.96	68.13±16.78	69.23±13.45	72.34±13.64	75.00±0.00	**75.00±7.86**	58.33±0.00	58.33±3.93
Soybeans-no till	97	871	78.56±0.91	**93.76±1.25**	88.72±1.46	90.62±0.83	83.23±0.00	92.15±1.40	87.78±1.06	89.39±1.40
Soybeans-min till	247	2221	82.62±0.78	94.38±0.52	91.18±0.36	90.17±0.67	86.55±0.00	**95.46±0.07**	92.58±0.20	91.58±0.23
Soybean-clean till	61	553	86.38±0.77	94.34±0.75	93.19±0.62	91.81±1.49	89.81±0.00	**96.39±1.12**	95.17±0.49	93.54±1.09
Wheat	21	191	97.52±0.28	**99.57±0.56**	97.65±1.59	98.73±0.83	99.73±0.00	99.32±0.19	99.46±0.38	99.19±0.00
Woods	129	1165	95.07±0.26	97.53±0.43	96.90±0.43	97.10±0.41	95.48±0.00	**98.13±0.33**	96.44±0.73	96.75±0.97
Bldg-grass-tree-drives	38	342	78.63±2.67	91.21±1.42	86.95±2.49	89.81±1.27	84.47±0.00	**93.75±1.19**	87.22±3.14	89.42±2.74
Stone-steel towers	10	85	91.97±1.52	**97.12±1.55**	91.29±0.97	93.16±3.15	90.85±0.00	92.51±0.49	91.27±0.28	93.05±0.57
Average accuracy	-	-	84.37±1.58	92.45±1.42	90.72±1.00	90.80±1.15	87.36±0.00	**94.17±0.69**	90.30±0.52	90.57±2.23
Overall accuracy	-	-	85.20±0.61	94.76±0.35	92.14±0.35	92.17±0.45	88.70±0.00	**95.64±0.21**	93.10±0.45	92.88±0.29
k statistic	-	-	0.830±0.007	0.940±0.004	0.910±0.004	0.910±0.005	0.871±0.000	**0.950±0.002**	0.921±0.005	0.919±0.003
Time(Seconds)	-	-	**2.21**	164.8+1.79	164.8+1.88	164.8+1.87	13.69	164.8+12.24	164.8+11.96	164.8+11.98
			PCA-based							
Average accuracy	-	-	75.75±1.20	77.79±1.26	77.45±1.53	72.15±1.56	80.03±0.57	81.52±0.00	**81.69±1.13**	73.57±3.04
Overall accuracy	-	-	77.55±0.51	79.10±0.32	78.69±0.72	73.48±0.49	82.01±0.91	83.02±0.00	**83.24±0.71**	75.89±0.30
k statistic	-	-	0.742±0.006	0.760±0.004	0.755±0.008	0.694±0.006	0.794±0.011	0.805±0.000	**0.808±0.008**	0.722±0.005
Time(Seconds)	-	-	**2.44**	135.6+2.33	135.6+2.25	135.6+2.35	14.71	135.6+14.65	135.6+15.88	135.6+15.11
			Origin-based							
Average accuracy	-	-	71.26±1.42	69.65±2.04	70.46±1.82	70.20±1.67	85.88±0.69	87.07±1.20	**87.67±1.23**	79.88±0.37
Overall accuracy	-	-	76.69±0.92	75.19±0.70	76.58±0.47	74.54±0.68	87.78±0.48	88.37±0.62	**88.84±0.38**	83.03±0.72
k statistic	-	-	0.732±0.011	0.715±0.008	0.731±0.005	0.708±0.008	0.860±0.005	0.867±0.007	**0.872±0.004**	0.806±0.008
Time(Seconds)	-	-	**3.62**	3054+3.76	3054+3.60	3054+5.75	261.68	3054+235.29	3054+222.66	3054+708.69

- Concerning the classifiers, we optimize the involved parameters of SVM by using the particle swarm optimization [43]. In order to evaluate the classification performance of the proposed scheme, we also include bagging [44] and implement other recently proposed spectral–spatial classifiers: MLRsubMLL [20], MPM-LBP [21], and LORSAL-MLL [23]. For bagging, we empirically set the ensemble size as 20, which was experimentally observed to produce good results and less computational complexity in our experiments. As for the considered spectral–spatial classification methods, we adopted the recommended parameter settings in their corresponding work.

- The training set used for experiments is randomly selected from the ground-truth image. The reported overall accuracy levels (OAs), average accuracy levels (AAs), kappa statistics κ and class individual accuracy levels are derived from averaging the results obtained after conducting five independent Monte Carlo runs with respect to the initial training set.

- Finally, we would like to emphasize that all the experiments were conducted using MATLAB R2012b in a desktop PC equipped with an Intel Core i7 CPU (at 3.4 GHz) and 16 GB of RAM.

The remainder of this section is organized as follows. In Section III-A, we introduce the data sets used for evaluation, which composite the AVIRIS Indian Pines and the ROSIS Pavia University data sets (two of the most widely used benchmarks for hyperspectral image classification). Section III-B describes the experiments with the AVIRIS Indian Pines data set. Section III-C reports the experimental results with the ROSIS University of Pavia data set.
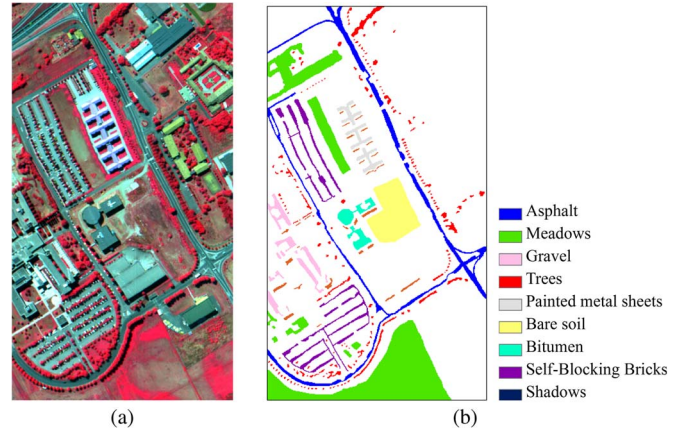

Fig. 3. (a) False color composition of the ROSIS University of Pavia scene (R: 102, G: 56, B: 31). (b) Ground-truth map containing nine mutually exclusive land-cover classes (right).

A. Hyperspectral Data Sets

Two hyperspectral data sets collected by different instruments are used in our experiments. The two data sets are available online:[3]

- The first hyperspectral image used in the experiments was recorded by the AVIRIS sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene, with a size of 145 by 145 pixels, was acquired over a mixed agricultural/forest area, early in the growing season. The scene comprises 220 spectral bands in the wavelength range from 0.4 to 2.5 μm, nominal spectral resolution
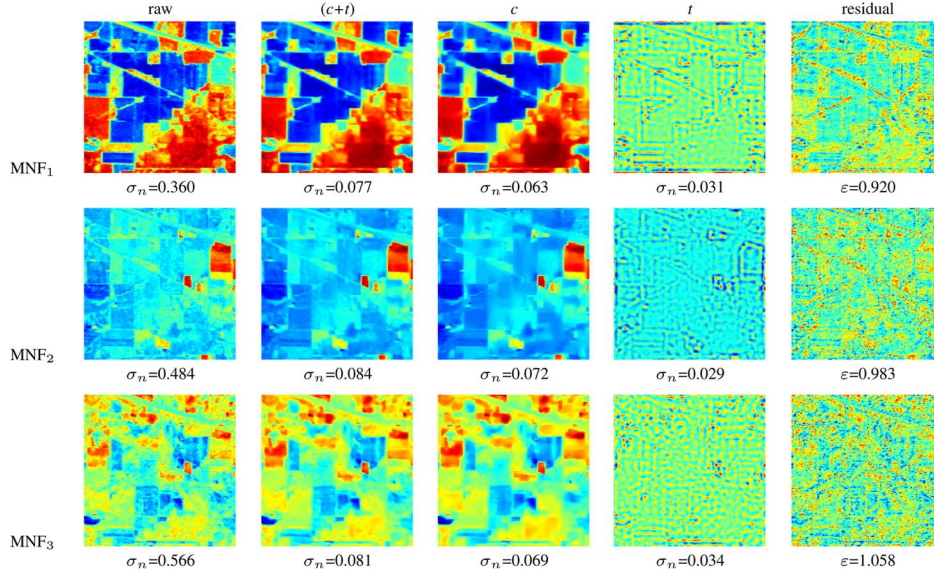
[3] Available online: http://www.ehu.es/ccwintco/index.php.

