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

Classification of Hyperspectral Images by Exploiting Spectral–Spatial Information of Superpixel via Multiple Kernels

Leyuan Fang, *Member, IEEE*, Shutao Li, *Senior Member, IEEE*, Wuhui Duan, Jinchang Ren, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—For the classification of hyperspectral images (HSIs), this paper presents a novel framework to effectively utilize the spectral–spatial information of superpixels via multiple kernels, which is termed as superpixel-based classification via multiple kernels (SC-MK). In the HSI, each superpixel can be regarded as a shape-adaptive region, which consists of a number of spatial neighboring pixels with very similar spectral characteristics. First, the proposed SC-MK method adopts an oversegmentation algorithm to cluster the HSI into many superpixels. Then, three kernels are separately employed for the utilization of the spectral information, as well as spatial information, within and among superpixels. Finally, the three kernels are combined together and incorporated into a support vector machine classifier. Experimental results on three widely used real HSIs indicate that the proposed SC-MK approach outperforms several well-known classification methods.**

***Index Terms*—Hyperspectral image (HSI), multiple kernels, spectral–spatial image classification, superpixel, support vector machines (SVMs).**

I. Introduction

HYPERSPECTRAL imaging has been widely used in the remote sensing, which can acquire images from hundreds of narrow contiguous bands, spanning the visible-to-infrared spectrum. In the hyperspectral image (HSI), each pixel is a high-dimensional vector, and its entries represent the spectral responses of different spectral bands. The highly informative spectral information of the HSI pixels has many applications, such as classification [1], target detection [2], anomaly detection [3], spectral unmixing [4], and others [5].

In the last decades, HSI classification has been a very active research topic in the remote sensing. Given a representative training set for each class, the objective of the classification is to assign each pixel to one of the classes, based on its spectral characteristics. To achieve this, many discriminative approaches have been developed. Among these, the support vector machine (SVM) [6], [7] and multinomial logistic regression (MLR) [8]–[10] have demonstrated to be very powerful. Dynamic or random subspace approaches [3], [11], which are new versions of random forest and exploit the inherent subspace structure of hyperspectral, have proved to be an effective way for analyzing and classifying HSIs. The sparse representation [12]–[16], which can sparsely decompose the input pixel on an overcomplete dictionary, is another widely used classifier. Recently, metric learning [17] has also been successfully explored in HSI processing, which has formulated a novel and adaptive metric learning method for classification and object recognition. In addition, some other classification approaches have focused on the design of effective feature extraction or reduction techniques, such as the principle component analysis (PCA) [18], clonal selection feature selection [19], kernel discriminative analysis [20], and semisupervised discriminative locally enhanced alignment [21]. Note that the kernel [22] has been widely used in the aforementioned approaches, since it can improve the class separability [23].

Although the aforementioned approaches can effectively utilize the spectral information, their classification results often appear very noisy. This is mainly due to the two facts that the number of reference training samples is often very limited and the spectral information of pixels from one class may be easily mixed with that of pixels from other classes. To further improve the classification performance, some recent works attempt to use both the spectral information and spatial information of the HSI, which is based on the assumption that pixels from a local spatial region should have very similar spectral characteristics and thus correspond to the same materials. In [24] and [25], the extended morphological profiles (EMPs) is used to exploit the spatial information, which can effectively improve the estimation. In [26], the spatial dependence of pixels within a local region is exploited by a postprocessing procedure on each individual pixel label. In [12], the spatial

Manuscript received March 26, 2015; revised May 19, 2015; accepted June 5, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61172161, by the National Natural Science Foundation for Distinguished Young Scholars of China under Grant 61325007, and by the Fundamental Research Funds for the Central Universities of Hunan University.

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

J. Ren is with the Centre for Excellence in Signal and Image Processing, Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, U.K. (e-mail: jinchang.ren@strath.ac.uk).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavík 107, Iceland (e-mail: benedikt@hi.is).

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

Digital Object Identifier 10.1109/TGRS.2015.2445767

0196-2892 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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

information is incorporated into the sparse representation technique by a joint sparse norm on pixels within a local region. In [27], multiple kernels are utilized for the exploration of the spatial information, which is modeled as the mean and variance of pixels within a local region. Although the aforementioned works [12], [26], [27] can provide promising classification accuracy, the size and shape of the adopted spatial region is fixed, and thus, the spatial context in the HSI may not be sufficiently exploited. That is, the shape of the regions should be changed according to different spatial structures of the HSI. For example, large region sizes should be selected for the smooth area, while heterogeneous spatial area requires small region sizes.

In computer vision, *superpixel* has been extensively investigated to facilitate visual recognition [28]–[30]. Each superpixel is a local region, whose size and shape can be adaptively adjusted according to local structures. In this paper, we introduce the superpixel in HSI classification and propose to effectively exploit both the spectral and spatial information of the superpixel via multiple kernels, which is denoted as the superpixel-based classification via multiple kernels (SC-MK). First, the SC-MK adopts an efficient oversegmentation algorithm [30] to cluster the HSI into many superpixels. Then, mean filtering and weighted average filtering are utilized to extract the spatial features within and among superpixels. Subsequently, three kernels are separately computed on the pixels extracted from the original spectral feature and the spatial features within and among superpixels. Finally, the three kernels are combined and incorporated into the SVM classifier. Note that, in some very recent works [31], [32], the HSI is also classified based on the superpixel. These works [31], [32] use histogram descriptors or K-means clustering for the classification, and do not consider the correlations among superpixels. In addition, some works have recently applied the multiple kernels for HSI classification [33], [34]. In [33], multiple kernels have been incorporated into a domain adaptation framework to reduce the bias between the source and the target domains. The work in [34] employed a learning algorithm to adaptively select multiple representative kernels for classification. In contrast, the proposed SC-MK method adopts multiple kernels to effectively capture spectral–spatial information within and among superpixels, which is different from the aforementioned superpixel-based and multiple-kernel-based works [31]–[34].

The rest of this paper is organized as follows: In Section II, the SVMs with multiple kernels for HSI classification are briefly reviewed. Section III introduces the proposed SC-MK method. Experimental results on three well-known HSI data are presented in Section IV. Section V concludes this paper and suggests future works.

II. SVMs With Multiple Kernels

SVM is a supervised learning model, and its objective in HSI classification is to find a decision rule, which can determine the class label for each test pixel [6], [7]. Since a real HSI pixel is linearly nonseparable, the SVM usually adopts a kernel function to map pixels into the high-dimensional feature spaces. To be specific, let $(\mathbf{y}_1, \ldots, \mathbf{y}_N) \in \mathbb{R}^M$ denote the training pixels and $(c_1, \ldots, c_N)$ represent the corresponding class labels. The SVM aims to solve the following problem:

$$\min_{\mathbf{w},\beta_i,b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + T\sum_i \beta_i \right\} \text{ subject to}$$
$$c_i\left(\langle \phi(\mathbf{y}_i), \mathbf{w}\rangle + b\right) \geq 1 - \beta_i, \qquad \forall i = 1, \ldots, N,$$
$$\beta_i \geq 0, \qquad \forall i = 1, \ldots, N \quad (1)$$

where $\mathbf{w}$ and b define the classifier in the feature space, β_i are the slack variables for the nonseparability of data, and T is a regularization parameter that controls the generalization ability of the classifier. $\phi(\bullet)$ is a mapping function, which transforms the input pixel $\mathbf{y}_1 \in \mathbb{R}^M$ into a higher dimensional feature space $\phi(\mathbf{y}_i) \in \mathbb{R}^{M^*}(M^* > M)$. Since the mapping $\phi(\bullet)$ in the SVM learning is represented by the inner function, a kernel function K can be defined by

$$K(\mathbf{y}_i, \mathbf{y}_j) = \langle \phi(\mathbf{y}_i), \phi(\mathbf{y}_j)\rangle . \quad (2)$$

Then, we can construct a nonlinear SVM by the kernel function, without considering the mapping $\phi(\bullet)$ explicitly. The most widely used kernel is the radial basis function (RBF) kernel, which is computed as

$$K(\mathbf{y}_i, \mathbf{y}_j) = \exp\left(-\|\mathbf{y}_i - \mathbf{y}_j\|^2/2\sigma^2\right). \quad (3)$$

