Spatial–Spectral Hyperspectral Image Classification Using Random Multiscale Representation

Jianjun Liu, *Member, IEEE*, Zebin Wu, *Member, IEEE*, Jun Li, *Member, IEEE*, Liang Xiao, *Member, IEEE*, Antonio Plaza, *Fellow, IEEE*, and Jon Atli Benediktsson, *Fellow, IEEE*

***Abstract*—This paper presents a novel spatial–spectral classification method for remotely sensed hyperspectral images. First of all, a multiscale representation technique based on random projection, referred as random multiscale representation (RMSR), is proposed to extract the spatial features from the given scene. The idea behind RMSR is to properly model the spatial characteristics comprised by each pixel vector and its neighbors by some criteria computed at all reasonable scales, and then compress the *implicit* high-dimensional spatial features by using a very sparse measurement matrix that approximately preserves the salient spatial information. The entire process is *explicitly* performed by computing simple criteria (i.e., the first two moments) at rectangular scales of random bands, according to the nonzero entries of the sparse measurement matrix. Subsequently, a composite kernel framework is utilized to balance the extracted spatial features and the original spectral features in the classifier. Our proposed method is shown to be effective for hyperspectral image classification purposes. Specifically, our experimental results with hyperspectral images collected by the airborne visible/infrared imaging spectrometer and the reflective optics spectrographic imaging system demonstrate the effectiveness of the proposed method as compared to other state-of-the-art spatial–spectral classifiers.**

***Index Terms*—Composite kernels, compressive sensing, hyperspectral, image classification, multiscale representation, random projection.**

Manuscript received February 2, 2016; revised May 24, 2016 and June 30, 2016; accepted July 1, 2016. Date of publication July 26, 2016; date of current version September 30, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61471199, in part by the Fundamental Research Funds for the Central Universities under Grant JUSRP11559, in part by the Open Project Program of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education under Grant JYB201505, and in part by the China Scholarship Fund under Grant 201406845012. *(Corresponding author: Jianjun Liu.)*

J. Liu is with the School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China, and also with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: liuofficial@163.com).

Z. Wu is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China, and also with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politcnica, University of Extremadura, Cáceres E-10003, Spain (e-mail: zebin.wu@gmail.com).

J. Li is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation and the Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: lijun48@mail.sysu.edu.cn).

L. Xiao is with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: xtxiaoliang@163.com).

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

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

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

Digital Object Identifier 10.1109/JSTARS.2016.2587678

I. Introduction

Supervised classification is an important task in hyperspectral image analysis [1], where pixels are assigned to one of the available classes according to a set of given training pixels. During the last decade, several classifiers have been extended for supervised classification of hyperspectral images, such as support vector machines (SVMs) [2], [3], multinomial logistic regression (MLR) [4], [5], and sparse representation classification [6], [7].

In order to improve the classification performance of these pixelwise classifiers, several approaches have included spatial-contextual information during the classification process. Generally, those spatial–spectral techniques can be roughly divided into two categories:

1) On one hand, some techniques extract the spatial features first, and then incorporate both the spatial and spectral features into the classification process. In this context, several techniques have been exploited, such as mean and median filters [8], [9], Gabor wavelets [10], morphological profiles [11], [12], and mean shift (MS) techniques [13]. After extracting the spatial features, composite kernels (CKs) [8], generalized CKs [14], and multiple kernels [15]–[17] have been used to perform the final classification by considering the spatial features in addition to the spectral features.
2) On the other hand, several other techniques integrate the spatial and spectral information separately. For instance, in [18] the probabilistic SVM is first used to estimate class conditional probability density functions, and then context-based class priors are estimated by using Markov random fields (MRF). In [19], a similar approach is proposed that can be solved in a Bayesian framework, which uses the MLR to learn the posterior probability distributions from the spectral information contained in the data, and then uses the MRF to include spatial-contextual information in the classifier. In [20]–[22], the original hyperspectral image is first classified by a pixelwise classifier and simultaneously segmented into some adaptive neighborhoods by using some segmentation techniques, such as partitional clustering [20], morphological watersheds [21], and minimum spanning forests [22], and then majority voting is adopted to integrate the pixelwise classification results and the obtained segmentation map.

This paper considers designing the spatial–spectral classification method by the first strategy, and mainly concentrates on the exploitation of spatial features that play a fundamental role in this classification process. Generally, the spatial characteristics of a pixel are related to its surrounding pixels. As a result, the spatial feature extraction techniques usually define a fixed spatial neighborhood by a sliding window [8] or create an adaptive spatial neighborhood relating to the parameter value of the filter considered [9], and then compute spatial criteria within the obtained neighborhoods/regions. In order to provide additional discriminant information related to the structures of different objects in a scene, multiscale representation techniques repeating the aforementioned process with different scales (i.e., neighborhoods) to perform a multiscale (or a multilevel) analysis, have been proposed in the literature. For instance, in [11] the derivative of the morphological profile is used to isolate different structures in the image, where the multiscale morphological profile is based on the repeated use of openings and closings with a structuring element of increasing size. In [23], a multilevel context-based system is used for the classification of very high spatial resolution images. In [24], a so-called extended multiattribute profile (EMAP), which is based on the sequential application of different types of attribute filters to a hyperspectral image, is proposed, and the multilevel spatial features are obtained by stacking all filtered images together in the same data structure. In [25], a multiscale MS analysis approach is adopted to compute the multiscale representation of hyperspectral images by using different spatial bandwidths. Several other strategies have been proposed in the literature for multiscale classification [26] and modeling of spatial-contextual information [27] that are not further described here for space considerations.

In the aforementioned multiscale representation techniques, a sequence of scales are usually necessary in order to properly model the spatial information of different objects. Nevertheless, considering a large number of scales can result in high-dimensional feature vectors that will bring a heavy burden on computation and storage. In order to address the problems of high-dimensional features, most methods based on multiscale representation consider to select a limited number of optimal scales empirically [11], [24], [25]. It is obvious that these methods are both unsuitable for an automatic analysis of data and suboptimal since they might not provide a complete characterization of the spatial information [28]. Indeed, if there are two pixels that respectively belong to two similar objects that are not separated by the limited scales, it is difficult for these methods to distinguish the two pixels. In addition, there are other methods that consider feature extraction [29] and feature selection [28] to solve the problem of high-dimensional features. However, these methods cannot decrease the memory demands and computational complexity significantly, as the feature reduction process is generally data dependent and time consuming, and one needs to store the spatial features in memory throughout the entire multiscale representation process.

In view of the aforementioned issues, this paper first proposes a new multiscale representation technique based on random projection for extracting the spatial features of hyperspectral images, and then designs a spatial–spectral classification method by utilizing a CK framework [8] to balance the extracted spatial features and the original spectral features in the classification process. The main idea behind the proposed multiscale representation technique, referred as random multiscale representation (RMSR), is to *implicitly*[1] construct a dense multiscale representation (DMSR) by the first two moments computed at all reasonable rectangular scales in order to provide a complete characterization of the spatial information of the given image, and then compress the *implicit* high-dimensional spatial features by using a very sparse measurement matrix [30] that approximately preserves the salient spatial information. The entire process of RMSR is *explicitly* performed by computing the investigated criteria (i.e., the first two moments) at rectangular scales of random bands, according to the nonzero entries of the sparse measurement matrix, thus avoiding the selection of optimal scales and decreasing the memory demands and computational complexity significantly. Moreover, the computation of the first two moments can be made independent of the scale size by taking advantage of the integral image method in [31]. The novelty of the proposed method relies on the exploitation of the sparse measurement matrix to propose a new multiscale representation technique RMSR for extracting the spatial features of hyperspectral images, which presents several innovative contributions with regards to the existing multiscale representation methods.

1) First and foremost, RMSR neither relies on the empirical selection of optimal scales nor on the feature reduction of spatial features, thus providing an alternative solution for the multiscale representation of hyperspectral images.
2) Second, RMSR can efficiently handle a large number of scales (even dense scales) and, therefore, it can benefit from the complementary information obtained in dense scales, so that the spatial characteristics of a pixel can be properly captured and the simple criteria can be utilized to describe the spatial information within a given scale.
3) Finally, RMSR is general since the criteria available for it are not limited only to the investigated first two moments, and one can easily extend the criteria used in the existing multiscale representation methods to it.

The remainder of this paper is organized as follows. Section II first briefly introduces the very sparse random measurement matrix that will be used in the proposed method, and then proposes RMSR for extracting the spatial features as well as the final spatial–spectral classification method based on the CK technique. The effectiveness of the proposed method is demonstrated in Section III by a series of experiments with two real hyperspectral images, collected by the airborne visible/infrared imaging spectrometer (AVIRIS) and the reflective optics spectrographic imaging system (ROSIS). These experiments demonstrate the effectiveness of the proposed method as compared to other state-of-the-art spatial-spectral classifiers. Finally, Section IV concludes the paper with some remarks and hints at plausible future research lines.