Fig. 4. Image separation results along with the noise variance σ_n and standard deviation ε for the first three MNF features ($\text{MNF}_1 - \text{MNF}_3$) of the AVIRIS Indian Pines data set.

of 10 nm, moderate spatial resolution of 20 m by pixel, and 16-b radiometric resolution. After an initial screening, several spectral bands were removed from the data set due to noise and water absorption phenomena, leaving a total of 200 radiance channels to be used in the experiments. A three-band false color image and the ground-truth image are shown in Fig. 2. A total of 10 366 samples containing 16 classes are available, which are detailed in Table I. This scene constitutes a very challenging classification problem due to the significant presence of mixed pixels and unbalanced labeled classes.

- The second hyperspectral data set was acquired by the ROSIS sensor over the urban area of the University of Pavia, Italy. The image size in pixels is 610×340, with very high spatial resolution of 1.3 m/pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). The ground-truth data contain nine classes of interest with a total of 42 776 labeled samples (see Table III). This training set is widely used in the hyperspectral image classification community and is provided by the University of Pavia, who conducted the ground-truth data collection and labeled sample generation for this particular scene. A three-band false color image and the ground-truth image are shown in Fig. 3.

B. Experiments With AVIRIS Indian Pines Data Set

Experiment 1: In this set of experiments, we first estimate the quality of image separation by adopting the proposed approach. Naturally, the peak signal-to-noise ratio (PSNR) could be a prior choice, which is effective in assessing the quality of noisy images [45]. However, we cannot obtain the pure image as an input to calculate the true PSNR. As an alternate, we estimate the noise level of MCs by adopting a fast noise variance estimation algorithm [46], which designs a zero-mean *Laplacian* operator and calculates the variance of its output as the noise variance with the form $\sigma_n \equiv \sqrt{\pi/2}(1/6(W-2)$

$(H-2))\sum\text{conv}(\mathbf{I}, \mathbf{M})$. Here, W and H represent the width and height of the input image $\mathbf{I}$, respectively. Function conv conducts the convolution of $\mathbf{I}$ with the designed *Laplacian* template $\mathbf{M} = [1\ -2\ 1; -2\ 4\ -2; 1\ -2\ 1]$.

Fig. 4 shows the values of σ_n and ε for the MCs obtained from the first three MNF features. It can be observed that the noise level of the input image is greatly reduced by the proposed image separation method. For instance, the value of σ_n is 0.36 for MNF_1, which is reduced to 0.077 and 0.063 for the summation component $(c+t)$ and the content component (c), respectively. Similar observations can be made for the second and third MNF features. Another observation is that the content and texture components are well separated and represented by the proposed approach, as shown in Fig. 4. Moreover, the obtained content components are much more smooth and have a high level of homogeneity, particularly for the classes of *corn-notill, soybeans-notill, and soybeans-mintill* in this scene.

These observations reveal that the content and texture components can maintain the major spatial information, which also hint at the potential performance improvement in terms of classification accuracy by using the proposed approach.

Experiment 2: In the second set of experiments, we evaluate the performance of the proposed approach under balanced training samples per class for the AVIRIS Indian Pines data set. Around 10% of the labeled samples per class (a total of 1036 samples) are randomly chosen for training, and the remaining 9330 labeled samples are used for testing. The number of input MNF features is set as $q = 15$. Table I reports the OAs, AAs, individual classification accuracy levels, and κ statistic along with the computational time in seconds obtained for different classification methods when applied to this scene. The table also reports the standard deviation of the five conducted Monte Carlo runs that were averaged in order to generate the classification scores reported in it. In the table, we include the results obtained by the SVM and bagging classifiers with different ingredients of MCs as their inputs. The classification results obtained from PCA features and the original full-band

TABLE II
OAs, AAs, AND INDIVIDUAL CLASS ACCURACY LEVELS (%), KAPPA STATISTICS κ ALONG WITH THE STANDARD DEVIATION OF
FIVE CONDUCTED MONTE CARLO RUNS, AND COMPUTATIONAL TIME IN SECONDS OBTAINED FOR DIFFERENT SPECTRAL–SPATIAL
CLASSIFICATION METHODS FOR THE INDIAN PINES DATA SET WITH A BALANCED TRAINING SET (10% OF THE LABELED SAMPLES
USED FOR TRAINING FOR A TOTAL OF 1036 SAMPLES AND THE REMAINING LABELED SAMPLES USED FOR TESTING)

Class	#Samples		Classification methods			
	Train	Test	SVM§	M^2 S^3 VM†	LORSAL-MLL‡	MPM-LBP$^{\pounds}$
Alfalfa	5	49	81.74±4.07	**89.81±5.79**	67.14±13.38	75.51±12.73
Corn-no till	143	1291	85.78±0.54	**95.16±0.79**	93.65±2.29	93.30±2.35
Corn-min till	83	751	82.53±1.25	**92.79±0.70**	88.12±5.41	88.48±4.65
Corn	23	211	72.26±3.10	92.56±3.20	93.22±8.62	**95.26±2.77**
Grass/Pasture	50	447	93.02±1.05	**96.63±0.33**	94.14±4.42	95.01±2.66
Grass/Trees	75	672	95.68±1.07	**99.89±0.05**	99.14±1.04	98.93±0.85
Grass/Pasture-mowed	3	23	86.20±2.86	**98.00±0.00**	55.65±48.28	81.74±12.93
Hay-windrowed	49	440	97.20±0.63	99.15±0.55	99.52±0.23	**99.57±0.27**
Oats	2	18	64.60±3.99	**75.00±7.86**	11.67±24.77	68.33±26.71
Soybeans-no till	97	871	83.07±0.65	92.15±1.40	89.61±4.31	**93.35±1.90**
Soybeans-min till	247	2221	87.16±0.47	95.46±0.07	**97.12±1.17**	96.13±1.49
Soybean-clean till	61	553	85.88±1.32	**96.39±1.12**	96.47±1.34	95.59±2.04
Wheat	21	191	96.25±0.48	99.32±0.19	**99.53±0.17**	99.37±0.33
Woods	129	1165	95.05±0.14	**98.13±0.33**	97.15±2.13	97.59±1.48
Bldg-grass-tree-drives	38	342	73.93±1.80	**93.75±1.19**	88.19±4.51	85.23±4.74
Stone-steel towers	10	85	**93.66±3.20**	92.51±0.49	76.94±14.22	78.24±9.32
Average accuracy	-	-	85.88±0.69	**94.17±0.69**	84.20±3.47	90.10±2.23
Overall accuracy	-	-	87.78±0.48	**95.64±0.21**	94.32±0.91	94.61±0.66
k statistic	-	-	0.860±0.005	**0.950±0.002**	0.935±0.010	0.939±0.008
Time(Seconds)	-	-	261.68	177.04	**4.08**	111.85

§ SVM results are obtained based on the original full-band image;
† M^{2}S^{3}VM algorithm is the proposed spectral-spatial classification method;
‡ LORSAL-MLL algorithm implements the maximum *a posteriori* (MAP) segmentation by using sparse multinomial logistic regression
(MLR) via variable splitting and augmented Lagrangian (LORSAL)and a multilevel logistic (MLL) prior in [23];
$^{\pounds}$ MPM-LBP algorithm maximizes the posterior marginal by using the loopy belief propagation (LBP) in [21].

image are also included for reference. We have highlighted the best results (the highest accuracy or the lowest computational time) with bold typeface in each case.