Then, by incorporating (2) into (1), we solve a dual Lagrangian problem and obtain the decision rule for any test pixels as follows:

$$f(\mathbf{y}) = \sum_{i=1}^{N} c_i \alpha_i K(\mathbf{y}_i, \mathbf{y}) + b \quad (4)$$

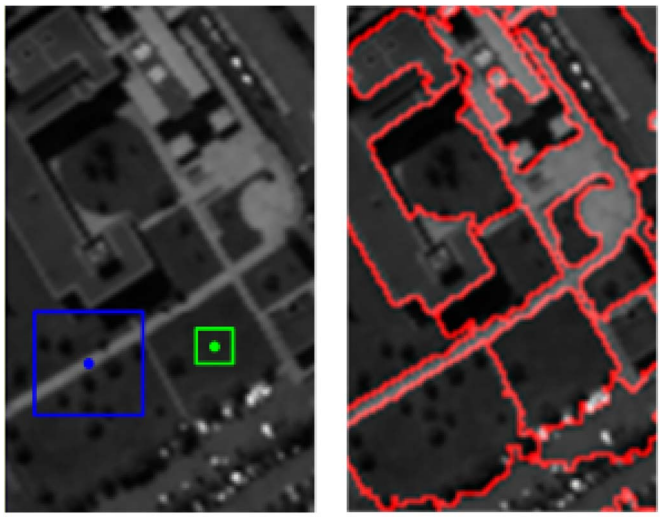
where α_i are the Lagrange multipliers in (1), which can be estimated by quadratic programming methods [35].

If the RBF kernel is created from the original spectral pixels $(\mathbf{y}_1^{\text{Spec}}, \ldots, \mathbf{y}_N^{\text{Spec}})$, the corresponding kernel is denoted as the spectral kernel $K_{\text{Spec}}(\mathbf{y}_i^{\text{Spec}}, \mathbf{y}_j^{\text{Spec}})$. In [27], to further exploit the spatial information of the HSI, one spatial region (of fixed size) is defined for each pixel in the HSI, and the mean or variance is computed for pixels within each region as the spatial feature. Then, a new RBF kernel can be computed on the pixels $(\mathbf{y}_1^{\text{Spat}}, \ldots, \mathbf{y}_N^{\text{Spat}})$ from the spatial feature, and it is referred to as the spatial kernel $K_{\text{Spat}}(\mathbf{y}_i^{\text{Spat}}, \mathbf{y}_j^{\text{Spat}})$. Finally, a composite kernel can be effectively computed by a weighted average of these two kernels as follows:

$$K_{\text{CW}}(\mathbf{y}_i, \mathbf{y}_j) = \mu_{\text{Spec}} K_{\text{Spec}}\left(\mathbf{y}_i^{\text{Spec}}, \mathbf{y}_j^{\text{Spec}}\right) + \mu_{\text{Spat}} K_{\text{Spat}}\left(\mathbf{y}_i^{\text{Spat}}, \mathbf{y}_j^{\text{Spat}}\right) \quad (5)$$

where μ_{Spec} and μ_{Spat} are the weights for the spectral kernel $K_{\text{Spec}}(\mathbf{y}_i^{\text{Spec}}, \mathbf{y}_j^{\text{Spec}})$ and spatial kernel $K_{\text{Spat}}(\mathbf{y}_i^{\text{Spat}}, \mathbf{y}_j^{\text{Spat}})$, respectively, and $\mu_{\text{Spec}} + \mu_{\text{Spat}} = 1$. The composite kernel $K_{\text{CW}}(\mathbf{y}_i, \mathbf{y}_j)$ can be directly incorporated into (4) to create a new decision rule for classification. Compared to the single

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



(a) (b)

Fig. 1. Spatial region selection by (a) fixed-size rectangles and (b) adaptive-size superpixels.

spectral kernel, the composite kernel considers the spatial information and, thus, can enhance the HSI classification performance. However, since the size of the adopted spatial region is fixed, the spatial information still may not be sufficiently exploited. For example, if the region size for the test pixel is selected too large in the detailed region [see blue region in Fig. 1(a)], some pixels uncorrelated to the test pixel might be included, thus deteriorating the classification accuracy. In contrast, if the region size is chosen too small for the pixel in the smooth region [see green region in Fig. 1(a)], the spatial information cannot be sufficiently exploited for the classification.

III. SC-MK

In computer vision, superpixels have been studied to provide an efficient representation, which can facilitate visual recognition [28]–[30]. Each superpixel is a perceptually meaningful region, whose shape and size can be adaptively changed according to different spatial structures [see an example in Fig. 1(b)] [29]. In this paper, the proposed SC-MK algorithm extends the superpixel for HSI classification and adopts the multiple kernels to effectively exploit spectral–spatial information within and among each superpixel. In general, the proposed SC-MK algorithm consists of two parts: 1) creation of superpixels in HSI and b) exploration of spectral–spatial information of superpixels via multiple kernels, which will be described in the following two subsections.

A. Creation of Superpixels in HSI

Unlike the single-band gray or three-band color image, the HSI usually has hundreds of spectral bands. To improve the computational efficiency, PCA [36] is first used to reduce the spectral bands of the HSI. Since the important information of the HSI exists in the principle components (e.g., first three principle components), they are used as the base images.

Second, the superpixel number L is selected based on the complexness of the structural texture in the HSI. Specifically, the Sobel filter [37], which is a simple texture detector, is first applied on the base images. Then, the number of nonzero elements in the filtered images is compared with the total number of pixels in the base images to create a texture ratio R_{texture}, which reflects the complexness of texture in the HSI. Finally, the superpixel number L is selected by the texture ratio R_{texture} and a predefined base superpixel number L_{base}, i.e.,

$$L = L_{\text{base}} \times R_{\text{texture}}. \tag{6}$$

Note that other advanced texture detectors might be used to enhance the performance, but this will involve more computational cost.

Third, given the superpixel number L, an oversegmentation algorithm called entropy rate superpixel (ERS) [30] is applied on the base images to generate a 2-D superpixel map. Specifically, the ERS is a graph-based clustering algorithm, which first constructs a graph $G = (V, E)$ on the base images. V is the vertex set corresponding to pixels of the base images, and E is the edge set representing the pairwise similarities between adjacent pixels. Then, the ERS segments the graph into L connected subgraphs (each corresponds to a superpixel) by selecting a subset of edges $A \subseteq E$. To form the compact, homogeneous, and balanced superpixels, an entropy rate term $H(\bullet)$ and a balancing term $B(\bullet)$ are incorporated into the objective function of the superpixel segmentation as follows:

$$\max_{A} \{H(A) + \lambda B(\bullet)\} \qquad \text{subject to } A \subseteq E \tag{7}$$

where $\lambda \geq 0$ is the weight for controlling the contribution of the entropy rate term and balancing term. In fact, the optimization problem in (7) can be efficiently solved by a greedy algorithm, as introduced in [38].

Finally, for the 2-D superpixel map, the position indexes of pixels within each superpixel can be obtained. Then, the position indexes for L superpixels in the 2-D map can be applied on the original HSI to extract the corresponding L nonoverlapping 3-D superpixels. The procedure for the creation of the superpixels in HSI is illustrated in Fig. 2.

B. Exploration of Spectral–Spatial Information of Superpixels via Multiple Kernels

Here, we will first introduce how to utilize the superpixels to create three feature images, which separately reflect the spectral information and spatial information within and among superpixels. Then, three kernels are computed on the pixels from the feature images to exploit the spectral–spatial information of superpixels.

Each superpixel is a group of neighboring spectral pixels $\mathbf{y}_i^z, z = 1, \ldots, Z$, which can be transformed into a matrix $\mathbf{Y}_i^{\text{SP}}$. As described in Section II, spectral pixels representing the spectral information of superpixels in the HSI can be directly used as the spectral feature. All the spectral pixels in the HSI constitute the spectral feature image $\mathbf{I}^{\text{Spec}}$.