[1]The modifier “implicitly” means the following process that just stands in theory does not exist in practice.

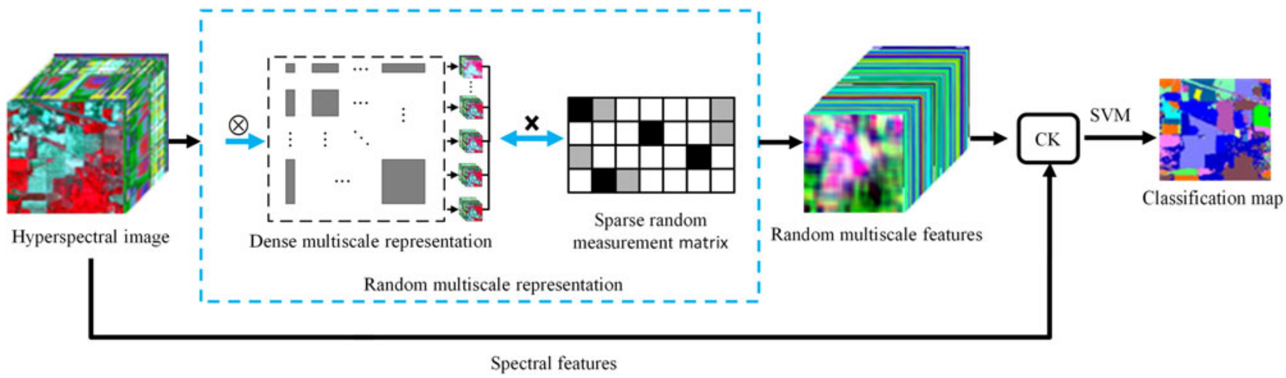


Fig. 1. Flowchart of the proposed spatial–spectral classification method.

II. Proposed Spatial–Spectral Classification Method

Fig. 1 shows a flowchart of the proposed spatial–spectral classification method, which consists of two main steps: 1) extraction of the spatial random multiscale features by using RMSR; and 2) integration of spatial and spectral information in a CK framework using the SVM as the baseline classifier. In the remainder of this section, we describe in more details the strategies adopted for RMSR and final spatial–spectral classification.

A. Very Sparse Random Measurement Matrix

This section briefly introduces the very sparse random measurement matrix that will be used in RMSR. In random projection, a high-dimensional feature vector $\mathbf{a} \in \mathbb{R}^m$ is projected to a lower dimensional feature vector $\mathbf{v} \in \mathbb{R}^n$ by using a random matrix $\mathbf{R} \in \mathbb{R}^{n \times m}$ as follows:

$$\mathbf{v} = \mathbf{R}\mathbf{a} \tag{1}$$

where the columns of $\mathbf{R}$ have unit length, and $n \ll m$. The Johnson–Lindenstrauss (JL) lemma [32] indicates that if the dimensionality n is suitably high, there exists such a random matrix $\mathbf{R}$ that can provide a stable embedding. Specifically, given an integer d and let $n \geq n_0 = \mathcal{O}(\epsilon^{-2}\ln(d))$ with $\epsilon > 0$, then for any two vectors $\mathbf{a}_i$, $\mathbf{a}_j$ in a finite collection $\mathcal{X}$ of d vectors in $\mathbb{R}^m$, we have

$$(1-\epsilon)||\mathbf{a}_i - \mathbf{a}_j||_2^2 \leq ||\mathbf{R}\mathbf{a}_i - \mathbf{R}\mathbf{a}_j||_2^2 \leq (1+\epsilon)||\mathbf{a}_i - \mathbf{a}_j||_2^2. \tag{2}$$

The random Gaussian matrix $\mathbf{R} \in \mathbb{R}^{n \times m}$ is a typical matrix satisfying the JL lemma, where the entry $\mathbf{R}_{ij}$ follows a zero-mean and unit-variance Gaussian distribution. Since the random Gaussian matrix $\mathbf{R}$ is dense, it involves significant memory and computation requirements when m is large. To overcome this problem, in [33] a sparse random matrix $\mathbf{R}$ is introduced with its entries defined as follows:

$$\mathbf{R}_{ij} = \sqrt{\rho} \times \begin{cases} 1 & \text{with probability } \frac{1}{2\rho} \\ 0 & \text{with probability } 1 - \frac{1}{\rho} \\ -1 & \text{with probability } \frac{1}{2\rho} \end{cases} \tag{3}$$

where $\rho = 3$,[2] and it is proved that if $n \geq (4 + 2\beta)(\epsilon^2/2 - \epsilon^3/3)^{-1}\ln(d)$ with $\beta > 0$, the statement (2) holds true with probability at least $1 - d^{-\beta}$.[3] This matrix R is easy to compute by using a uniform random generator. In [34], it is proved that, for $\rho = \mathcal{O}(m)$ ($\mathbf{a} \in \mathbb{R}^m$) even $\rho = m/\ln(m)$, the sparse random matrix $\mathbf{R}$ is almost as effective as the conventional random Gaussian matrix. Moreover, in [30] good results have been shown for visual tracking by setting $\rho = m/c$ with a small c, i.e., there are only about c nonzero entries in each row of $\mathbf{R}$. Although a large c is beneficial to improve the accuracy as shown in [34], we set $c = 4$ in our experiments as a tradeoff between accuracy and computational cost (see Section III-B). Therefore, the adopted random matrix $\mathbf{R}$ with $\rho = m/4$ is very sparse, and the computational cost is very low (only $\mathcal{O}(4n)$). In addition, the memory requirements are also very light, as we only need to store the nonzero entries of $\mathbf{R}$.

B. Random Multiscale Representation

Recent works on multiscale representations for hyperspectral imagery have shown that it is necessary to include a large number of scales during the representation process [11]–[13], [24], [25]. Inspired by this observation, we intend to use dense rectangular scales to provide a complete characterization of the spatial information, i.e., for each pixel vector, all possible rectangular neighborhoods centered at the pixel should be ideally included. Intuitively, increasing the number of scales is beneficial for the classification performance, since the more scales are used, the more spatial characteristics are provided.

Let us denote a hyperspectral image by $\mathbf{X} \in \mathbb{R}^{L \times W \times H}$, with L being the number of bands and $W \times H$ being the spatial dimensions. For each pixel vector $\mathbf{x} \in \mathbb{R}^L$ of $\mathbf{X}$, its DMSR is constructed in this paper by computing some criteria of its rectangular neighborhoods at dense scales $\{\mathbf{F}_{1,1}, \mathbf{F}_{1,2}, ..., \mathbf{F}_{w,h}\}$, i.e.,

$$\mathbf{F}_{w,h}(\mathbf{x}) = f(\mathbf{\Omega}_{\mathbf{x}}^{w,h}) \in \mathbb{R}^L \tag{4}$$

where $f(\cdot)$ denotes the criterion function per spectral channel, $\mathbf{\Omega}_{\mathbf{x}}^{w,h}$ is the rectangular neighborhood centered at $\mathbf{x}$,

[2] $\rho = 1$ also holds true, but the random matrix $\mathbf{R}$ is not sparse.

[3] The columns of the sparse random matrix $\mathbf{R}$ are unnormalized, thus the statement (2) should be rewritten as $(1-\epsilon)||\mathbf{a}_i - \mathbf{a}_j||_2^2 \leq (1/n)||\mathbf{R}\mathbf{a}_i - \mathbf{R}\mathbf{a}_j||_2^2 \leq (1+\epsilon)||\mathbf{a}_i - \mathbf{a}_j||_2^2$.

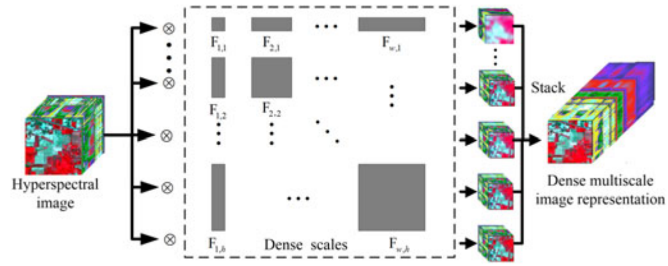


Fig. 2. Graphical illustration of the procedure adopted for dense multiscale hyperspectral image representation.