Several conclusions can be made from Table I. First of all, it is remarkable that the MCs extracted from the MNF features by using the proposed *MCASUnSAL* method can greatly improve the accuracy levels. For instance, for the MNF-based content component, SVM^c obtained OA and AA of 95.64% and 94.17%, respectively, which are around 7% higher than those obtained directly from the MNF features. Similar observation can be made for the bagging-based classification results. In fact, the proposed approach is more accurate when using SVM for classification.

Furthermore, it is noticeable that, when considering the dimensionality reduction methods, MNF leads to the best classification accuracy levels in this experiment in comparison with PCA. This observation matches the expectation that, as a dimensionality reduction method, MNF is very powerful for jointly preserving the spectral information and reducing noise. For instance, the highest OA based on the PCA features is 83.24%, which is obtained by SVM$^{(c+t)}$ in this category. Similar observation can be made for the original full-band image based classification methods. Although the improvements of classification accuracy are not so significant for PCA and the original full-band image based methods, the results obtained based on the extracted MCs are indeed the best. We now can confirm that the obtained results involving the spatial information (MCs) are much better in terms of accuracy than those obtained only with the spectral information.

Finally, to illustrate the computational complexity of the presented method, we also reported the computational time in this experiment. The computational time consists of two independent parts: the MCs' extraction time and the classification time, among which the first part tells the computational complexity indirectly. As shown in the table, it takes around 164 s to complete the separation for MNF-based features, and the time is around 135 s for the PCA-based features. However, it takes a much longer time for the original image, which is around 3054 s. To sum up, the computational complexity is in a relatively low level.

The experiment demonstrates that image-separation-induced component extraction contributes to the improvement of classification performance. This is consistent with the former observations resulted from Fig. 4, i.e., the content and texture components have decreased the noise level and preserved the major spatial information. Since the proposed image separation approach *MCASUnSAL* obtained the best classification performance based on MNF features and the SVM classifier, we denote by M^{2}S^{3}VM the proposed classification scheme for simplicity.

Experiment 3: In the third experiment, we compare the proposed classification scheme M^{2}S^{3}VM with several recently proposed spectral–spatial classification methods in the literature to illustrate the performance of our method. We recall that around 10% of the labeled samples per class have been used for training (a total of 1036 samples), and the remaining labeled samples are used for testing. We again set the dimensionality of the input MNF features as $q = 15$. Table II reports the OAs,
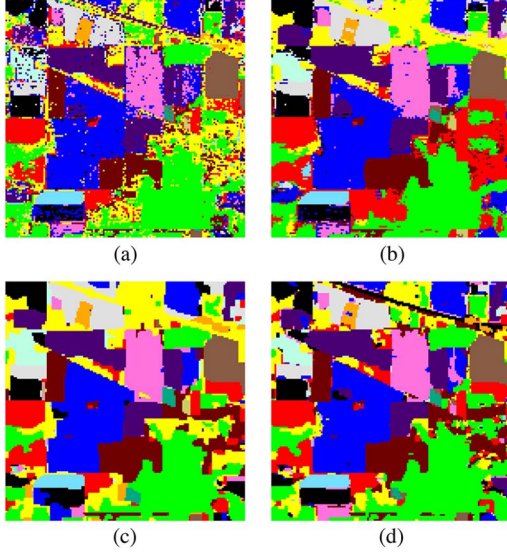
Fig. 5. Classification maps obtained by different spectral–spatial classification methods for the AVIRIS Indian Pines data set (the OAs are shown in the parentheses for each case). (a) SVM (87.78%). (b) M^2S^3VM (95.64%). (c) LORSAL-MLL (94.32%). (d) MPM-LBP (94.61%).

AAs, and individual classificaiton accuracy levels (%) and the κ statistic obtained for different spectral–spatial classification methods when applied to this scene. In the table, we also include the results obtained by the SVM classifier with the original full-band image as the baseline. The table also includes the processing time in seconds. As shown in Table II, the higher classification accuracy levels are always obtained by using the spectral–spatial classification methods, among which the proposed method M^2S^3VM provides the best performance for this scene. Precisely, the OA obtained by the proposed method is around 95.64%, which is 8% higher than the baseline obtained by the SVM when applied to spectral information. However, MPM-LBP produces competitive performance, which is 94.61% in terms of OA. Another observation is that LORSAL-MLL provides the lowest computational time. Although the computational complexity of the proposed method is a bit high, it is still within a reasonable level that can be bearable when applying our method for classification purpose. For illustrative purposes, Fig. 5 shows the classification maps corresponding to one of the Monte Carlo runs conducted for the AVIRIS Indian Pines data set using the balanced training set in Table II. The advantages obtained by adopting the proposed image separation and classification scheme with regard to other considered methods can be visually appreciated in the maps displayed in Fig. 5, which also reports the OAs obtained for each case in the parentheses.

Experiment 4: In order to show the performance of the proposed approach under different training conditions and scenarios, in the fourth experiment, we evaluate the classification accuracy levels with regard to different sizes of the initial training set and different numbers of dimensionality of the input MNF features. Let L_k be the total number of labeled samples in the ground-truth map for a single class k, and let l_k be the number of selected training samples per class. Here, if $l_k > L_k$ (i.e., the number of labeled samples per class is very small), we take $l_k = L_k/2$.
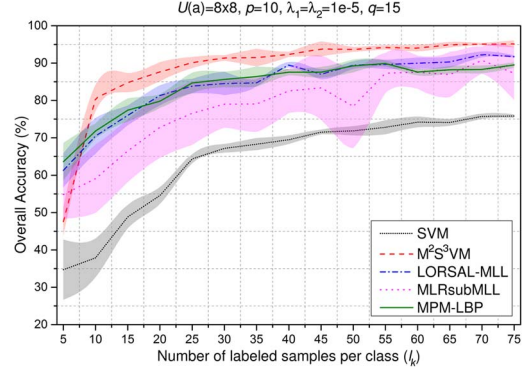

Fig. 6. OAs and standard deviation (colored area along each line) (as a function of the number of labeled samples per class) achieved by different spectral–spatial classification methods for the AVIRIS Indian Pines data set.
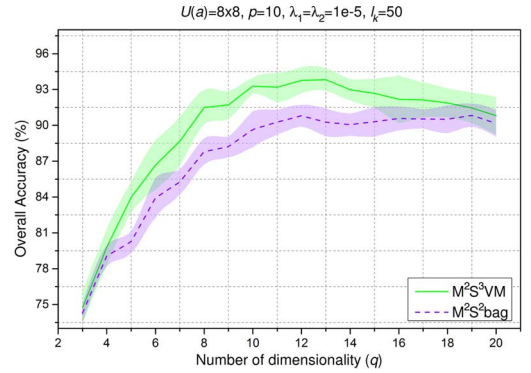

Fig. 7. OAs and standard deviation (colored area along each line) (as a function of the number of dimensionality of MNF features) achieved by M^2S^3VM (top) (our method) and M^2S^2bag (bottom) (using bagging for classification in the proposed scheme) for the AVIRIS Indian Pines data set.