To exploit the spatial information within each superpixel, a mean operation is first applied on the spectral pixels $[\mathbf{y}_i^1, \ldots, \mathbf{y}_i^Z]$ within each superpixel $\mathbf{Y}_i^{\text{SP}}$, and then, the mean pixel $\mathbf{y}_i^{\text{Mean}}$ is assigned to all pixels in each superpixel. Here, $\mathbf{Y}_i^{\text{SP}}$ is

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

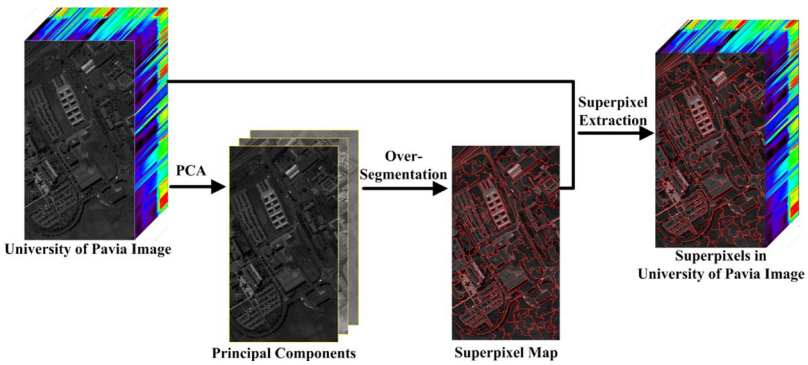


Fig. 2. Procedure for the creation of superpixels in HSI.

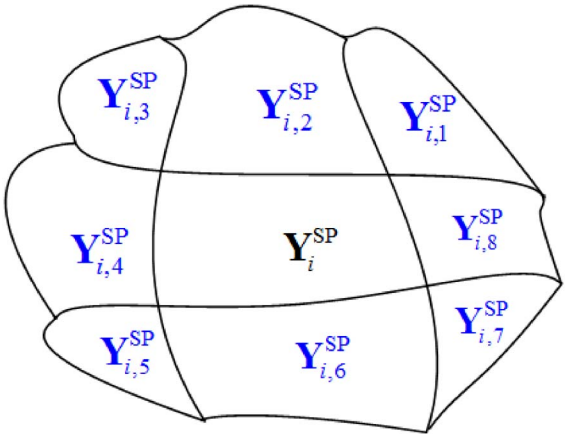


Fig. 3. Example showing the current processed superpixel $\mathbf{Y}_i^{\mathrm{SP}}$ and its neighboring superpixels $\mathbf{Y}_{i,1}^{\mathrm{SP}}, \ldots, \mathbf{Y}_{i,8}^{\mathrm{SP}}$.

still the superpixel, which consists of a number of spectral pixels. This operation is the same as mean filtering (which is also adopted in the work [39]) and can reduce the interferences (e.g., noise) in each superpixel. All the filtered superpixels can constitute a mean feature image $\mathbf{I}^{\mathrm{Mean}}$. Note that adopting other more powerful filtering approaches (e.g., guided filtering [40] and nonlocal filtering [41]) might enhance the performance but increase the computational cost.

To exploit the spatial information among superpixels, a weighted average operation is conducted on the neighboring superpixels $\mathbf{Y}_{i,j}^{\mathrm{SP}}, j = 1, \ldots, J$ of the current processed superpixels $\mathbf{Y}_i^{\mathrm{SP}}$ (see an example in Fig. 3), where J is the number of neighboring superpixels. Since the mean pixel is the representative feature of each superpixel, the weighted average operation can also be applied on the mean pixels $\mathbf{y}_{i,j}^{\mathrm{Mean}}$, $j = 1, \ldots, J$ of neighboring superpixels, and a weighted average pixel can be obtained by

$$\mathbf{y}_i^{\mathrm{WA}} = \sum_{j=1}^{J} w_{i,j} \times \mathbf{y}_{i,j}^{\mathrm{Mean}} \tag{8}$$

where $w_{i,j}$ is the weight, which is estimated as [42], [43]

$$w_{i,j} = \frac{\exp\left(-\left\|\mathbf{y}_{i,j}^{\mathrm{Mean}} - \mathbf{y}_i^{\mathrm{Mean}}\right\|_2^2 \Big/ h\right)}{\mathrm{Norm}}. \tag{9}$$

In (9), Norm is defined as $\sum_{j=1}^{J} \exp(-\|\mathbf{y}_{i,j}^{\mathrm{Mean}} - \mathbf{y}_i^{\mathrm{Mean}}\|_2^2/h)$, and h is a predefined scalar. Then, the $\mathbf{y}_i^{\mathrm{WA}}$ is assigned to all pixels in each superpixel $\mathbf{Y}_i^{\mathrm{SP}}$, and all the superpixels constitute a weighted average feature image $\mathbf{I}^{\mathrm{Weigh}}$.

In the training stage, first, a set of spectral pixels $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$ are randomly (or manually) selected from the original HSI. Then, the position indexes for these selected pixels are used to extract pixels from the spectral feature image $\mathbf{I}^{\mathrm{Spec}}$, the mean feature image $\mathbf{I}^{\mathrm{Mean}}$, and the weighted average feature image $\mathbf{I}^{\mathrm{Weigh}}$, respectively. The extracted pixels can separately constitute the corresponding spectral feature training data $(\mathbf{y}_1^{\mathrm{Spec,\ Train}}, \ldots, \mathbf{y}_N^{\mathrm{Spec,\ Train}})$, mean feature training data $(\mathbf{y}_1^{\mathrm{Mean,\ Train}}, \ldots, \mathbf{y}_N^{\mathrm{Mean,\ Train}})$, and weighted average feature training data $(\mathbf{y}_1^{\mathrm{Weigh,\ Train}}, \ldots, \mathbf{y}_N^{\mathrm{Weigh,\ Train}})$. Subsequently, the RBF kernel function in (3) can be applied on the three kinds of training data to compute a spectral kernel $K_{\mathrm{Spec}}^{\mathrm{Train}} \times (\mathbf{y}_i^{\mathrm{Spec,\ Train}}, \mathbf{y}_j^{\mathrm{Spec,\ Train}})$, an intrasuperpixel spatial kernel $K_{\mathrm{IntraS}}^{\mathrm{Train}}(\mathbf{y}_i^{\mathrm{IntraS,\ Train}}, \mathbf{y}_j^{\mathrm{IntraS,\ Train}})$, and an intersuperpixel spatial kernel $K_{\mathrm{InterS}}^{\mathrm{Train}}(\mathbf{y}_i^{\mathrm{InterS,\ Train}}, \mathbf{y}_j^{\mathrm{InterS,\ Train}})$, as follows:

$$\begin{aligned}
&K_{\mathrm{Spec}}^{\mathrm{Train}}\left(\mathbf{y}_i^{\mathrm{Spec,\ Train}}, \mathbf{y}_j^{\mathrm{Spec,\ Train}}\right) \\
&\quad = \exp\left(-\left\|\mathbf{y}_i^{\mathrm{Spec,\ Train}} - \mathbf{y}_j^{\mathrm{Spec,\ Train}}\right\|^2 \Big/ 2\sigma^2\right) \\
&K_{\mathrm{IntraS}}^{\mathrm{Train}}\left(\mathbf{y}_i^{\mathrm{IntraS,\ Train}}, \mathbf{y}_j^{\mathrm{IntraS,\ Train}}\right) \\
&\quad = \exp\left(-\left\|\mathbf{y}_i^{\mathrm{Mean,\ Train}} - \mathbf{y}_j^{\mathrm{Mean,\ Train}}\right\|^2 \Big/ 2\sigma^2\right) \\
&K_{\mathrm{InterS}}^{\mathrm{Train}}\left(\mathbf{y}_i^{\mathrm{InterS,\ Train}}, \mathbf{y}_j^{\mathrm{InterS,\ Train}}\right) \\
&\quad = \exp\left(-\left\|\mathbf{y}_i^{\mathrm{Weigh,\ Train}} - \mathbf{y}_j^{\mathrm{Weigh,\ Train}}\right\|^2 \Big/ 2\sigma^2\right).
\end{aligned} \tag{10}$$