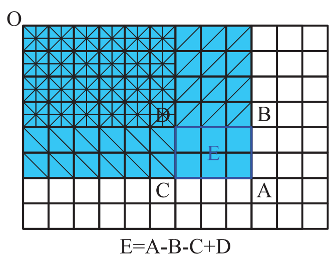


Fig. 3. Graphical illustration of the procedure used to obtain an integral image. Symbol "O" denotes the origin; the blue blocks denote region "A", and the blocks "/", "\", and "+" represent regions "B", "C", and "D", respectively. It takes only three additions and four memory accesses to calculate the summation of intensities inside a region "E" of any size.

and $2w+1$ and $2h+1$ are the width and height of $\mathbf{\Omega}_{\mathbf{x}}^{w,h}$, respectively. Here, we pad $\mathbf{X}$ with minor reflections of itself to deal with the outliers. Then, we concatenate these vectors $\mathbf{F}_{w,h}(\mathbf{x})$ to form a high-dimensional multiscale feature vector $\mathbf{a} = [\mathbf{F}_{1,1}(\mathbf{x}); \mathbf{F}_{1,2}(\mathbf{x}); ...; \mathbf{F}_{w,h}(\mathbf{x})] \in \mathbb{R}^m$ where $m = whL$. Fig. 2 shows a graphical illustration of the aforementioned procedure to construct the DMSR of the hyperspectral image.

Once the neighborhoods of each pixel vector have been defined, the DMSR features are extracted by computing some criteria associated to the considered neighborhoods. As discussed in [8], the simplest (but still powerful) spatial features are based on the application of moment criteria. For simplicity, the criteria investigated in the proposed method are the first two moments, i.e., mean and standard deviation. Note that, the integral image method can be used to accelerate the computation of the mean [31]. Specifically, for the given hyperspectral image $\mathbf{X} \in \mathbb{R}^{L\times W\times H}$, the entry of its integral image $\mathbf{I} \in \mathbb{R}^{L\times W\times H}$ at a spatial location (x, y) represents the summation of all pixels in $\mathbf{X}$ per spectral channel within a rectangular region formed by the origin and (x, y)

$$\mathbf{I}(l, x, y) = \sum_{i=1}^{x}\sum_{j=1}^{y} \mathbf{X}(l, i, j), l = 1, 2, ..., L. \tag{5}$$

After computing the integral image, we just need three additions to calculate the summation of intensities over any rectangular region, as illustrated graphically in Fig. 3. Then, for a given neighborhood $\mathbf{\Omega}_{\mathbf{x}}^{w,h}$ centered at $\mathbf{x} = \mathbf{X}(:, x, y) \in \mathbb{R}^L$, the mean function $f_\mu(\cdot)$ can be calculated by the following expression:

$$\begin{aligned} f_\mu(\mathbf{\Omega}_{\mathbf{x}}^{w,h}) &= [\mathbf{I}(:, x+w, y+h) - \mathbf{I}(:, x+w, y-h-1) \\ &- \mathbf{I}(:, x-w-1, y+h) + \mathbf{I}(:, x-w-1, y-h-1)]/N \end{aligned} \tag{6}$$

where $N = (2w+1)(2h+1)$. As for the standard deviation function $f_\sigma(\cdot)$, it can be written as follows:

$$\begin{aligned} f_\sigma(\mathbf{\Omega}_{\mathbf{x}}^{w,h}) &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_{(i)} - \bar{\mathbf{m}}\right)^2} \\ &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\mathbf{x}_{(i)}^2 - \frac{N}{N-1}\bar{\mathbf{m}}^2} \end{aligned} \tag{7}$$

where $\mathbf{x}_{(i)}$ is the ith pixel vector in $\mathbf{\Omega}_{\mathbf{x}}^{w,h}$, $\bar{\mathbf{m}} = f_\mu(\mathbf{\Omega}_{\mathbf{x}}^{w,h})$ and $(\cdot)^2$ denotes the component-wise application of the square function. It is easy to see that the integral image method can also be used to solve (7) if we precompute the integral images of $\mathbf{X}$ and $\mathbf{X}^2$. Hence, the computational complexity of the first two moments is independent of the size of the given rectangular scale. This is important for the proposed method, as we intend to use dense scales in order to properly model the spatial-contextual information.

After obtaining the DMSR features for all pixel vectors, we can use them as spatial features for classification purposes. However, as recommended in the existing multiscale representation methods [11], [24], [25], [28], [29], it is impractical to directly use the obtained DMSR features since they lie in a high-dimensional space $\mathbb{R}^m$. As discussed in the Section I, the methods based on the manual setting of optimal scales might not provide a complete characterization of the spatial information, and the methods based on feature reduction cannot decrease the memory demands and computational complexity significantly. Thus, it is necessary to design an effective mechanism to address the problems of the DMSR features. Moreover, the dimensionality m of the DMSR features is typically in the order of 10^6 (if a given image contains 100 spectral bands with size of 200×200 pixels, then $m = whL \approx 100 \times 100 \times 100 = 10^6$) and, therefore, the feature reduction techniques adopted in the existing multiscale representation methods are impractical for the proposed DMSR due to the limited memory load and compute capacity.

In order to circumvent the aforementioned limitations, we construct a very sparse random measurement matrix $\mathbf{R} \in \mathbb{R}^{n\times m}$ introduced in Section II-A to project every DMSR feature vector $\mathbf{a} \in \mathbb{R}^m$ onto a low-dimensional vector $\mathbf{v} \in \mathbb{R}^n$. The entire process composed by the DMSR and the random matrix $\mathbf{R}$ is referred as RMSR. Since the random matrix $\mathbf{R}$ is data independent and very sparse, RMSR can be organically performed by computing the investigated criteria (i.e., the first two moments) at random rectangular scales of random bands according to the nonzero entries of $\mathbf{R}$ as shown in Fig. 4. Thus, we do not need to store the DMSR features, and the computational cost of RMSR is very light.

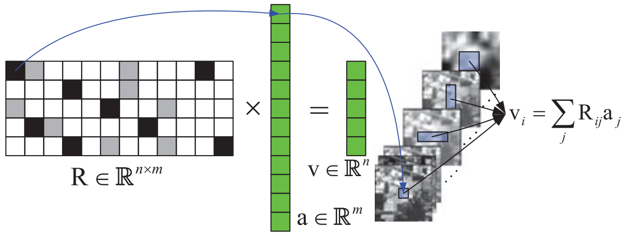


Fig. 4. Illustration of the procedure used for projecting a high-dimensional vector $\mathbf{a}$ onto a low-dimensional vector $\mathbf{v}$. In the matrix $\mathbf{R}$, white, gray, and black blocks denote zero, positive, and negative entries, respectively. The arrows show that for each pixel vector of an input hyperspectral image, one of the nonzero entries in each row of $\mathbf{R}$ (multiplied by an element in $\mathbf{a}$) is equivalent to computing a criterion at a particular band and scale.

It is worth noting that, although the very sparse measurement matrix $\mathbf{R}$ can provide a stable embedding that approximately preserves the salient spatial information, the dimensionality n of it should be suitably high as stated in the JL lemma [32]. In [33], a JL-bound (introduced in Section II-A) has been given for the sparse measurement matrix. For hyperspectral images with about $d \approx 10^5$ pixels, $\epsilon = 0.2$, and $\beta = 1$, the lower bound for n is approximately 4000. However, this JL-bound is conservative in many applications [34]. In [35], the results show that the JL-bound is much higher than that, which is sufficient to give good results on image and text data. As for hyperspectral images, this also holds true as shown in Section III-B.