First, we fix the input MNF feature number as $q = 15$ while evaluating the impacts of different values of l_k on the OAs. Fig. 6 illustrates the evolution of OAs with different values of l_k for all the considered methods. As shown in the figure, the proposed approach outperforms other considered methods when $l_k > 10$, which illustrates that the proposed method has good generalization performance under a very small training set. For instance, the OA obtained by our method is around 80% with a total of 160 training samples ($l_k = 10$). Another observation is that all methods converge to a stable value when $l_k > 60$. Moreover, the considered methods are distinct from each other, except LORSAL-MLL and MPM-LBP, which produce similar performance under different values of l_k. MLRsubMLL[4] shows lower OA than these two methods. Furthermore, the classification methods involving the spatial information are always better than the method that only uses spectral information. Second, we fix the value of $l_k = 50$ while evaluating the impacts of different values of q on the OAs. Fig. 7 illustrates the evolution of OAs with different values of q for the proposed method. We also include the bagging classifier for comparison in the proposed image separation and classification scheme. We then denote by M^2S^2bag this involved method. As depicted in Fig. 7, the proposed method always achieves better classification accuracy levels under different numbers of dimensionality. The peak

[4]The MLRsubMLL [20] method combines MLR with subspace projection and MLL prior for spectral–spatial classification of hyperspectral image.

values of OA are met when q is around 12 for these two methods, and the OAs decrease slowly when $q > 13$.

Experiment 5: Let sparsity $\equiv n_0/n \times 100\%$, where n is the total number of coefficients in $\hat{\mathbf{x}}$ (i.e., $\hat{\mathbf{x}}_c$ or $\hat{\mathbf{x}}_t$), and n_0 is the number of small values, i.e., $|\hat{x}_i| \leq 1e-3$ for any i in $\hat{\mathbf{x}}$. In a final experiment conducted with the AVIRIS Indian Pines data set, we investigate the relationship between sparsity, ε, accuracy, and σ_n with parameters $\lambda_1, \lambda_2, U(a)$, and p, in which 50 samples per class are used for training, and the remaining samples are used for validation.

In Fig. 8, plots (a), (b), and (c), respectively, illustrate the sparsity (the class of oats), ε, and σ_n as a function of parameters λ_1 and λ_2 with fixed $U(a) = 8 \times 8$ and $p = 10$. It can be observed that sparsity is sensitive to λ_1 and λ_2, which decays as the joint decrease in λ_1 and λ_2, and a relatively small value of sparsity can be obtained when these two parameters are lower than $1e - 4$, which results in a better representation. Standard deviation ε seems to be insensitive to λ_2, whereas it is highly related to λ_1 when it is in the range of $1e - 4$ to 1. However, when λ_1 is lower than $1e - 4$, ε comes to a stable level, which is around 1. Different from this observation, σ_n has a complex relationship with the two parameters. However, we can roughly notice that the value of σ_n is higher when λ_1 and λ_2 are around $1e - 4$. Plots (d) and (e), respectively, show the OAs (as a function of parameters λ_1 and λ_2) obtained by $\text{M}^2\text{S}^3\text{VM}$ and $\text{M}^2\text{S}^2\text{bag}$, respectively, again with fixed U and p. For visibility purpose, the λ_1-axes of these two plots are reversed. These two plots tell the similar story: OA is sensitive to λ_1, OA comes to a stable level when λ_1 is lower than $1e - 4$, and the evolution trend of OA is opposite to the trend of ε. Plots (f), (g), and (h), respectively, illustrate the sparsity versus OA, sparsity versus ε, and sparsity versus σ_n as a function of parameters $U(a)$ and p with fixed $\lambda_1 = \lambda_2 = 1e - 5$. It is interesting to observe that the two methods are still performing well with a small size of $U(a)$ and p. Therefore, in order to reduce the computational complexity, a small dictionary size is recommended. On the contrary, sparsity is closely related to dictionary size, and a larger size of dictionary can yields higher sparsity. Standard deviation ε also has close relationship with sparsity, and higher sparsity always leads to higher ε. Similar observations can be made for σ_n. However, when $p < 4$, σ_n decreases when p increases with $U(a) = 8 \times 8$, whereas σ_n remains stable when using a smaller window size with $U(a) = 4 \times 4$.

C. Experiments With ROSIS University of Pavia Data Set

Experiment 1: In this experiment, we first evaluate the quality of image separation by applying the proposed approach to the ROSIS University of Pavia data set. Fig. 9 shows the separation results along with σ_n and ε for MCs obtained from the first three MNF features. It can be observed that the MCs always have lower σ_n compared with the raw features. For instance, the value of σ_n for MNF_1 is around 0.47, which is reduced to 0.081 and 0.0035 for the $(c + t)$ and the c components. The well separation and noise removal of the proposed approach can be appreciated in Fig. 9 by visual inspection. Again, the obtained content components are much more smooth and have a high level of homogeneity for the class of *bare soil*, *meadows*,

and *bitumen*, which preliminarily suggests the improvement of classification accuracy by using the MCs that preserving the major spatial information. This observation is matched in the classification maps shown in Fig. 10.

Experiment 2: In the second set of experiments, we evaluate the classification performance of the proposed approach based on the ROSIS University of Pavia data set, where around 5% of the labeled samples per class (a total of 2138) are used for training, and the remaining 40 638 labeled samples are used for testing. We also set the number of input MNF features as $q = 15$. Table III reports the OAs, AAs, individual classification accuracy levels, and κ statistic along with the computational time in seconds obtained for different classification methods when applied to this scene. The table also reports the standard deviation of the five conducted Monte Carlo runs. Several conclusions can be drawn from Table III. First of all, the MCs extracted from the MNF features can significantly improve the classification accuracy. For instance, for the MNF-based content component, SVM^c, obtained OA of 99.01%, which is around 8% higher than those obtained directly from the MNF features. Similar observation can be made for the bagging-based classification results. This is a high classification accuracy level for the considered scene that comprises a challenging classification scenario dominated by the complex urban features. Furthermore, MNF still outperforms PCA in this scene. However, no significant improvement can be appreciated for the original full-band image in terms of classification accuracy. Nevertheless, the MCs extracted from the original image still produce better results in the associated category.

Experiment 3: In the third experiment, we illustrate the performance of our method with regard to other spectral–spatial classification methods in the literature. It is worth noting that 5% of the labeled samples per class are used for training (a total of 2138 samples) and the remaining labeled samples are used for testing. The number of the input MNF features is also set as $q = 15$. Table IV reports the OAs, AAs, individual classificaiton accuracy levels (%), and the κ statistic obtained for different spectral–spatial classification methods when applied to this scene. As shown by Table IV, the higher classification accuracy is obtained by using the proposed method $\text{M}^2\text{S}^3\text{VM}$. Precisely, the OA obtained by the proposed method is around 99.01%, which is 5% higher than the baseline obtained by SVM when applied to spectral information. However, MPM-LBP produces competitive performance, which is 97.71% in terms of OA. Another observation is that LORSAL-MLL provides the lowest computational time in this scene. Although the computational complexity of the proposed method is a bit high, it is still within a reasonable level that can be bearable when applying our method for classification purpose. For illustrative purposes, Fig. 10 shows the classification maps corresponding to one of the Monte Carlo runs conducted for the ROSIS Univeristy of Pavia data set using the balanced training set in Table IV. The advantages obtained by adopting the proposed image separation and classification scheme with regard to other considered methods can be visually inspected in Fig. 10, which also reports the OAs obtained for each case in the parentheses.