Then, the aforementioned three kernels are combined by a weighted average, i.e.,

$$\begin{aligned}
&K_{\mathrm{CompSup}}^{\mathrm{Train}}(\mathbf{y}_i, \mathbf{y}_j) \\
&\quad = \mu_{\mathrm{Spec}} K_{\mathrm{Spec}}^{\mathrm{Train}}\left(\mathbf{y}_i^{\mathrm{Spec,\ Train}}, \mathbf{y}_j^{\mathrm{Spec,\ Train}}\right) \\
&\qquad + \mu_{\mathrm{IntraS}} K_{\mathrm{IntraS}}^{\mathrm{Train}}\left(\mathbf{y}_i^{\mathrm{IntraS,\ Train}}, \mathbf{y}_j^{\mathrm{IntraS,\ Train}}\right) \\
&\qquad + \mu_{\mathrm{InterS}} K_{\mathrm{InterS}}^{\mathrm{Train}}\left(\mathbf{y}_i^{\mathrm{InterS,\ Train}}, \mathbf{y}_j^{\mathrm{InterS,\ Train}}\right)
\end{aligned} \tag{11}$$

where μ_{Spec}, μ_{IntraS}, and μ_{InterS} are the weights for the three different kernels, respectively, and $\mu_{\mathrm{Spec}} + \mu_{\mathrm{IntraS}} + \mu_{\mathrm{InterS}} = 1$. The composite kernel $K_{\mathrm{CompSup}}^{\mathrm{Train}}(\mathbf{y}_i, \mathbf{y}_j)$ can be incorporated into (4) to create a decision rule.

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

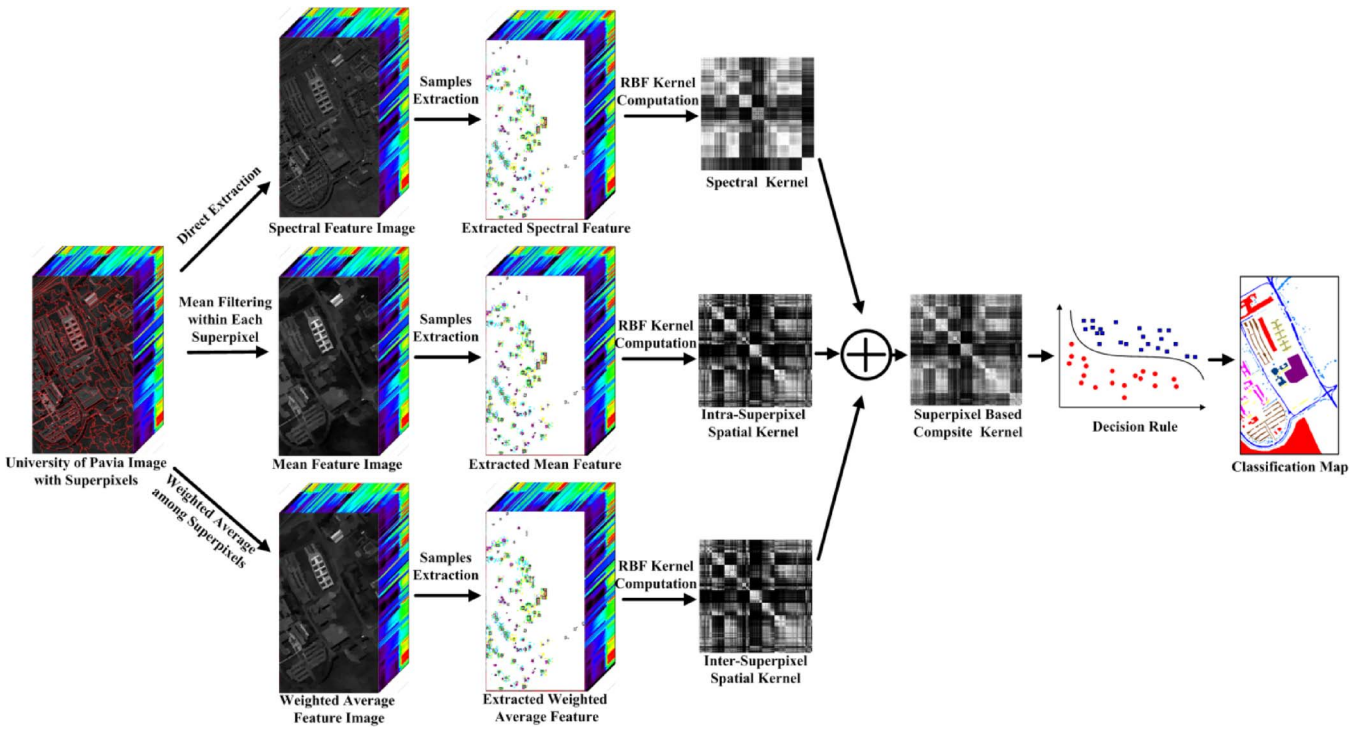


Fig. 4. Schematic for utilizing spectral–spatial information of superpixels via multiple kernels.

In the test stage, for the classification of one pixel in the HSI, as indicated in (4), the kernel transformation (same as the aforementioned training stage) requires to be first applied, and then, the decision rule can determine the class label for this pixel. A schematic for utilizing the spectral–spatial information of superpixels via multiple kernels is illustrated in Fig. 4.

IV. Experimental Results

To verify the effectiveness of the proposed SC-MK algorithm, it is tested on three real hyperspectral datasets[1]: Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines image, AVIRIS Salinas image, and Reflective Optics System Imaging Spectrometer (ROSIS-03) University of Pavia image. The performance of the proposed SC-MK algorithm is compared with those of seven competing classification algorithms: SVM [7], EMP [24], SVM-composite kernel (SVM-CK) [27], logistic regression via variable splitting and augmented Lagrangian-multilevel logistic (LORSAL-MLL) [44], sparse representation-based classification (SRC) [12], MLR-generalized composite kernel (MLR-GCK) [25], and IntraSC-MK. The IntraSC-MK is a simplified version of the proposed SC-MK method, which only exploits the spectral–spatial information within each superpixel for the HSI classification. The SVM classifier does not consider the spatial information, which was implemented with the spectral-only Gaussian kernel and functions in the library for support vector machines (LIBSVM) library [45]. For EMP and LORSAL-MLL, the spatial context of the HSI was utilized by the EMP and the multilevel logistic prior-based segmentation technique, respectively. For SRC and SVM-CK, the spatial information within a fixed-size local region is utilized by the joint sparse regularization and composite kernel, respectively. In the MLR-GCK, the spatial context of the HSI was exploited by the extended multiattribute profile [46] and the generalized composite kernel.

[1]Datasets can be downloaded at: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

A. Data Set Description

The Indian Pines image, which captures the agricultural Indian Pines test site of Northwestern Indiana, was acquired by the AVIRIS sensor. The image is of size $145 \times 145 \times 220$, which has a spatial resolution of 20 m per pixel and a spectral coverage ranging from 0.2 to 2.4 μm. Before the classification, 20 water absorption bands [47] are discarded. Fig. 5(a) and (b) shows the color composite of the Indian Pines image and the corresponding reference data, which contains 16 reference classes from different types of crops (e.g., corns, soybeans, and wheat).

The Salinas image was also captured by the AVIRIS sensor over the area of Salinas Valley, California. The image is of size $512 \times 217 \times 224$, which has a spatial resolution of 3.7 m per pixel. Same as in the Indian Pines image, 20 water absorption spectral bands are removed. Fig. 6(a) and (b) shows the color composite of the Salinas image and the corresponding reference data, which contains 16 different classes.

The University of Pavia image was acquired with the ROSIS-03 sensor over the campus at the University of Pavia, Italy. The image is of size $610 \times 340 \times 120$, with a spatial resolution of 1.3 m per pixel and a spectral coverage ranging from 0.43 to 0.86 μm [48]. Before the classification, 12 spectral bands were removed due to high noise. Fig. 7(a) and (b) shows the color composite of the University of Pavia image and the corresponding reference data, which considers nine classes of interest.