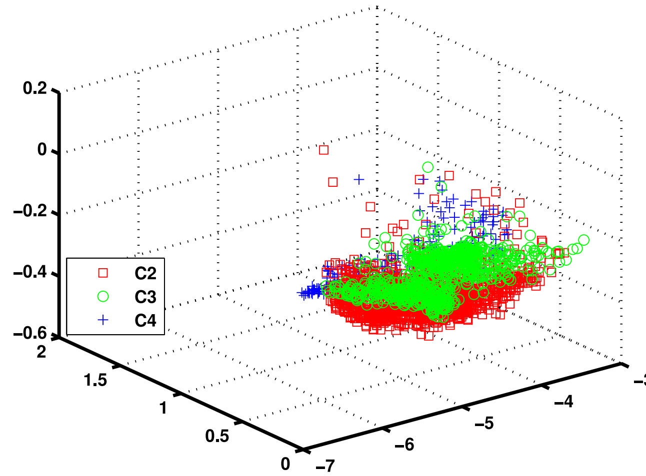
To interpret the aforementioned issue, one can resort to the perspective of compressive sensing. In compressive sensing [36]–[38], it is shown that, for random projection (1), if $\mathbf{R}$ satisfies the restricted isometry property (RIP) [36], [39] and $\mathbf{a}$ is sparse, then the compressive measurement $\mathbf{v}$ approximately preserves the salient information in $\mathbf{a}$, i.e.,

$$(1-\epsilon)||\mathbf{a}||_2^2 \leq ||\mathbf{Ra}||_2^2 \leq (1+\epsilon)||\mathbf{a}||_2^2. \tag{8}$$

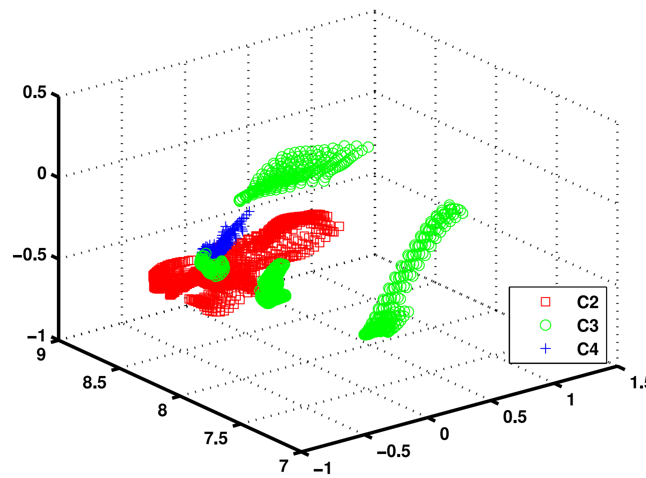
In [40], it is proved that any random matrix $\mathbf{R}$ satisfying the JL lemma can also satisfy the RIP in compressive sensing. The random matrix $\mathbf{R}$ adopted in RMSR asymptotically satisfies the JL lemma, thus holding true for the statement (8). Moreover, the statement (2) is equal to (8) by setting $\mathbf{a} = \mathbf{a}_i - \mathbf{a}_j$. In other words, if $\mathbf{a}_i - \mathbf{a}_j$ is sparse, the bound $n \geq \kappa\beta \log(m/\beta)$ [30] (with constants κ and β) derived from the RIP is available. It is apparent that the pairwise differences of the DMSR features are very sparse and highly compressible, since dense scales are used and related features are very similar to each other. In this paper, as recommend in [30], it is expected that $n \geq 50$ when $m = 10^6$, $\kappa = 1$ and $\beta = 10$, which is much tighter than the JL-bound.

Furthermore, we should highlight that the proposed RMSR is general, since the criteria available for it are not limited only to the investigated first two moments, and other criteria that can be used to filter an image within a sliding window, such as median and singular value, are all available.[4] In order to illustrate the generalization of RMSR, the median criterion is also used for evaluation purposes in our experiments. Moreover, one can easily extend the adaptive filters used in the existing multiscale representation methods to RMSR by replacing the rectangle neighborhoods with the adaptive neighborhoods, but this is beyond the scope of this paper.

[4]The integral image method cannot be used to accelerate the computation of the mentioned criteria except the first two moments.



(a)



(b)

Fig. 5. Feature vectors obtained by PCA. (a) Spectral feature vectors. (b) RMSR feature vectors.

Next, we graphically illustrate how the proposed RMSR features separate the classes. Specifically, the AVIRIS Indian Pines dataset is tested, and three classes C2, C3, and C4 (see Section III-A) are considered. Both the spectral features and the RMSR features of the three classes are reduced by principal component analysis (PCA), and the first three principal components (PCs) are retained. The scatterplots depicting the reduced vectors of the spectral features and the RMSR features are shown in Fig. 5(a) and (b), respectively. It is clear that the RMSR features can increase the margin between the classes and improve the smoothness within class.

C. Spatial–Spectral Classification With CK

After performing the spatial feature extraction process of RMSR, each pixel has three kinds of features: spectral features $\mathbf{x}$, which are the original values of each pixel vector's elements; and spatial features $\mu_{\mathbf{x}}$ and $\sigma_{\mathbf{x}}$, which are the projective mean and standard deviation values computed within each pixel's neighborhoods, respectively. In this paper, we adopt the CK framework [8] to combine the spatial and spectral features for the final classification, where the used kernel is the Gaussian radial basis function (RBF)

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma ||\mathbf{x}_i - \mathbf{x}_j||_2^2\right), \gamma \in \mathbb{R}^+. \tag{9}$$

In our context, the spatial–spectral CK $\mathcal{K}$ is defined as

$$\mathcal{K}(i,j) = \nu K(\mathbf{x}_i, \mathbf{x}_j) + (1-\nu) K\left([\mu_{\mathbf{x}_i}; \sigma_{\mathbf{x}_i}], [\mu_{\mathbf{x}_j}; \sigma_{\mathbf{x}_j}]\right) \tag{10}$$

where $0 \leq \nu \leq 1$ controls the relative weight of the spatial and spectral information.

Once the spatial–spectral kernel $\mathcal{K}$ has been constructed, the next step is to perform the final classification procedure. Here, we adopt the SVM (embedded in the aforementioned CK framework) to conduct the final classification. We have selected the SVM because it is one the most successful kernel classifiers and is intrinsically less sensitive to the high dimensionality of the feature space.

It is worth noting that in this paper, the integration of RMSR into a RBF-based CK has been taken with overall consideration of the following three issues:

1) Recent works [41] have highlighted the importance of kernel-based methods in the classification of hyperspectral images, thus we only consider investigating the proposed RMSR features by using kernel-based classifiers.

2) The incorporation of spectral features is beneficial to the classification accuracy of spatial features, and thus the CK, which is widely used to combine the spectral and spatial features [1], is adopted.

3) Only the pairwise l_2-norm distance of the DMSR features is maintained by RMSR as shown in (2) and the RBF kernel is just the suitable one that is based on l_2-norm.

III. Experimental Results and Analysis

In this section, we first introduce the two hyperspectral datasets used in experiments, and then present a series of experimental results to validate the proposed method. As mentioned before, an SVM classifier (embedded in a CK framework for spatial–spectral classification) is used, where the LIBSVM software package [42] has been used for the specific implementation.[5] The free parameters of the SVM have been carefully optimized via a cross-validation procedure. More specifically, the RBF-kernel parameter is varied in the range $\gamma \in \{2^{-4}, ..., 2^4\}$, and the regularization parameter is varied in the range $C \in \{10^0, ..., 10^5\}$. As for the free parameter ν of the CK framework, it is varied in the range $[0, 1]$. A detailed explanation of the procedure followed to optimize parameter settings is given in [2] and [8], and is not repeated here for space considerations. To evaluate the performance of the classification methods, three widely used quality metrics are adopted, i.e., the overall accuracy (OA), the average accuracy (AA), and the Kappa coefficient (κ). In addition, the original data have been scaled in the range $[0, 1]$ before the experiments and, unless otherwise specified, the quantitative measures are obtained by averaging ten Monte Carlo runs. Besides the first two moments, another criterion (i.e., the median) is also used to implement the proposed method. In the following, the proposed method implemented with moment criteria is abbreviated by MOM and the one implemented using the median criterion is abbreviated by MED.

[5]In LIBSVM, a "one-against-one" strategy is implemented for multiclass classification.

TABLE I
Sixteen Ground Reference Classes in AVIRIS Indian Pines and the Training and Test Sets

Class		Samples	
NO	Name	Train	Test
C1	Alfalfa	3	51
C2	Corn-no till	72	1362
C3	Corn-min till	42	792
C4	Corn	12	222
C5	Grass/pasture	25	472
C6	Grass/trees	38	709
C7	Grass/pasture-mowed	2	24
C8	Hay-windrowed	25	464
C9	Oats	2	18
C10	Soybeans-no till	49	919
C11	Soybeans-min till	124	2344
C12	Soybean-clean till	31	583
C13	Wheat	11	201
C14	Woods	65	1229
C15	Bldg-grass-tree drives	19	361
C16	Stone-steel towers	5	90
Total		525	9841

A. Hyperspectral Imagery Datasets

The first hyperspectral image dataset used in our experiments is an image collected by the AVIRIS over NW Indiana's Indian Pines region in 1992. The AVIRIS sensor collects 220 bands, covering the wavelength range of 0.4–2.5 μm, and the number of retained bands is 200 (after removing 20 water absorption bands). This image consists of 145×145 pixels, with a nominal spectral resolution of 10 nm and moderate spatial resolution of 20 m by pixel. It contains 16 ground reference classes ranging from 20 to 2468 pixels in size and, for each class, we randomly choose around 5% of the labeled samples for training and use the remaining 95% for testing. The details of training and test sets are shown in Table I. For illustrative purposes, Fig. 6 shows a false color composite of the scene and the associated ground reference map.