Experiment 4: In the fourth experiment, we evaluate the classification accuracy levels with regard to different sizes
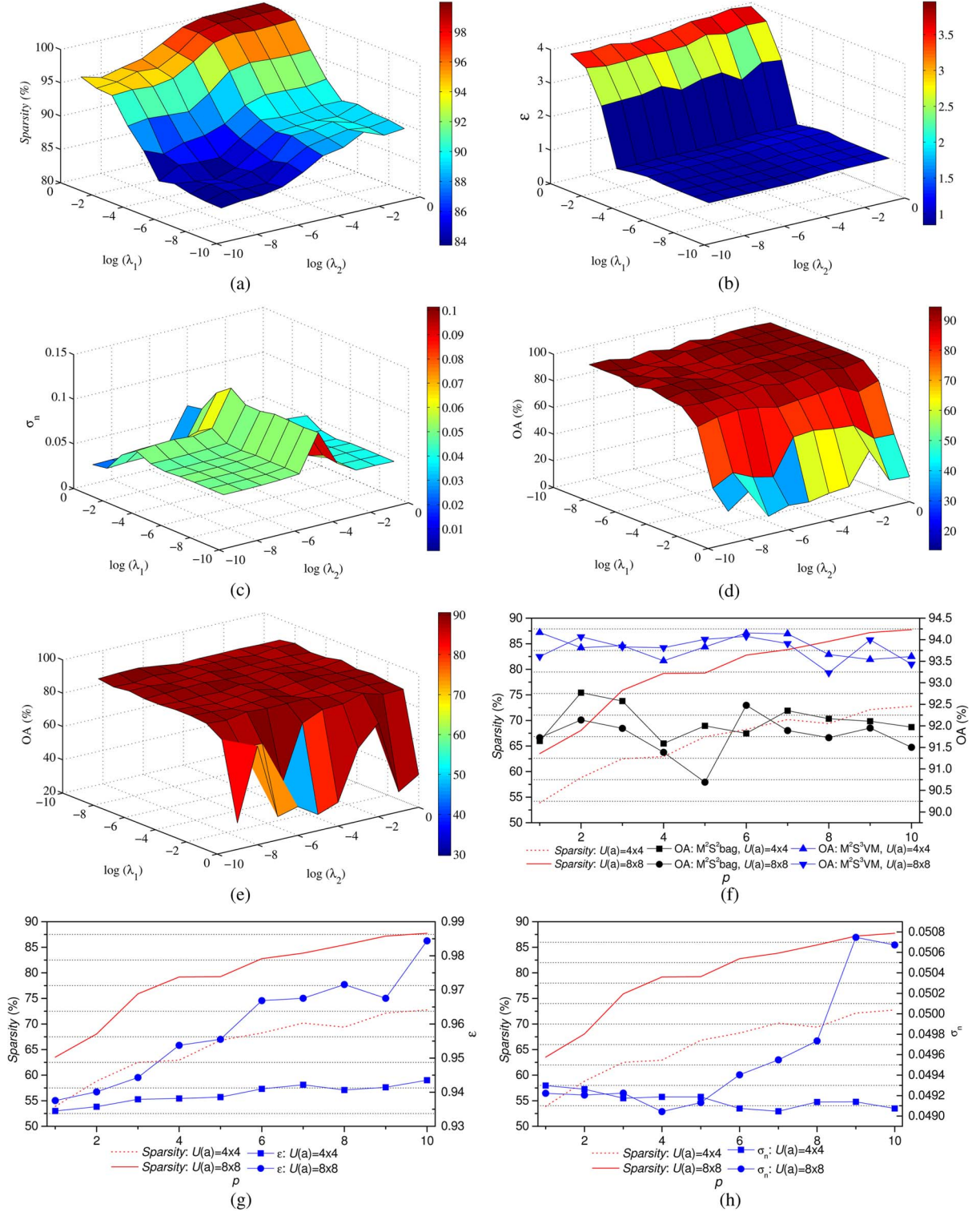
Fig. 8. Sparsity (the class of oats), standard deviation ε, noise variation σ_n, and OA as a function of parameters λ_1, λ_2, $U(a)$, and p for the AVIRIS Indian Pines scene with $l_k = 50$. Plots (a), (b), and (c), respectively, represent sparsity, ε, and σ_n (as a function of λ_1 and λ_2) obtained by the proposed image separation method with fixed $U(a) = 8 \times 8$ and $p = 10$. Plots (d) and (e) represent the OAs (as a function of λ_1 and λ_2) obtained by M^2S^3VM and M^2S^2bag, respectively. Plots (f), (g), and (h) represent the sparsity versus OA, sparsity versus ε, and sparsity versus σ_n as a function of $U(a)$ and p, respectively, with fixed $\lambda_1 = \lambda_2 = 1e - 5$. (a) $U(a) = 8 \times 8$, $p = 10$. (b) $U(a) = 8 \times 8$, $p = 10$. (c) $U(a) = 8 \times 8$, $p = 10$. (d) M^2S^3VM, $U(a) = 8 \times 8$, $p = 10$. (e) M^2S^2bag, $U(a) = 8 \times 8$, $p = 10$. (f) $\lambda_1 = 1e - 5$, $\lambda_2 = 1e - 5$. (g) $\lambda_1 = 1e - 5$, $\lambda_2 = 1e - 5$. (h) $\lambda_1 = 1e - 5$, $\lambda_2 = 1e - 5$.

of the initial training set and different dimensionality of the input MNF features to show the performance of the proposed approach under different training conditions and scenarios. First, we fix $q = 15$ while evaluating the impacts of different

values of l_k on the OAs. Fig. 11 illustrates the evolution of OAs with different values of l_k for the considered methods. As shown in the figure, the proposed approach (M^2S^3VM) outperforms other considered methods when $l_k > 15$, which
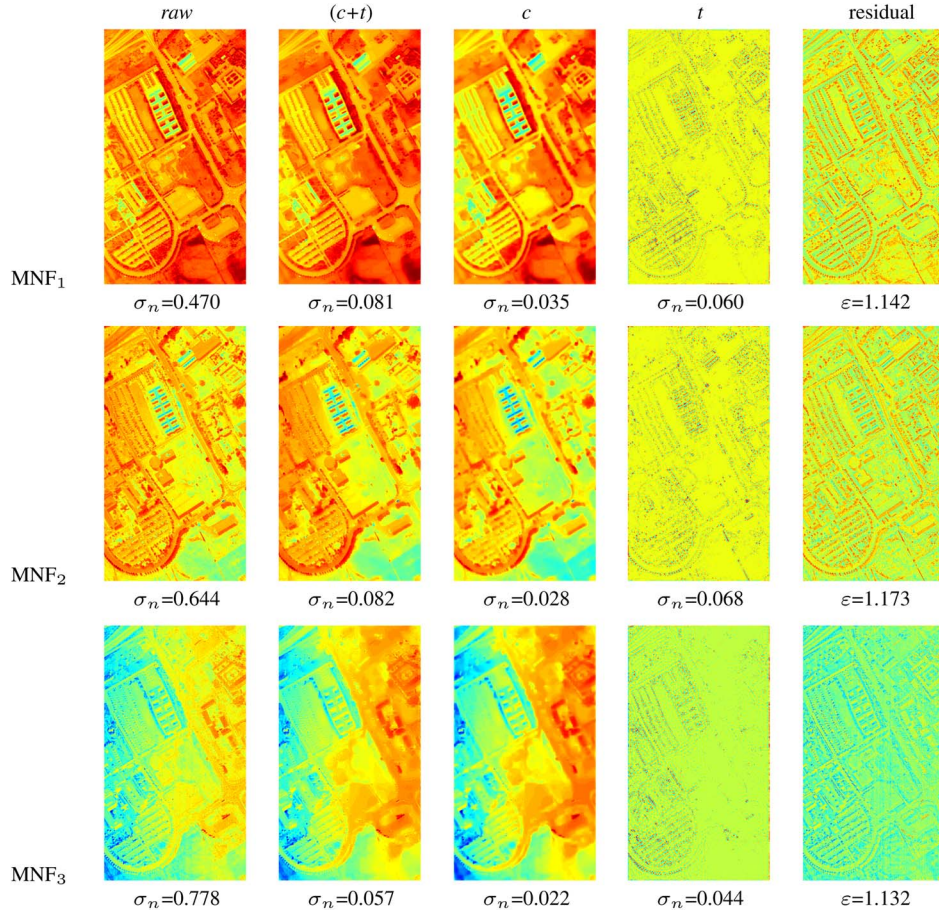
Fig. 9.   Obtained image separation results along with the noise variation σ_n and standard deviation ε for the first three MNF features (MNF$_1$$-MNF_3$) of the ROSIS University of Pavia data set.
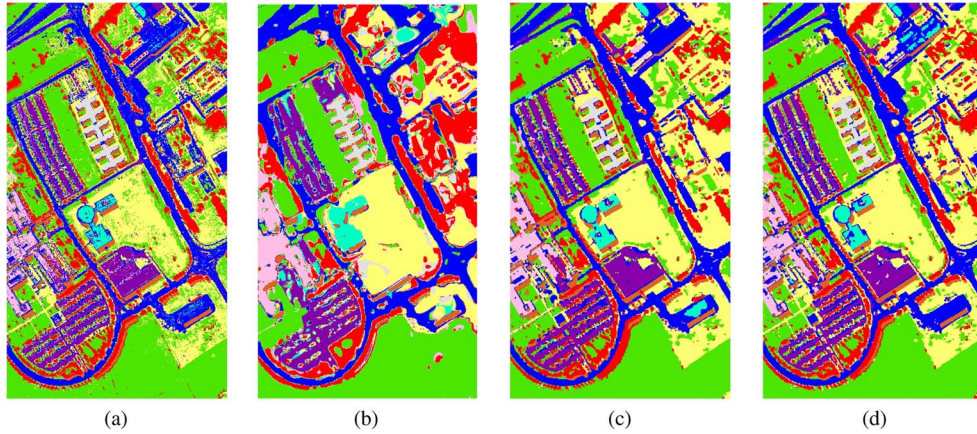