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

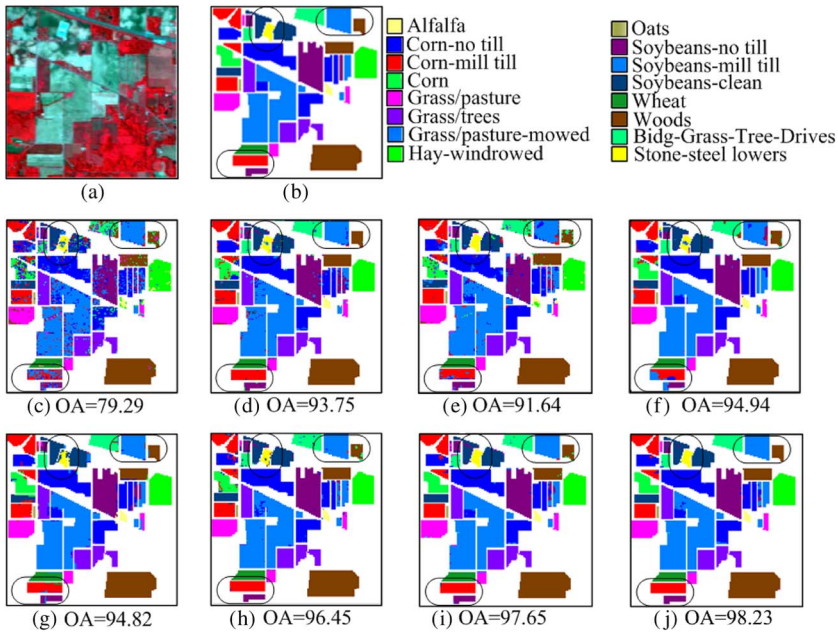


Fig. 5. Indian Pines image. (a) Three-band color composite image. (b) Reference image and the classification results (OA in %) obtained by the (c) SVM [7], (d) EMP [24], (e) SVM-CK [27], (f) LORSAL-MLL [44], (g) SRC [12], (h) MLR-GCK [25], (i) IntraSC-MK, and (j) SC-MK methods.

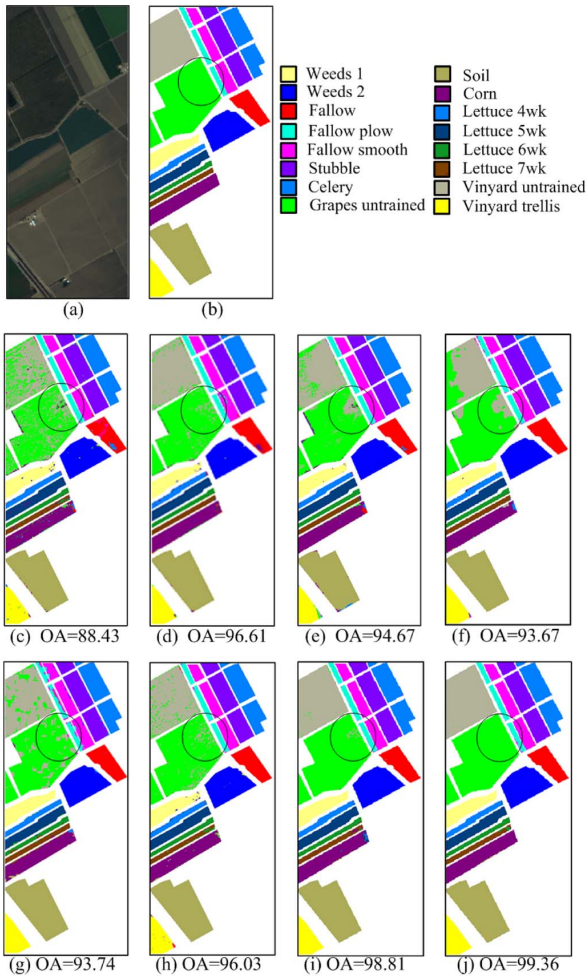


Fig. 6. Salinas image. (a) Three-band color composite image. (b) Reference image and the classification results (OA in %) obtained by the (c) SVM [7], (d) EMP [24], (e) SVM-CK [27], (f) LORSAL-MLL [44], (g) SRC [12], (h) MLR-GCK [25], (i) IntraSC-MK, and (j) SC-MK methods.

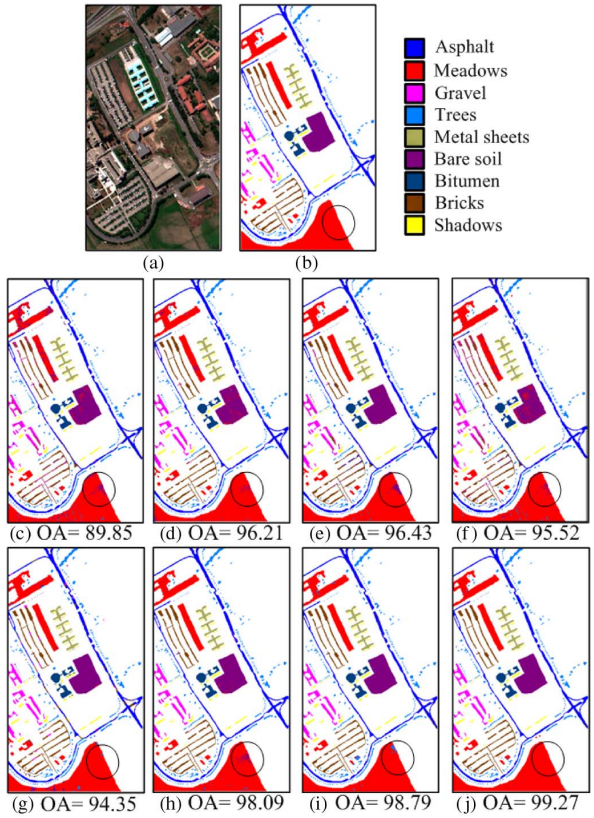


Fig. 7. University of Pavia image. (a) Three-band color composite image. (b) Reference image and the classification results (OA in %) obtained by the (c) SVM [7], (d) EMP [24], (e) SVM-CK [27], (f) LORSAL-MLL [44], (g) SRC [12], (h) MLR-GCK [25], (i) IntraSC-MK, and (j) SC-MK methods.

B. Classification Results

In the experiments, the parameters for the proposed SC-MK and IntraSC-MK methods are empirically selected and kept unchanged for the three test images. The base superpixel number

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

TABLE I
CLASSIFICATION ACCURACY OF INDIAN PINES IMAGE OBTAINED BY THE SVM [7], EMP [24], SVM-CK [27], LORSAL-MLL [44], SRC [12], MLR-GCK [25], INTRASC-MK, AND SC-MK METHODS. CLASS-SPECIFIC ACCURACY VALUES ARE IN PERCENTAGE

Class	Training/Test	SVM	EMP	SVM-CK	LORSAL-MLL	SRC	MLR-GCK	IntraSC-MK	SC-MK
Alfalfa	10/36	68.80	97.50	91.66	83.88	96.03	95.67	99.25	**100**
Corn-no till	143/1285	71.26	92.18	88.81	92.12	94.47	93.22	96.12	**97.11**
Corn-min till	83/747	73.91	88.47	86.66	89.05	92.35	95.92	97.08	**97.65**
Corn	24/213	62.28	79.24	83.38	95.58	92.55	94.00	96.93	**97.82**
Grass/Pasture	48/435	88.30	94.57	93.56	90.85	93.33	94.78	95.24	**96.38**
Grass/Trees	73/657	86.44	98.04	99.08	99.72	94.87	99.81	99.98	**100**
Grass/Pasture-mowed	10/18	88.07	61.24	93.33	92.22	88.88	98.14	97.24	**100**
Hay-windrowed	48/430	90.89	**100**	98.27	99.90	99.55	**100**	**100**	**100**
Oats	10/10	77.77	82.54	**100**	98.00	80.71	**100**	**100**	**100**
Soybeans-no till	97/875	74.42	92.57	86.66	91.86	91.93	91.17	**93.69**	93.35
Soybeans-min till	246/2209	78.79	92.58	92.10	95.89	96.36	97.91	98.48	**99.02**
Soybean-clean	59/534	69.31	88.76	83.80	97.15	90.61	95.13	95.88	**97.80**
Wheat	21/184	91.84	**100**	98.58	99.56	89.13	99.45	99.52	99.60
Woods	127/1138	92.60	99.24	97.82	97.66	98.21	99.39	99.70	**99.98**
Building-Grass-Trees-Drives	39/347	68.84	**98.50**	85.53	93.14	94.23	96.02	96.84	97.56
Stone-steel Towers	10/83	99.05	**99.13**	98.31	82.41	81.23	82.59	98.79	97.15
OA (Mean in %)		79.53	93.56	91.51	94.73	94.66	96.29	97.53	**98.06**
AA (Mean in %)		80.01	91.54	92.35	93.69	92.15	95.83	97.80	**98.34**
Kappa		0.77	0.93	0.90	0.94	0.94	0.96	0.97	**0.98**