The second hyperspectral image dataset used in our experiments was collected over the University of Pavia, Italy, by the reflective optics system imaging spectrometer (ROSIS). The ROSIS sensor collects 115 bands, covering the wavelength range of 0.43–0.86 μm. After removing the noisy bands, we retained

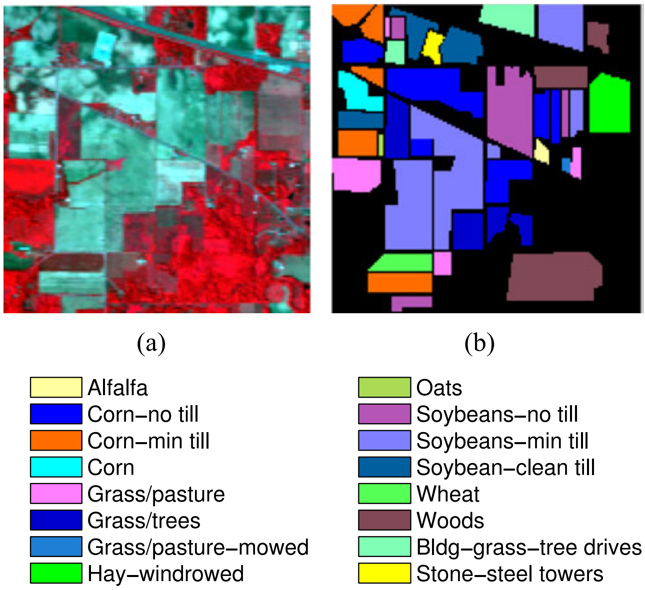


Fig. 6. AVIRIS Indian Pines dataset. (a) RGB composite image of three bands. (b) Ground reference map.

TABLE II
NINE GROUND REFERENCE CLASSES IN ROSIS UNIVERSITY OF PAVIA AND THE TRAINING AND TEST SETS

Class		Samples	
NO	Name	Train	Test
C1	Asphalt	40	6812
C2	Meadow	40	18 646
C3	Gravel	40	2167
C4	Trees	40	3396
C5	Metal sheets	40	1338
C6	Bare soil	40	5064
C7	Bitumen	40	1316
C8	Bricks	40	3838
C9	Shadows	40	986
Total		360	43563

103 bands in the acquired image. The scene contains 610 $\times$ 340 pixels, with very high spatial resolution of 1.3 m per pixel. There are nine ground reference classes of interest, and we randomly choose 40 samples per class for training and use the rest for testing. The details of training and test sets are shown in Table II. In addition, a false color composite image and the associated ground reference map are shown in Fig. 7.

B. Different Sparse Random Measurement Matrices

There are two tuning parameters in the construction of the very sparse random measurement matrix $\mathbf{R} \in \mathbb{R}^{n\times m}$: c controls the number of the nonzero entries in each row of $\mathbf{R}$, and n is the dimensionality of the projection vectors (i.e., the number of spatial features). The first set of experiments is designed to analyze the impact of these two parameters, and the criterion is the distortion error measured by comparing the l_2-norm distance between two projection data vectors to their l_2-norm distance in

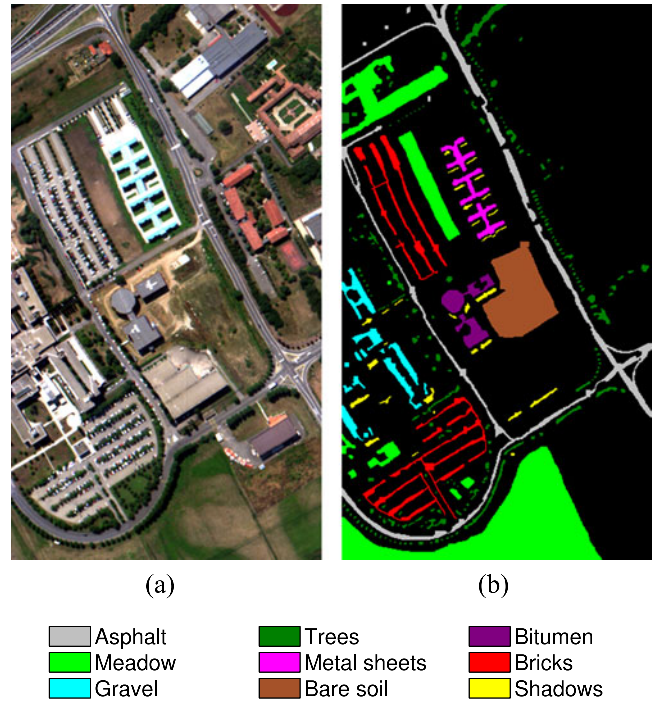


Fig. 7. ROSIS University of Pavia dataset. (a) RGB composite image of three bands. (b) Ground reference map.

the original high-dimensional space, which is defined as

$$\epsilon(\mathbf{a}_i, \mathbf{a}_j) = \left|1 - \frac{1}{\sqrt{n}} \cdot \frac{\|\mathbf{R}\mathbf{a}_i - \mathbf{R}\mathbf{a}_j\|_2}{\|\mathbf{a}_i - \mathbf{a}_j\|_2}\right|. \quad (11)$$

Here, we randomly choose 1000 pairs of data vectors per dataset for testing, and compute the error between members of a pair of data vectors, averaged over these pairs, for each $\mathbf{R}$. c is varied to be 2, 4, 10, 50, and 100, and n is varied in steps of 20 in the range [20, 1000]. At each pair c and n, $\mathbf{R}$ is generated anew 100 times to obtain the average error. Fig. 8 shows the results for the two given datasets. As for the free parameters w and h in the multiple scales $\{\mathbf{F}_{i,j}\}_{i=1,j=1}^{w,h}$, they were set equal to each other. More specifically, for Indian Pines dataset we used, $w = h = 50$, and for University of Pavia dataset we used, $w = h = 100$.

As shown by Fig. 8, it is clearly seen that random projection can yield very accurate results. When n is analyzed, the performance obtained by using more features is better than that obtained using less features, but the differences observed between both cases are quite slight when $n > 100$. When c is analysed, the results obtained by using more nonzero entries are better than that obtained using less nonzero entries in most cases, but the gaps among the results of different c are not significant. As a result, we can conclude that the construction of $\mathbf{R}$ is robust to parameter settings, and almost all the information provided by the DMSR features is preserved by RMSR. In the following, we assume that c is fixed to 4 and n is fixed to 200 as a tradeoff between accuracy and computational cost.

In the next set of experiments, we investigate the sparse random projection by statistically analyzing the errors of using different $\mathbf{R}$. Different from the first set of experiments, the

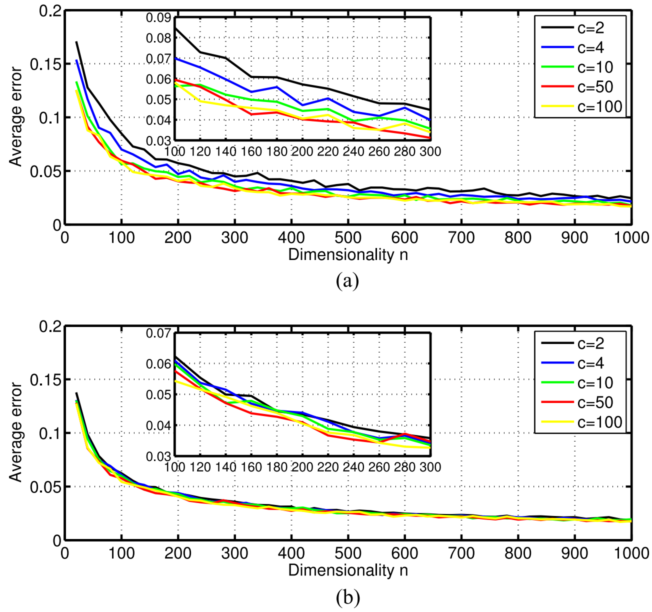


Fig. 8. Influence of the parameters of $\mathbf{R}$. (a) AVIRIS Indian Pines scene. (b) ROSIS University of Pavia scene.