Fig. 10.   Classification maps obtained by different spectral–spatial classification methods for the ROSIS University of Pavia data set (the OAs are shown in the parentheses for each case). (a) SVM (94.33%). (b) M^{2}S^{3}VM (99.01%). (c) LORSAL-MLL (97.75%). (d) MPM-LBP (97.71%).

illustrates that the proposed method has good generalization performance under a very small training set. For instance, the OA obtained by our method is around 90% with a total of 135 training samples ($l_k = 15$). Another observation is that all methods converge slowly along with the increase in l_k. Moreover, the considered methods are distinct from each other. However, when $l_k > 45$, LORSAL-MLL and MPM-LBP produce competitive performance with regard to the proposed method. Furthermore, the classification methods involving the spatial information are always better than the method that only using spectral information. Second, we fix $l_k = 100$ while evaluating the impacts of different values of q on the OAs. Fig. 12 illustrates the evolution of OAs with different values of q for the proposed method. Again, bagging is included

TABLE III
OAs, AAs, AND INDIVIDUAL CLASS ACCURACY LEVELS (%), KAPPA STATISTICS κ ALONG WITH THE STANDARD DEVIATION OF FIVE CONDUCTED MONTE CARLO RUNS, AND COMPUTATIONAL TIME IN SECONDS OBTAINED FOR BAGGING AND SVM BASED ON DIFFERENT MCs OF UNIVERSITY OF PAVIA DATA SET WITH A BALANCED TRAINING SET (5% OF THE LABELED SAMPLES USED FOR TRAINING (A TOTAL OF 2138 SAMPLES) AND THE REMAINING LABELED SAMPLES USED FOR TESTING)

Class	#Samples		Classification methods							
			MNF-based							
	Train	Test	bagging^{raw}	bagging^{c}	bagging$^{(c+t)}$	bagging$^{(c,t)}$	SVM^{raw}	SVM^{c}	SVM$^{(c+t)}$	SVM$^{(c,t)}$
Asphalt	332	6299	91.22±0.39	**98.69±0.28**	98.00±0.18	97.29±0.18	91.60±0.66	98.51±0.33	97.17±0.27	97.19±0.22
Meadows	932	17717	95.41±0.17	99.68±0.04	99.42±0.05	99.48±0.20	95.74±0.22	**99.81±0.08**	99.68±0.05	99.68±0.12
Gravel	105	1994	75.45±0.54	95.67±2.10	92.95±2.08	90.86±1.41	75.32±1.27	**96.52±0.83**	89.92±1.99	88.49±1.54
Trees	153	2911	89.87±0.43	96.79±0.33	96.44±0.46	96.78±0.76	92.60±0.52	97.88±0.31	97.26±0.90	**98.14±0.24**
Painted metal sheets	67	1278	99.77±0.15	99.94±0.05	**99.96±0.00**	99.94±0.05	99.89±0.07	99.92±0.03	99.77±0.36	99.87±0.13
Bare soil	251	4778	88.48±0.60	**99.88±0.07**	99.76±0.10	99.46±0.36	87.61±0.52	**99.88±0.09**	99.87±0.10	99.54±0.15
Bitumen	67	1263	75.85±2.92	98.71±1.12	96.67±1.03	95.58±0.91	73.44±3.24	**99.79±0.13**	96.38±1.10	98.01±0.72
Self-Blocking Bricks	184	3498	80.91±1.30	96.55±0.84	94.31±0.84	92.52±0.48	83.51±0.79	**96.74±0.84**	91.89±0.90	91.31±0.98
Shadows	47	900	**99.26±0.29**	97.12±0.92	98.40±0.59	96.61±1.00	99.78±0.09	97.70±0.74	99.01±0.22	98.33±0.81
Average accuracy	-	-	88.47±0.39	98.11±0.37	97.32±0.47	96.50±0.43	88.83±0.73	**98.53±0.13**	96.77±0.39	96.73±0.37
Overall accuracy	-	-	90.95±0.26	98.79±0.18	98.17±0.25	97.74±0.29	91.41±0.48	**99.01±0.14**	97.87±0.24	97.81±0.24
k statistic	-	-	0.879±0.004	0.984±0.002	0.976±0.003	0.970±0.004	0.885±0.006	**0.987±0.002**	0.972±0.003	0.971±0.003
Time(Seconds)	-	-	**6.94**	2535+5.83	2535+6.10	2535+6.19	61.02	2535+33.81	2535+36.08	2535+33.91
			PCA-based							
Average accuracy	-	-	84.38±0.10	84.58±0.42	84.67±0.33	85.24±0.23	86.74±0.34	86.74±0.38	86.76±0.57	**88.57±0.75**
Overall accuracy	-	-	86.35±0.06	86.90±0.29	86.51±0.28	87.45±0.22	88.52±0.31	88.77±0.26	88.42±0.37	**90.23±0.55**
k statistic	-	-	0.816±0.001	0.823±0.004	0.818±0.004	0.831±0.003	0.846±0.004	0.849±0.003	0.844±0.005	**0.869±0.008**
Time(Seconds)	-	-	**3.71**	557+3.57	557+3.59	557+4.06	87.78	557+83.31	557+93.36	557+157.50
			Origin-based							
Average accuracy	-	-	85.81±0.40	85.50±0.25	86.01±0.19	84.73±0.60	92.76±0.28	92.61±0.44	93.16±0.16	**93.94±0.15**
Overall accuracy	-	-	87.66±0.22	87.83±0.33	87.72±0.15	86.87±0.38	94.16±0.22	94.36±0.33	94.50±0.18	**95.29±0.07**
κ statistic	-	-	0.833±0.003	0.835±0.005	0.834±0.002	0.823±0.005	0.922±0.003	0.925±0.004	0.927±0.002	**0.938±0.001**
Time(Seconds)	-	-	**43.85**	8078+78.41	8078+78.23	8078+148.70	439.74	8078+498.70	8078+553.73	8078+1124.82