L_{base} for the proposed SC-MK and IntraSC-MK methods is set to 800. The parameter h in (9) is set to 500. Setting a larger h will give more power to the neighboring superpixels, which improves the proposed method for large homogeneous regions while decreasing the classification accuracy in heterogeneous regions. When h is varied from 100 to 1000, the performance of the proposed method changes very slightly in the three test images. The σ in (10) is set to 1. As indicated in [27], the spatial kernel should be assigned with slightly larger weight, as compared with the spectral kernel. Therefore, for the IntraSC-MK method, the spectral kernel weight μ_{Spec} is set to 0.4, while the intrasuperpixel kernel weight μ_{IntraS} is set to 0.6. For the SC-MK method, the spectral kernel weight μ_{Spec}, intrasuperpixel kernel weight μ_{IntraS}, and intersuperpixel kernel weight μ_{InterS} are set to 0.2, 0.4, and 0.4, respectively. In the following subsection, the influences of the base superpixel number L_{base} and kernel weights to the performances of the proposed SC-MK approach will be further analyzed. For the SVM method, the parameters C and σ are obtained by fivefold cross-validation. For the EMP, MLR-GCK, and LORSAL-MLL methods, their parameters are set to the default values as in [24], [25], and [44]. The parameters for the SRC and SVM-CK methods are tuned to reach their best results in the experiments.

The first experiment was performed on the Indian Pines image. In this experiment, training samples are randomly selected to account for about 10% of the labeled reference data (see the second column in Table I). The visual classification results from different classifiers are shown in Fig. 5. As can be observed, the SVM classifier that only considers the spectral information exhibits very noisy estimations in its classification map. By utilizing the spatial context of the HSI, the EMP, LORSAL-MLL, SRC, SVM-CK, and MLR-GCK methods can deliver a smoother appearance in their classification results. However, these approaches cannot accurately classify pixels in the detailed and near-edge regions (e.g., ellipse regions in Fig. 5). By contrast, the proposed superpixel-based IntraSC-MK and SC-MK approaches not only provide a smoother appearance but also achieve more accurate estimations in the detailed area (e.g., ellipse regions in Fig. 5). For the quantitative comparison, our experiments adopt the overall accuracy (OA), average accuracy (AA), and Kappa coefficient as the metrics to evaluate the classification results. Quantitative results for various classifiers on the Indian Pines image are tabulated in Table I. Note that the classification accuracy values reported in Table I are the average results over ten experiments with different randomly selected training data. As can be seen, the proposed IntraSC-MK and SC-MK methods perform better than the other compared methods, in terms of OA, AA, and the Kappa coefficient. In addition, we can observe that the SC-MK method outperforms the IntraSC-MK method that only considers the correlations within each superpixel. This demonstrates that, in addition to the spectral–spatial information within each superpixel, further utilizing the spatial information among superpixels in the SC-MK method can enhance the classification performance.

The second and third experiments are conducted on the Salinas and University of Pavia images, respectively. In the experiment on the Salinas image, only 1% of the labeled reference data were randomly selected as the training samples and the remaining 99% of data as the test set (see the second column in Table II). In the experiment on the University of Pavia image, as in some recent papers [49]–[51], a fixed number (200) of training samples for each class were randomly selected as the training samples and the rest as the test samples (see the second column in Table III). The selected training samples account for about 4% of the whole labeled reference data, which provide a challenging test set. The visual classification maps and quantitative results (averaged over ten experiments) obtained by various classifiers on the Salinas and University of Pavia images are shown in Figs. 6 and 7 and Tables II and III. As can be observed, the proposed superpixel-based SC-MK and IntraSC-MK classification methods deliver better performances than the other compared classifiers, in terms of visual quality and objective

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

TABLE II
CLASSIFICATION ACCURACY OF SALINAS IMAGE OBTAINED BY THE SVM [7], EMP [24], SVM-CK [27], LORSAL-MLL [44], SRC [12], MLR-GCK [25], INTRASC-MK, AND SC-MK METHODS. CLASS-SPECIFIC ACCURACY VALUES ARE IN PERCENTAGE

Class	Training/Test	SVM	EMP	SVM-CK	LORSAL-MLL	SRC	MLR-GCK	IntraSC-MK	SC-MK
Weeds_1	20/1989	99.74	99.84	99.09	99.44	**100**	98.75	**100**	**100**
Weeds_2	37/3689	99.01	99.76	99.37	99.95	99.98	99.35	**100**	**100**
Fallow	20/1956	91.05	93.15	98.69	99.78	97.61	97.54	99.92	**100**
Fallow plow	14/1380	97.04	98.49	**99.00**	98.34	83.24	98.84	98.44	98.62
Fallow smooth	27/2651	98.07	99.16	98.04	98.78	97.10	97.92	**98.80**	98.74
Stubble	40/3919	**99.98**	**99.98**	99.81	99.83	97.63	99.49	99.76	99.74
Celery	36/3543	98.89	**99.92**	99.34	99.66	99.57	99.51	**99.92**	**99.92**
Grapes	113/11158	75.96	92.96	89.86	90.76	88.61	92.21	99.32	**99.81**
Soil	62/6141	98.87	99.25	99.23	**99.97**	99.97	99.94	99.89	99.95
Corn	33/3245	88.86	93.37	95.00	94.15	96.11	96.46	96.49	**97.65**
Lettuce 4wk	11/1057	91.77	**98.80**	95.19	95.34	97.37	93.55	94.78	95.77
Lettuce 5wk	19/1908	95.75	96.53	99.84	99.99	95.52	99.88	98.59	**100**
Lettuce 6wk	9/907	94.78	98.01	**99.12**	97.83	95.08	98.38	98.11	98.15
Lettuce 7wk	11/1059	96.47	**97.30**	94.89	95.95	94.64	93.90	91.87	91.31
Vinyard untrained	73/7195	72.35	91.74	84.94	73.55	84.07	91.30	97.19	**99.78**
Vinyard trellis	18/1789	98.64	98.30	94.97	98.92	99.33	95.04	**100**	**100**
OA (Mean in %)		89.33	96.23	94.78	93.75	93.96	96.16	98.79	**99.38**
AA (Mean in %)		93.58	97.29	96.65	96.39	95.36	97.01	98.32	**98.72**
Kappa		0.88	0.95	0.94	0.93	0.93	0.96	**0.99**	**0.99**

TABLE III
CLASSIFICATION ACCURACY OF UNIVERSITY OF PAVIA IMAGE OBTAINED BY THE SVM [7], EMP [24], SVM-CK [27], LORSAL-MLL [44], SRC [12], MLR-GCK [25], INTRASC-MK, AND SC-MK METHODS. CLASS-SPECIFIC ACCURACY VALUES ARE IN PERCENTAGE