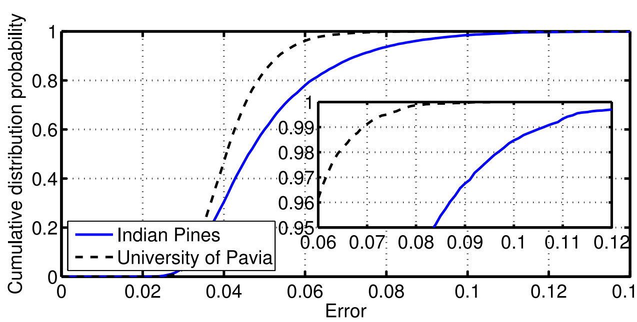


Fig. 9. Cumulative distribution probability of error.

suggested parameters c and n are used and $\mathbf{R}$ is generated anew 10^4 times to avoid any bias induced by random generator. Fig. 9 shows the cumulative distribution probability of error for the two given datasets. It is apparent that the sparse random projection can achieve a low error with a high probability. Thus, we can conclude that the sparse random projection process is robust.

C. Different Numbers of Scales

In this set of experiments, we investigate the impact of the input parameters on RMSR. There are only two parameters: w and h intended for the definition of the multiple scales $\{\mathbf{F}_{i,j}\}_{i=1,j=1}^{w,h}$, which control the scale size of the largest rectangle neighborhood. The product of w and h gives the number of scales used by RMSR, and their values should not be more than half of the width and height of the given dataset, respectively. Several values for w and h have been tested in our experiments with the two considered hyperspectral datasets, and the obtained classification accuracies are reported in Fig. 10(a) and (b), respectively. In these experiments and

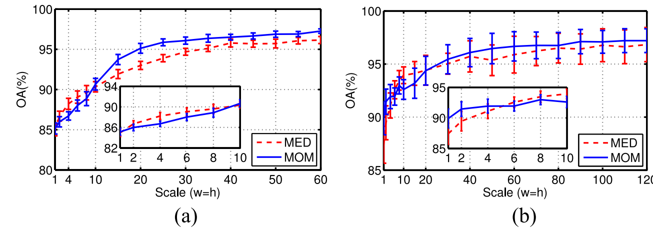


Fig. 10. Classification accuracies using different numbers of scales. (a) AVIRIS Indian Pines scene. (b) ROSIS University of Pavia scene.

the following, the parameters w and h are set equal to each other, and the error bars indicate the standard deviation for ten random samplings.

From Fig. 10, it can be seen that for both the two given datasets, when the number of scales is small, MED performs better than MOM in most cases; whereas when the number of scales is large, MOM outperforms MED consistently. As the number of scales increases, the classification accuracies of both MED and MOM increase almost monotonically. This indicates that a large number of scales have advantages in improving classification performance. However, the accuracies reach saturation asymptotically when the number of scales is relatively large. This is because the rectangle neighborhoods with large widths or heights might cover many objects that the computed criteria of different classes are almost the same. In view of the above issues, a moderate number of scales are used in the experiments, since a larger number of scales will increase the computational complexity of MED significantly and will not bring significant performance improvements. More specifically, we hereinafter set parameters w and h to a fixed value of 50 for the AVIRIS Indian Pines scene and to a fixed value of 100 for the University of Pavia scene. These empirical parameter settings were found appropriate after simple experimentation, which indicate that these parameters are not difficult to set in practice.

D. Influence of Kernel Weight

The kernel weight ν in (10) balances the spectral and spatial information in the construction of the proposed method. In this set of experiments, we examine the influence of it to the performance of the investigated MED and MOM by varying ν from 0 to 1. The classification results for the two given datasets are shown in Fig. 11. It can be seen that when ν is set to 0 or 1, i.e., only spatial information or spectral information is used, the investigated two methods do not perform very good results on both datasets. The optimal values of ν are different for the two given datasets, but they are all in the range [0.1, 0.5]. When ν goes from 0.6 to 1, almost all cases degrade performance apparently. All these indicate that the spatial features play an important role in the spatial–spectral classification, and the integration of the spectral features into the spatial features can improve the classification performance when compared with the cases only using the spatial features.

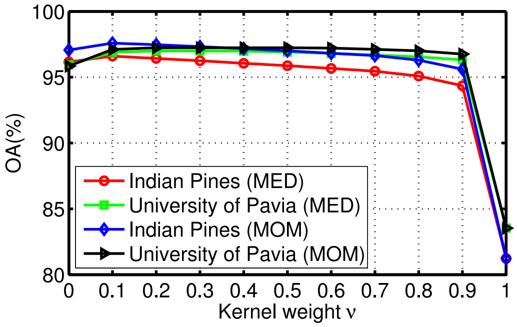
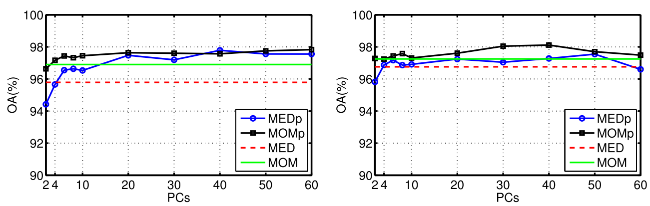


Fig. 11. OA as a function of the kernel weight ν.



(a) (b)

Fig. 12. Influence of base images. (a) AVIRIS Indian Pines scene. (b) ROSIS University of Pavia scene.

E. Influence of Base Images

In traditional multiscale methods, it is suggested to use dimensionality reduction methods to generate the characteristic images, and then use the obtained characteristic images as base images to construct the multiscale representations [11]–[13], [24], [25]. In this set of experiments, we analyze the impact of using and not using base images on RMSR. As with most multiscale methods, PCA is adopted as the dimensionality reduction method, and the q most significant PCs are used to produce the base images. In our experiments, several values for q were tested, and the corresponding methods for MED and MOM are denoted by MEDp and MOMp, respectively. As for RMSR without using base images, the original spectral features were used, and the corresponding methods are MED and MOM themselves.

A comparison of the results obtained using and not using bases images for the classification of the two considered datasets is given in Fig. 12(a) and (b), respectively. From the results reported in Fig. 12, it is clear that using base images could be beneficial to RMSR. However, this is not the main goal of our approach as we intend to conduct the classification with all the information available at hand, without any prior dimensionality reduction. Thus, in this paper we do not consider using bases images for the construction of the proposed method.

F. Comparison of Different Classification Methods

In this set of experiments, we compare the proposed MED, MOM, and DMSR (i.e., MOM without random projection) with three widely used spatial–spectral classification methods based on both the SVM classifier and the CK framework. The first one is the SVMCK, where the spatial features are extracted by computing the mean and standard deviation of the neighborhood pixels in a window per spectral signature, and more details can be seen in [8]. The second one is denoted by MSCK, where a multiscale MS technique is utilized to extract the spatial features, and the details of this methodology are given in [25]. The last one is denoted by EMAPCK, since EMAP is used to build the spatial features. As reported in [14], the first three PCs are retained, and the area and standard deviation are considered to build the morphological attribute profiles. The corresponding parameters and codes of EMAP are provided by [12] and [24].

TABLE III
CLASSIFICATION ACCURACIES FOR AVIRIS INDIAN PINES USING DIFFERENT CLASSIFICATION METHODS

Class	SVM	SVMCK	MSCK	EMAPCK	DMSR	MED	MOM
C1	63.92	77.25	89.02	84.31	90.59	88.04	**91.57**
C2	77.25	85.10	88.91	89.60	95.60	94.38	**96.26**
C3	73.52	90.76	91.49	93.40	97.58	96.88	**97.98**
C4	53.02	73.20	83.24	72.34	87.79	86.35	**89.10**
C5	89.79	91.76	90.51	91.84	93.96	93.69	**94.47**
C6	93.98	96.12	95.95	97.01	98.39	97.64	**98.67**
C7	85.83	90.83	62.08	92.08	94.58	93.75	**95.00**
C8	95.97	97.37	99.38	99.20	99.48	99.38	**99.55**
C9	55.56	81.67	73.33	61.11	82.22	**88.33**	82.78
C10	71.75	82.66	86.75	87.33	92.84	92.62	**93.36**
C11	81.01	90.74	91.55	94.67	96.86	96.13	**97.69**
C12	80.19	84.27	86.96	88.59	**96.30**	95.27	95.83
C13	98.41	99.40	99.35	98.86	99.10	97.56	**99.45**
C14	93.86	95.97	98.67	98.00	99.29	99.06	**99.43**
C15	52.41	76.40	81.52	90.86	96.20	**96.34**	96.20
C16	85.56	**93.67**	88.67	90.67	92.78	91.22	93.11
OA	81.23	89.42	91.44	92.77	96.45	95.79	**96.90**
	(0.53)	(1.00)	(0.63)	(0.71)	(0.32)	(0.41)	(0.32)
AA	78.25	87.95	87.96	89.37	94.60	94.17	**95.03**
	(1.57)	(1.23)	(1.94)	(1.15)	(1.13)	(1.30)	(1.27)
κ	78.62	87.95	90.26	91.76	95.95	95.21	**96.47**
	(0.61)	(1.14)	(0.72)	(0.81)	(0.37)	(0.47)	(0.36)
Time(s)	**0.67**	1.97	30.10	4.82	–	35.50	2.54

The standard deviation (in the parentheses) of the ten Monte Carlo runs is also reported in each case.