TABLE IV
OAs, AAs, AND INDIVIDUAL CLASS ACCURACY LEVELS (%), KAPPA STATISTICS κ ALONG WITH THE STANDARD DEVIATION OF FIVE CONDUCTED MONTE CARLO RUNS, AND COMPUTATIONAL TIME IN SECONDS OBTAINED FOR DIFFERENT SPECTRAL–SPATIAL CLASSIFICATION METHODS FOR THE ROSIS UNIVERSITY OF PAVIA DATA SET WITH A BALANCED TRAINING SET (5% OF THE LABELED SAMPLES USED FOR TRAINING FOR A TOTAL OF 2138 SAMPLES AND THE REMAINING LABELED SAMPLES USED FOR TESTING)

Class	#Samples		Classification methods			
	Train	Test	SVM	M^{2}S^{3}VM	LORSAL-MLL	MPM-LBP
Asphalt	332	6299	94.51±0.14	98.51±0.33	**98.83±0.52**	98.13±0.73
Meadows	932	17717	97.17±0.16	99.81±0.08	**99.89±0.06**	99.74±0.17
Gravel	105	1994	82.94±0.88	**96.52±0.83**	79.18±6.29	81.57±1.94
Trees	153	2911	96.12±0.60	**97.88±0.31**	96.11±0.97	96.56±1.11
Painted metal sheets	67	1278	99.41±0.23	**99.92±0.03**	99.30±0.33	99.06±0.42
Bare soil	251	4778	91.06±0.48	**99.88±0.09**	99.53±0.27	98.23±1.01
Bitumen	67	1263	87.26±0.63	**99.79±0.13**	92.07±4.12	92.92±2.64
Self-Blocking Bricks	184	3498	88.32±0.61	96.74±0.84	95.45±0.94	**96.77±0.61**
Shadows	47	900	**99.94±0.04**	97.70±0.74	99.71±0.27	99.84±0.10
Average accuracy	-	-	92.97±0.19	**98.53±0.13**	95.56±1.15	95.87±0.41
Overall accuracy	-	-	94.33±0.19	**99.01±0.14**	97.75±0.48	97.71±0.20
k statistic	-	-	0.925±0.003	**0.987±0.002**	0.970±0.006	0.970±0.003
Time(Seconds)	-	-	437.96	2568.81	**35.75**	1084.94

here for comparison. As depicted in Fig. 12, the proposed method roughly achieves better classification accuracy levels under different numbers of dimensionality. The OAs of the two methods meet their peak values when $q = 15$.

Several conclusions can be concluded from this section.

1) The content and texture components maintain the major spatial information and can significantly improve the classification performance under complex training conditions.

2) The proposed method can yield much more homogenous and smoother classification maps.

3) Sparsity, ε, and σ_n are sensitive to λ_1, $U(a)$, and p, whereas OA is also sensitive to λ_1 but insensitive to $U(a)$ and p.

4) The stacked MCs show competitive performance compared with the direct summation or independent content component. However, texture information cannot significantly improve the results provided by the independent content component. Moreover, the additional texture information could greatly increase the computational complexity, particularly for the stacked MCs.

5) MNF outperforms PCA in terms of the classification performances derived from the corresponding extracted features. It is interesting that we found the classification methods based on the original spectral features outperform PCA-based ones, although the former ones are time-consuming. However, both the PCA-based and original-based classification methods are inferior to the MNF-based ones in all the conducted experiments. We should also emphasize that, in the respective category,
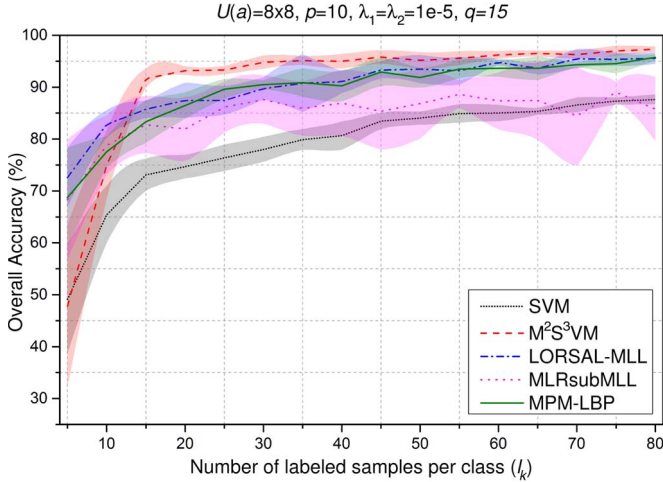
Fig. 11. OAs and standard deviation (colored area along each line) (as a function of the number of labeled samples per class) achieved by different spectral–spatial classification methods for the ROSIS University of Pavia data set.
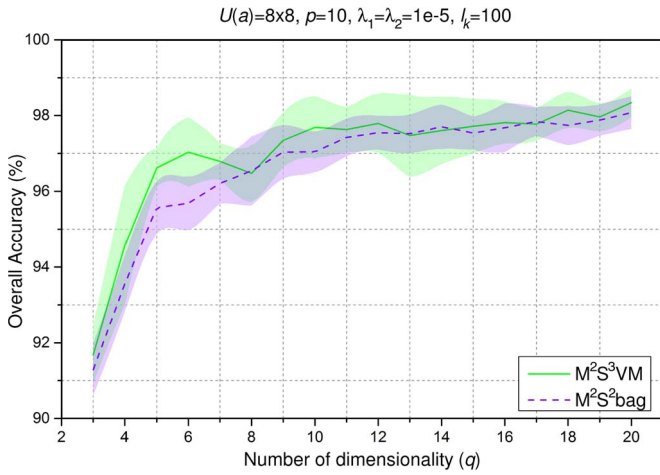

Fig. 12. OAs and standard deviation (colored area along each line) (as a function of the number of dimensionality of MNF features) achieved by M^2S^3VM (top) (our method) and M^2S^2bag (bottom) (using bagging for classification in the proposed scheme) for the ROSIS University of Pavia data set.

such as PCA-based classification methods, the content component can still yield the best results. Hence, our proposed method presents a good tradeoff between accuracy and the computational complexity.

6) The experiments provide a good comparison of different spatial resolution images: one is 20 m/pixel and the other is 1.3 m/pixel. The obtained results illustrate that the proposed approach is able to deal with images with different spatial resolutions.

7) In the proposed method, the reason that we conduct local transform to build the dictionary is twofold: first, conducting local transform can obtain a more representative dictionary since more detailed information will be extracted; second, conducting local transform can greatly decrease the computational complexity, which well confirms the efficiency of sparse representation. We admit that this significance will be soften when processing a high-resolution image. However, if we apply the same transform to the entire image to obtain a huge dictionary,

the computational complexity will be exponentially increased. Meanwhile, we cannot guarantee the equal or even better performance.

IV. CONCLUSION AND FUTURE RESEARCH LINKS

In this paper, we have developed a new method for spectral–spatial classification of the hyperspectral image based on the proposed sparse representation based MCA. The proposed approach relies on a new image separation framework, in which the *curvelet* and *Gabor* transforms are used to generate the content and texture dictionaries. Then, the *SUnSAL* algorithm is adopted for performing sparse representation based image separation. In order to reduce the computational complexity and retain the spectral information, we perform dimensionality reduction based on the MNF transform. The proposed approach for spatial and spectral feature extraction is then combined with a widely used SVM classifier. Our experimental results, obtained by two real hyperspectral images collected using AVIRIS and ROSIS indicate that the proposed approach leads to the best performance compared with other considered methods. Although the presented approach provides very good performance for the two considered scenes, further experiments with additional scenes and comparison methods should be conducted. Furthermore, we also envisage the following future perspectives for the development of this paper.