Class	Training/Test	SVM	EMP	SVM-CK	LORSAL-MLL	SRC	MLR-GCK	IntraSC-MK	SC-MK
Asphalt	200/6431	96.82	98.69	94.61	92.02	81.49	98.50	98.37	**98.84**
Meadows	200/18449	97.50	98.96	97.25	97.72	97.16	97.53	98.83	**99.14**
Gravel	200/1899	77.18	94.72	90.51	87.60	99.21	99.05	97.45	**99.70**
Trees	200/2864	87.90	96.05	98.08	97.31	88.30	**98.62**	98.12	98.19
Metal sheets	200/1145	97.38	98.47	99.98	99.68	97.29	99.43	**100**	**100**
Bare soil	200/4829	77.75	84.52	97.64	95.84	99.33	97.91	99.54	**99.59**
Bitumen	200/1130	64.57	87.97	97.58	96.65	98.93	98.51	**99.96**	**99.96**
Bricks	200/3482	85.91	98.29	93.23	91.48	97.32	98.98	99.56	**99.66**
Shadows	200/747	99.91	99.90	99.89	99.96	83.53	99.71	**100**	99.98
OA (Mean in %)		90.57	96.16	96.43	95.64	94.25	98.11	98.88	**99.22**
AA (Mean in %)		87.22	95.29	96.53	95.36	93.62	98.69	99.09	**99.46**
Kappa		0.88	0.95	0.95	0.94	0.92	0.97	0.98	**0.99**

TABLE IV
RUN TIME (IN SECONDS) FOR EACH STEP OF THE PROPOSED SC-MK METHOD ON THE INDIAN PINES, SALINAS, AND UNIVERSITY OF PAVIA IMAGES

	Indian Pines	Salinas	University of Pavia
Superpixel Creation	0.09	0.85	1.35
Mean Feature	0.08	0.41	0.44
Weighted Average Feature	0.15	1.89	2.02
SVM Training including kernels computation	149.05	36.86	128.65
SVM Test including kernels computation	9.83	9.91	42.31
Total	159.34	49.92	174.77

metrics. In addition, compared with the IntraSC-MK method that only uses the information within each superpixel, the SC-MK method that considers the correlations within and among superpixels, can further eliminate the disturbances and improve the estimations [see the ellipse regions in Figs. 6(i, j) and 7(i, j)].

In the aforementioned experiments, all the programs are operated on a laptop computer with an Intel Core i7-3720 CPU 2.60 GHz and 16 GB of RAM. Table IV reports the computational time of each step for the proposed SC-MK method on the Indian Pines, Salinas, and University of Pavia images, respectively. As can be observed, the main computational cost is occupied by the SVM training process. The step for creating the superpixels does not consume much computational cost (0.09, 0.85, and 1.35 s for the Indian Pines, Salinas, and University of Pavia images, respectively). This is because the PCA greatly reduces the dimension of the HSI and the adopted oversegmentation algorithm in [30] is very efficient. In addition, the main processes for exploiting the information within and among superpixels (e.g., mean feature within each superpixel, weighted average feature among superpixels, and multiple-kernel computations) also do not create much computational complexity. Note that, since the SVM training process consumes too much

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

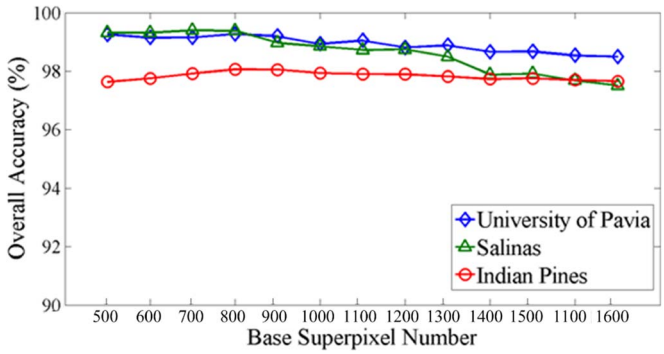


Fig. 8. Effect of the base superpixel number when applying the proposed SC-MK algorithm to the three test images.

computational cost, one of our ongoing works is to adopt the general-purpose graphics processing unit (GPU) to greatly accelerate this process.

C. Effect of the Number of Superpixels and Kernel Weights

Here, the effect of the base superpixel number and kernel weights on the performance of the proposed SC-MK method will be analyzed. In this analysis, the numbers of training and test samples are selected to the same as in the aforementioned experiments on the Indian Pines, Salinas, and University of Pavia images. Note that the reported accuracy values for this analysis experiment are also the average results over ten runs with randomly selected samples.

The base superpixel number was selected from 500 to 1600. Fig. 8 illustrates the OA values of the proposed SC-MK method under different base superpixel numbers on the three test images. We can observe that, as the base superpixel number varies from 500 to 1600, the OA values of the proposed SC-MK method generally show very good performances (OA values are over 97.5%) on the three test images. When the base superpixel number increases from 800 to 1600, the OA values of the proposed SC-MK method will slightly decrease on the three images. This is mainly due to the reason that if the superpixel number increases, the size of each superpixel will become small, and hence, the spatial information (e.g., in large homogeneous regions of the Salinas image) will not be sufficiently exploited for classification.

To examine the influences of the kernel weights to the performance of the proposed SC-MK method, the spectral kernel weight is first varied from 0 to 1, while the intra- and intersuperpixel kernel weights are selected to the corresponding equal values, as shown in Fig. 9(a). If the spectral kernel weight is set to 1 or 0 (which means that only spectral information or spatial information of the superpixel is utilized), the proposed SC-MK method does not show very good performances on the three test images. This indicates that both the spectral information and the spatial information of the superpixels should be utilized for the HSI classification. In addition, when the spectral kernel weight goes from 0.2 to 0.9, the performances of the proposed SC-MK method generally degrade on the three test images. This demonstrates that comparatively large weight values should be assigned to the two spatial kernels of the superpixels. Therefore, in Fig. 9(b), the spectral kernel weight is fixed to 0.2, while the intra- and intersuperpixel kernel weights are selected to the

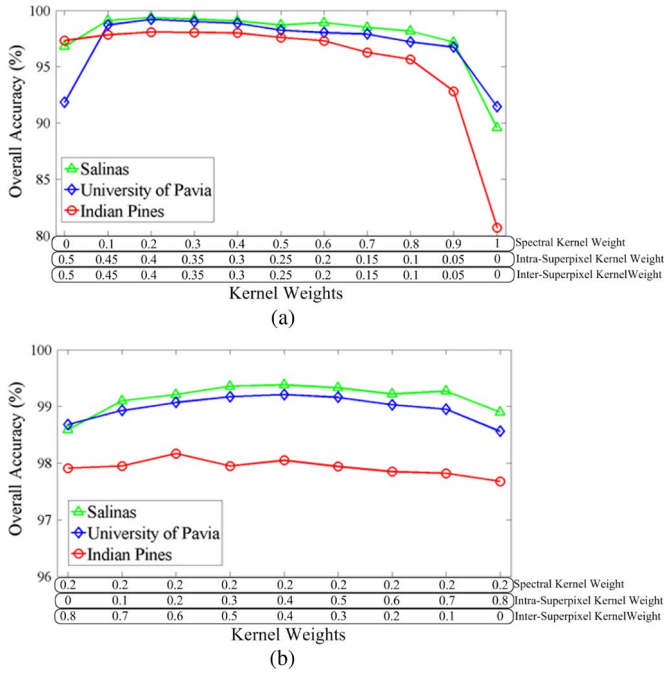


Fig. 9. Effect of the kernel weights when applying the proposed SC-MK algorithm to the three test images. (a) Effect of the spectral kernel weight. (b) Effect of the intra- and intersuperpixel kernel weights.

corresponding different values. As can be observed, if the intra-superpixel kernel weight is set from 0.2 to 0.6, the performances of the proposed SC-MK method are excellent and kept comparatively stable on the three test images. This shows that the spectral information and spatial information within and among superpixels should be all considered for the classification.

D. Effect of Different Number of Training Samples

Here, the effect of the number of training samples on several classifiers will be examined on the Indian Pines, Salinas, and University of Pavia images. The parameters for all the classifiers are kept the same as that in Section IV-B. For the Indian Pines and Salinas images, different percentages (from 2.5% to 30% for Indian Pines and from 0.25% to 3% for Salinas) of the labeled data were randomly selected as the training samples. For the University of Pavia image, various numbers (from 50 to 600 pixels for each class) of pixels were randomly chosen as the training set.