TABLE IV
CLASSIFICATION ACCURACIES FOR ROSIS UNIVERSITY OF PAVIA USING DIFFERENT CLASSIFICATION METHODS

Class	SVM	SVMCK	MSCK	EMAPCK	DMSR	MED	MOM
C1	76.25	87.41	90.48	**96.32**	94.69	95.22	94.06
C2	83.22	93.54	92.12	92.65	96.42	96.22	**97.20**
C3	79.04	85.43	96.86	87.36	97.96	98.16	**98.76**
C4	93.21	97.76	92.97	95.71	97.28	96.57	**98.04**
C5	99.07	99.91	99.09	99.25	99.98	99.77	**99.99**
C6	83.16	93.91	93.57	94.83	**99.73**	99.37	99.26
C7	92.27	95.43	94.41	97.63	99.88	99.70	**99.92**
C8	79.96	87.43	95.40	91.72	**96.73**	95.26	96.52
C9	99.16	99.48	98.58	99.30	98.98	**99.56**	99.44
OA	83.53	92.40	93.05	93.87	96.98	96.76	**97.25**
	(2.02)	(1.07)	(1.32)	(1.43)	(1.35)	(1.54)	(1.17)
AA	87.26	93.37	94.83	94.97	97.96	97.76	**98.13**
	(1.19)	(0.46)	(0.64)	(0.58)	(0.43)	(0.39)	(0.38)
κ	78.91	90.12	90.97	92.04	96.06	95.78	**96.41**
	(2.43)	(1.35)	(1.68)	(1.80)	(1.73)	(1.97)	(1.50)
Time(s)	**3.26**	8.87	398.70	92.06	–	1200.64	16.34

The standard deviation (in the parentheses) of the ten Monte Carlo runs is also reported in each case

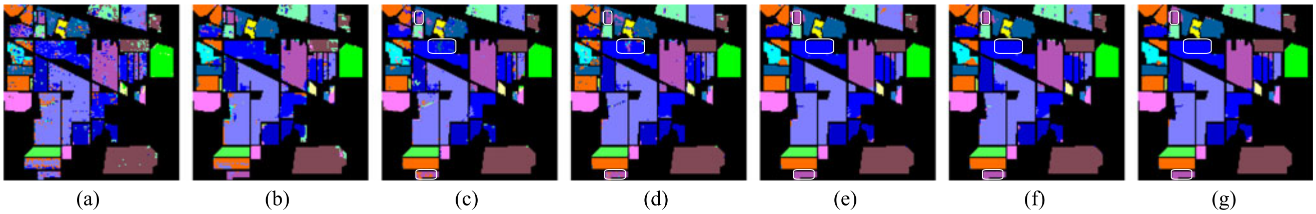


Fig. 13. Classification maps and overall classification accuracies (in the parentheses) obtained for the AVIRIS Indian Pines scene using different classification methods. (a) SVM (80.92). (b) SVMCK (89.35). (c) MSCK (91.70). (d) EMAPCK (92.98). (e) DMSR (96.18). (f) MED (95.47). (g) MOM (96.72) .

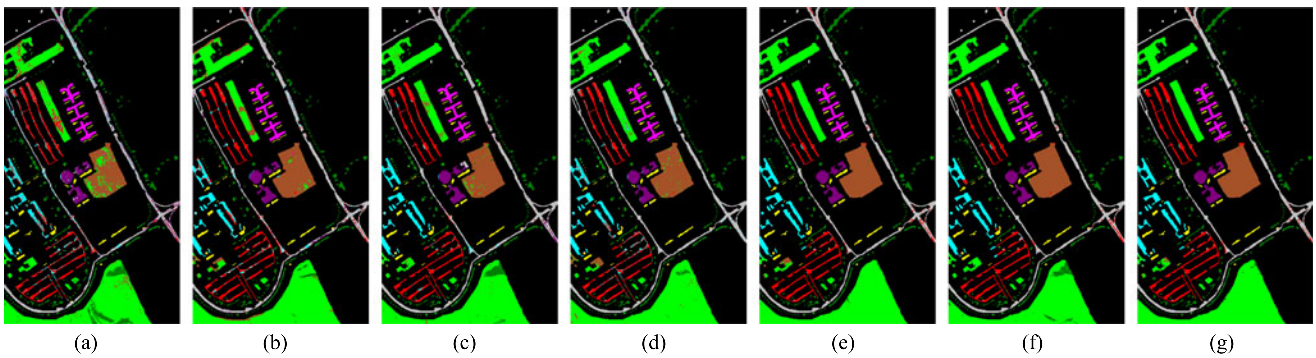


Fig. 14. Classification maps and overall classification accuracies (in the parentheses) obtained for the ROSIS University of Pavia scene using different classification methods. (a) SVM (83.62). (b) SVMCK (91.05). (c) MSCK (93.87). (d) EMAPCK (94.31). (e) DMSR (97.46). (f) MED (96.38). (g) MOM (97.96).

In order to emphasize the assessment of the proposed RMSR features, the kernel weights ν of the aforementioned methods are all set to 0.5 for a fair comparison of their spatial features. Moreover, the pixelwise SVM that only uses spectral features [2] is also included as a baseline classifier. The global and class-specific accuracies obtained by all the considered classification methods for the two given datasets are reported in Tables III and IV, and the corresponding classification maps are illustrated in Figs. 13 and 14, respectively. In addition, the processing time in seconds, measured in a 64-b quad-core Intel CPU 2.40-GHz processor, is also included for reference. Here, the processing time of DMSR is not included since it is far more than the others.

For the AVIRIS Indian Pines scene, as reported in Table III, all spatial–spectral methods yield higher classification accuracies when compared with the pixelwise SVM. Among these methods, MOM gives the highest global and most of the best class-specific accuracies, while DMSR can be regarded as the second-best method followed by MED. Note that, DMSR should ideally perform better than MOM, but the experimental results are just the opposite. This is because the dimensionality of DMSR features is so high that the classifier may suffer from the high dimensionality of the feature space. However, this impact is not serious, since the gap between MOM and DMSR is not significant. The classification maps reported in Fig. 13 confirm these observations conducted on Table III. It is also clear that the spatial–spectral methods provide smoother classification maps than the spectral-only SVM. The classification maps obtained by the proposed MOM, DMSR, and MED contain smoother regions than those produced by the other methods, and the misclassified pixels mainly concentrate in the corners that are difficult to be described by all compared spatial features, due to the information asymmetry among the classes lying in these irregular regions. Moreover, although MSCK and EMAPCK are also the methods based on multiscale spatial features, the classification maps of them are not as smooth as those of MOM, DMSR, and MED, and contain some block-like misclassified structures, especially in the flat regions (e.g., rectangle regions in Fig. 13), that are correctly classified by the proposed three methods.

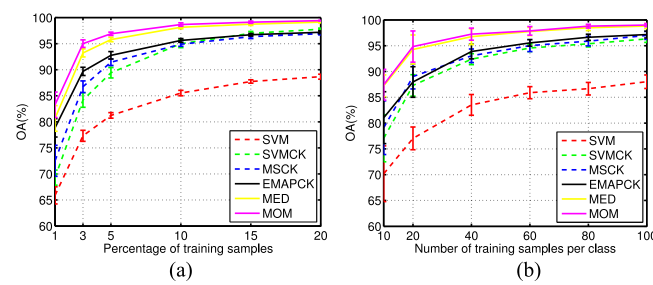


Fig. 15. OA as a function of the number of training samples for different classification methods when applied to the (a) AVIRIS Indian Pines scene and (b) ROSIS University of Pavia scene.

From Table IV, it is also clear that for the ROSIS University of Pavia scene, all spatial–spectral methods perform better than SVM and MOM yields the best classification accuracies among all these methods. The numerical results reported on Table IV can be confirmed by visual inspection of the classification maps shown in Fig. 14.

G. Different Numbers of Training Samples

In this set of experiments, we analyze how the number of training samples affects the classification performance of the classification methods compared in Section III-F for the two considered datasets. Here, DMSR is not included since it is the coarse version of MOM. The parameters of these methods are fixed to be the same as those used in Section III-F. For the Indian Pines dataset, in each test we randomly choose 1–20% of the labeled pixels per class for training and the remaining pixels for testing. For very small classes, we take a minimum of two training samples per class. For the ROSIS University of Pavia scene, we build training sets by randomly choosing 10, 20, 40, 60, 80, and 100 training samples per class. Fig. 15(a) and (b) shows the classification accuracies obtained in this set of experiments with the AVIRIS Indian Pines and ROSIS University of Pavia scenes, respectively. From Fig. 15, it is obvious that the OA increases monotonically, and the standard deviation decreases as the number of training samples increases. The proposed MED and MOM consistently yield higher OAs than the other methods considered in experiments.