1) As the *curvelet* and *Gabor* transforms are adopted to generate content and texture dictionaries in the proposed image separation method, in future work, we will consider including other transforms to generate these two dictionaries, i.e., *biorthogonal wavelet transforms*, *isotropic à trous algorithm*, and *local ridgelet transform* [47] for content, and *local discrete cosine transform* for texture. Given the flexibility of our proposed method in sparse representation, we will explore the impact of including different dictionary learning and sparse coding approaches.

2) Since the separation model assumption in the proposed method is a simple linear superposition of two components represented by a single dictionary for each of them, we could extend this method to multiple weighted representations for content and texture based on the aforementioned transforms.

3) Another possibility for future expansion of the presented method is to obtain a segmentation map by detecting edges based on the extracted texture component in order to improve the classification performance, as described in [15].

4) Since image separation and classification are two separated steps, we will conduct further comparison with other classification methods beyond bagging and SVM.

5) Last but not the least, we will address the tough relationship between spatial resolution and classification performance [48], and the necessity or effects of considering local feature for classification with a relatively high-resolution image [49].
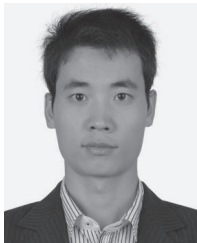
ACKNOWLEDGMENT

REFERENCES

[1] C. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.

[2] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York, NY, USA: Wiley, 2003.

[3] J. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[4] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. S110–S122, Sep. 2009.

[5] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[6] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, pp. 1–32, 2000.

[7] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2005.

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

[9] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.

[10] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[11] 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.

[12] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-14, no. 1, pp. 19–26, Jan. 1976.

[13] 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.

[14] 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.

[15] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.

[16] J. C. Tilton, Y. Tarabalka, P. M. Montesano, and E. Gofman, "Best merge region-growing segmentation with integrated nonadjacent region object aggregation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4454–4467, Nov. 2012.

[17] K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral data based on a Stochastic minimum spanning forest approach," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2008–2021, Apr. 2012.

[18] 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.

[19] S. Le Hegarat-Mascle, A. Kallel, and X. Descombes, "Ant colony optimization for image regularization based on a nonstationary Markov modeling," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 865–878, Mar. 2007.

[20] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[21] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.

[22] G. Camps-Valls, T. V. B. Marsheva, and D. Y. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Jul. 2007.

[23] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.

[24] J. Bai, S. M. Xiang, and C. H. Pan, "A graph-based classification method for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 803–817, Feb. 2013.

[25] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[26] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsso, "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.

[27] 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.

[28] S. Velasco-Forero and J. Angulo, "Classification of hyperspectral images by tensor modeling and additive morphological decomposition," *Pattern Recognit.*, vol. 46, no. 2, pp. 566–577, Feb. 2013.

[29] Y. Chong, Q. Qinfu, Z. Yilu, and C. Xiong, "Satellite image classification using morphological component analysis of texture and cartoon layers," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1109–1113, Sep. 2013.

[30] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 882–889, Aug. 2003.

[31] E. Michael, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer-Verlag, 2010.

[32] M. J. Fadili, J. L. Starck, J. Bobin, and Y. Moudden, "Image decomposition and separation using sparse representations: An overview," *Proc. IEEE*, vol. 98, no. 6, pp. 983–994, Jun. 2010.

[33] J. L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.

[34] J. Bobin, J. L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2662–2674, Nov. 2007.

[35] L. Rudin, L. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlin. Phenom.*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[36] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. 2nd WHISPERS*, 2010, pp. 1–4.

[37] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[38] J. Wang and C. I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[39] B. Zhang and L. R. Gao, *Hyperspectral Image Classification and Target Detection*. Beijing, China: Science Press, 2011.

[40] W. Y. Ma and B. S. Manjunath, "A comparison of wavelet transform features for texture image annotation," in *Proc. Int. Conf. Image Process.*, 1995, pp. I–III, pp. B256–B259.

[41] F. Malgouyres, "Minimizing the total variation under a general convex constraint for image restoration," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1450–1456, Dec. 2002.

[42] R. Tibshirani, "Regression shrinkage and selection via the LASSO: A retrospective," *J. R. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 73, no. 3, pp. 273–282, Jun. 2011.

[43] R. C. Eberhart and Y. H. Shi, "Particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 201–203, Jun. 2004.

[44] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[45] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imaging*, vol. 11, no. 2, pp. 206–223, Apr. 2002.

[46] J. Immerkaer, "Fast noise variance estimation," *Comput. Vis. Image Understand.*, vol. 64, no. 2, pp. 300–302, Sep. 1996.

[47] J. L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. New York, NY, USA: Springer-Verlag, 2002.

[48] J. R. Irons *et al.*, "The effects of spatial-resolution on the classification of thematic mapper data," *Int. J. Remote Sens.*, vol. 6, no. 8, pp. 1385–1403, Aug. 1985.

[49] W. Su *et al.*, "Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery," *Int. J. Remote Sens.*, vol. 29, no. 11, pp. 3105–3117, Jun. 2008.

Zhaohui Xue received the B.S. degree in geomatics engineering from Shandong Agriculture University, Taian, China, in 2009 and the M.E. degree in remote sensing from China University of Mining and Technology, Beijing, China, in 2012. He is currently working toward the Ph.D. degree in the Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing, China.

His research interests include hyperspectral image classification, time series analysis, pattern recognition, and machine learning.

Mr. Xue received the Outstanding Graduate Award for his B.S and M.E. degrees in 2009 and 2012, respectively. He has been honored as an "Excellent Student" from 2009 to 2012, and received several scholarships.

Jun Li (M'13) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004; the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007; and the Ph.D. degree in electrical engineering from Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotécnica e de Computadores and the Instituto de Telecomunicaes, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing (SIGNAL). She has also been actively involved in the Hyperspectral Imaging Network, a Marie Curie Research Training Network involving 15 partners in 12 countries and intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Since 2011, she has been a Postdoctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li has been a Reviewer of several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Pattern Recognition*, *Optical Engineering*, *Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*. She received the 2012 Best Reviewer Award of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Liang Cheng received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2008.

He is currently an Associate Professor with Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing, China. He is the author of over 20 papers in international journals. His research interests include LiDAR remote sensing and image understanding.

Peijun Du (M'07–SM'12) received the B.S. and Ph.D. degrees in geomatics engineering from China University of Mining and Technology, Xuzhou, China, in 1997 and 2001, respectively.

From 2002 to 2004, he was a Postdoctoral Fellow with the Laboratory of Pattern Analysis and Machines Intelligence, Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai, China. From 2006 to 2007, he was a Visiting Scholar with the University of Nottingham, Nottingham, U.K. He is currently a Professor of photogrammetry and remote sensing, and geographical information science with the Department of Geographical Information Science, Nanjing University, Nanjing, China. He is the author of nine books in Chinese and more than 100 research articles about remote sensing and geospatial information processing and applications. His research interests include remote sensing image processing and pattern recognition, remote sensing applications, hyperspectral remote sensing information processing, multisource geospatial information fusion and spatial data handling, integration and applications of geospatial information technologies, and environmental information science (environmental informatics).

Dr. Du is a member of the International Society of Environment Information Science and of the IEEE Geoscience and Remote Sensing Society (IEEE GRSS), and a council member of China Society for Image and Graphics, China Association for Remote Sensing Applications, and Jiangsu Provincial Society of Remote Sensing and Geographic Information System. He was the Cochair of the Technical Committee of the Fifth IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, and the Cochair of the local organizing committee of the 2009 Joint Urban Remote Sensing Event. He also served as the member of scientific committees or local organizing committees of other international conferences, for example, Accuracy 2008, ACRS 2009, WHISPERS 2010, 2011 and 2012, URBAN 2011, MultiTemp 2011, and ISDIF 2011. He is also an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.