IV. Conclusion and Future Lines

This paper developed a RMSR technique for extracting the spatial information of hyperspectral images and hereby presented a spatial–spectral classification method. The RMSR technique allows using dense scales to represent the spatial characteristics of the hyperspectral images, so that we can benefit from the complementary information collected at various scales to capture the spatial information around a pixel. The spatial features are extracted by computing criteria on the dense rectangular scales. After obtaining the spatial features on dense scales, we concatenate them as high-dimensional multiscale features to keep all the spatial information extracted in the process. Then, a very sparse random measurement matrix is introduced to compress the high-dimensional spatial features into lower dimensional features without the loss of salient information. The entire process of RMSR is organically performed by computing criteria at random scales of random bands according to the nonzero entries of the very sparse measurement matrix. The criteria investigated in RMSR are the first two moments, which are simple and can be significantly accelerated by the integral image method. Thanks to these effective mechanisms, the computation load of RMSR becomes very light. The final classification is conducted by a CK-based approach that appropriately weights the spatial features with regards to spectral features. The proposed method has been tested on two widely used hyperspectral datasets. Our experiments indicate that use of dense scales by the proposed method can be very helpful to improve the obtained classification results, and RMSR is computationally manageable and highly effective. Although the results obtained by the proposed method are very encouraging, further enhancements such as the use of adaptive neighborhoods and additional criteria to describe spatial neighborhoods should be pursued in future developments.

Acknowledgment

The authors would like to thank Prof. D. Landgrebe from Purdue University for providing the Indian Pines dataset, and Prof. P. Gamba from the University of Pavia, Italy, for providing the ROSIS University of Pavia dataset.

References

[1] 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.

[2] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[3] 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.

[4] P. Zhong, P. Zhang, and R. Wang, "Dynamic learning of SMLR for feature selection and classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 280–284, Apr. 2008.

[5] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.

[6] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[7] J. Liu, Z. Wu, Z. Wei, L. Xiao, and L. Sun, "Spatial–spectral kernel sparse representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2462–2471, Dec. 2013.

[8] 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.

[9] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial–spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recog.*, vol. 45, no. 1, pp. 381–392, 2012.

[10] 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.

[11] 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.

[12] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[13] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[14] 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.

[15] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.

[16] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.

[17] J. Li *et al.*, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.

[18] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM-and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[19] 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.

[20] 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.

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

[22] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.

[23] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, Sep. 2006.

[24] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, 2010.

[25] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy," *Int. J. Remote Sens.*, vol. 30, no. 12, pp. 3205–3221, 2009.

[26] J. A. Dos Santos, P.-H. Gosselin, S. Philipp-Foliguet, R. d. S. Torres, and A. X. Falao, "Multiscale classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3764–3775, Oct. 2012.

[27] G. Moser, S. B. Serpico, and J. A. Benediktsson, "Land-cover mapping by Markov modeling of spatial-contextual information in very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 631–651, Mar. 2013.

[28] M. Pedergnana, P. R. Marpu, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "A novel technique for optimal feature selection in attribute profiles based on genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, Jun. 2013.

[29] M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of hyperspectral images with extended attribute profiles and feature extraction techniques," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 76–79.

[30] K. Zhang, L. Zhang, and M. Yang, "Fast compressive tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2002–2015, Oct. 2014.

[31] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2001, vol. 1, pp. 511–518.

[32] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Proc. Conf. Mod. Anal. Probability*, 1984, pp. 189–206.

[33] D. Achlioptas, "Database-friendly random projections: Johnson–Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, 2003.

[34] P. Li, T. J. Hastie, and K. W. Church, "Very sparse random projections," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2006, pp. 287–296.

[35] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2001, pp. 245–250.

[36] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[37] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[38] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[39] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[40] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[41] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. Atli Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

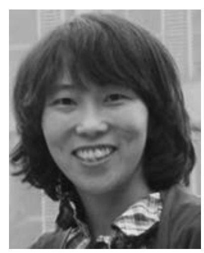
[42] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm

Jianjun Liu (M'15) received the B.S. degree in applied mathematics and the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology, Nanjing, China, in 2009 and 2014, respectively.

He is currently a Lecturer at the School of Internet of Things Engineering, Jiangnan University, Wuxi, China. His research interests include the area of spectral unmixing, hyperspectral image classification, image processing, sparse representation, and compressive sensing.

Zebin Wu (M'13) received the B.S. and Ph.D. degrees in computer science from Nanjing University of Science and Technology (NUST), Nanjing, China, in 2003 and 2008, respectively.

He is currently a Professor at the School of Computer Science, NUST. His research interests include hyperspectral image processing, high-performance computing, and computer simulation.



Jun Li (M'13) received the B.S. degree in geographic information systems from Hunan Normal University, Hunan, 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 the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisboa, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotcnica e de Computadores and the Instituto de Telecomunicações, IST, 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 at the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Badajoz, Spain. She is currently a Professor at Sun Yat-Sen University, Guangzhou, China. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing. She has been a Reviewer for several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the *Pattern Recognition*, the *Optical Engineering*, the *Journal of Applied Remote Sensing*, and the *Inverse Problems and Imaging*.

Dr. Li received the 2012 Best Reviewer Award of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Liang Xiao (M'11) received the B.S. degree in applied mathematics and the Ph.D. degree in computer science from Nanjing University of Science and Technology (NUST), Nanjing, China, in 1999 and 2004, respectively.

From 2009 to 2010, he was a Postdoctoral fellow at Rensselaer Polytechnic Institute, USA. He is currently a Professor at the School of Computer Science, NUST. His main research interests include inverse problems in image processing, scientific computing, data mining, and pattern recognition.

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

He is an Associate Professor (with accreditation for Full Professor) at the Department of Technology of Computers and Communications, University of Extremadura, where he is also the Head of the Hyperspectral Computing Laboratory (HyperComp). He was the Coordinator of the Hyperspectral Imaging Network. He authored more than 400 publications, including 119 Journal Citation Reports (JCR) journal papers (71 in IEEE journals), 20 book chapters, and more than 240 peer-reviewed conference proceeding papers (94 in IEEE conferences). He was an Associate Editor for IEEE ACCESS, and a Member of the Editorial Board of theIEEE GEOSCIENCE ANDREMOTE SENSING NEWSLETTER during 2011–2012 and theIEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE in 2013. He was also a Member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS andREMOTE SENSING in 2012. He was the Director of Education Activities for theIEEE GEOSCIENCE AND REMOTE SENSING SOCIETY (GRSS) from 2011 to 2012, and since November 2012, he has been the President of the Spanish Chapter of IEEE GRSS. Since January 2013, he has been the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE ANDREMOTE SENSING JOURNAL.

Dr. Plaza has guest edited seven special issues on JCR journals (three in IEEE journals). He has been the Chair for the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, in 2011. 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 TRANSACTIONSONGEOSCIENCE AND REMOTE SENSING in 2010, a journal for which he has served as an Associate Editor from 2007 to 2012.

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

He is currently a Rector and a Professor of electrical and computer engineering at the University of Iceland. He is the cofounder of the biomedical start-up company Oxymap. His research interests include remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing. He has published extensively in those fields. He is a Fellow of The International Society of Optics and Photonics and a Member of the Association of Chartered Engineers in Iceland (VFI) and Societas Scinetiarum Islandica. During 2011–2012, he was the President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 2000. He has been an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE and REMOTE SENSING (TGRS) since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and the IEEE ACCESS since 2013. He was an Editor of TGRS from 2003 to 2008. He was on the International Editorial Board of the *International Journal of Image and Data Fusion* and was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2007 to 2010.

Prof. Benediktsson received the Stevan J. Kristof Award from Purdue University as outstanding graduate student in remote sensing in 1991, the Icelandic Research Councils Outstanding Young Researcher Award in 1997, and the IEEE Third Millennium Medal in 2000. He was a corecipient of the University of Iceland's Technology Innovation Award in 2004, the 2012 TGRS Paper Award, and the IEEE GRSS Highest Impact Paper Award in 2013. He also received the yearly research award from the Engineering Research Institute, University of Iceland in 2006, the Outstanding Service Award from the IEEE GRSS in 2007, and the IEEE/VFI Electrical Engineer of the Year Award in 2013. He is also a Member of Tau Beta Pi.