Fig. 10 illustrates the overall classification accuracy (averaged over ten runs) for each classifier under different numbers of training samples. We can observe that the performances for all the classifiers generally improve as the numbers of training samples increase. Furthermore, the proposed SC-MK method consistently provides superior performances over the other compared methods for all the test numbers of training samples.

V. Conclusion

In this paper, we present an SC-MK method for HSI classification. Instead of using a fixed-size region as some previous works, the SC-MK adopts the superpixel, whose size and shape can be adaptively adjusted according to the spatial structures of the HSI. Then, the SC-MK uses the multiple kernels to

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

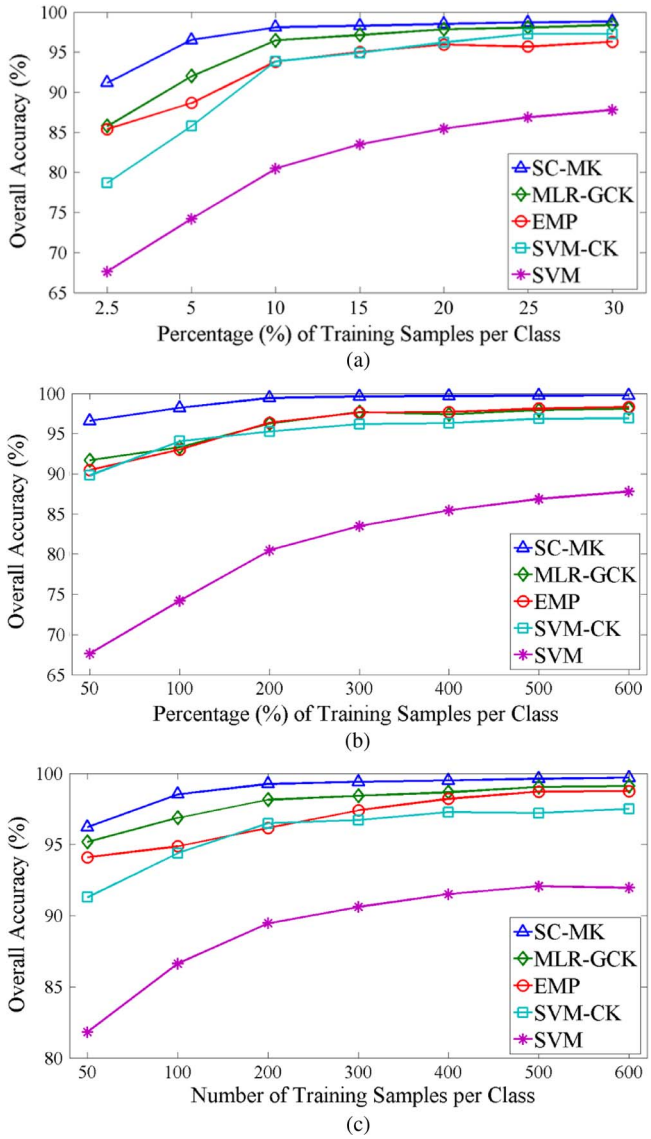


Fig. 10. Effect of the number of training samples on SC-MK, MLR-GCK, EMP, SVM-CK, and SVM for the (a) Indian Pines Image, (b) Salinas Image, and (c) University of Pavia Image.

effectively exploit the spectral–spatial information within and among superpixels. The experimental results on three real HSI images demonstrate the superiority of the proposed SC-MK over several well-known classifiers, in terms of both visual quality on the classification map and quantitative metrics.

In the experiments, the kernel weights were empirically selected to fixed values and not optimized for each test image. Therefore, we will systematically research on how to adaptively select the optimal kernel weights (e.g., based on the distributions of the materials in local regions of the test image) for different images. In addition, our future work will apply the superpixel-based kernel model to other hyperspectral applications (e.g., denoising, unmixing, and object recognition).

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Li and Dr. M.-Y. Liu for providing the software for the LORSAL-MLL and over-segmentation methods, respectively, on their websites.[2,3] They would also like to thank the editors and all of the anonymous reviewers for their constructive comments.

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] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

[3] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.

[4] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[5] N. Younan, S. Aksoy, and R. King, "Foreword to the special issue on pattern recognition in remote sensing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1331–1334, Oct. 2012.

[6] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer-Verlag, 1995.

[7] 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.

[8] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992.

[9] 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.

[10] 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.

[11] B. Du and L. Zhang, "Target detection based on a dynamic subspace," *Pattern Recog.*, vol. 47, no. 1, pp. 344–358, Jan. 2014.

[12] 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.

[13] U. Srinivas, Y. Chen, V. Monga, N. M. Nasrabadi, and T. D. Tran, "Exploiting sparsity in hyperspectral image classification via graphical models," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 505–509, May 2013.

[14] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[15] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[16] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.

[17] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.

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

[19] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4172–4186, Dec. 2007.

[20] H. Li, Z. Ye, and G. Xiao, "Hyperspectral image classification using spectral-spatial composite kernels discriminant analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published.

[21] Q. Shi, L. Zhang, and B. Du, "Semisupervised discriminative locally enhanced alignment for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4800–4815, Sep. 2013.

[2]http://www.lx.it.pt/~jun/

[3]https://sites.google.com/site/seanmingyuliu/

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

[22] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. New York, NY, USA: Wiley, 2009.
[23] J. Li, I. Dópido, P. Gamba, and A. Plaza, "Complementarity of discriminative classifiers and spectral unmixing techniques for the interpretation of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2899–2912 May 2015.
[24] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
[25] 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.
[26] 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.
[27] 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.
[28] G. Mori, X. Ren, A. A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *Proc. IEEE Conf. Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2004, pp. II-326–II-333.
[29] R. Achanta *et al.*, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[30] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 2097–2104.
[31] G. Zhang, X. Jia, and N. M. Kwok, "Super pixel based remote sensing image classification with histogram descriptors on spectral and spatial data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 4335–4338.
[32] J. Liu, W. Yang, S. Tan, and Z. Wang, "Remote sensing image classification based on random projection super-pixel segmentation," in *Proc. SPIE*, 2013, Art ID. 89210T.
[33] Z. Sun, C. Wang, H. Wang, and J. Li, "Learn multiple-kernel SVMs for domain adaptation in hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1224–1228, Sep. 2013.
[34] Y. Gu *et al.*, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.
[35] V. N. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998, vol. 2.
[36] I. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Wiley, 2005.
[37] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2009.
[38] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Math. Program.*, vol. 14, no. 1, pp. 265–294, Jan. 1978.
[39] 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.
[40] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[41] L. Fang *et al.*, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 3, no. 5, pp. 927–942, May 2012.
[42] L. Fang *et al.*, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Trans. Med. Imag.*, vol. 32, no. 11, pp. 2034–2049, Nov. 2013.
[43] L. Fang, S. Li, X. Kang, J. A. Izatt, and S. Farsiu, "3-D adaptive sparsity based image compression with applications to optical coherence tomography," *IEEE Trans. Med. Imag.*, vol. 34, no. 6, pp. 1306–1320, Jun. 2015.
[44] 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.
[45] 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, Jul. 2011.
[46] M. Dalla Mura, J. Atli 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, Jul. 2010.
[47] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," in *Proc. SPIE*, 1999, vol. 3584, pp. 221–232.
[48] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.
[49] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014.
[50] B. Song *et al.*, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, Aug. 2014.
[51] R. Ji *et al.*, "Spectral-spatial constraint hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1811–1824, Mar. 2014.

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

Since June 2013, he has been an Assistant Professor with the College of Electrical and Information Engineering, Hunan University. From September 2011 to September 2012, he was a Visiting Ph.D. Student in the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. His research interests include sparse representation and multiresolution analysis in remote sensing and medical image processing.

Dr. Fang was a recipient of the Scholarship Award for Excellent Doctoral Student granted by the Chinese Ministry of Education in 2011.

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

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong. From November 2002 to November 2003, he was a Postdoctoral Fellow at the Royal Holloway College, University of London, working with Prof. J. Shawe-Taylor. During April 2005 to June 2005, he was a Visiting Professor at the Department of Computer Science, Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored more than 160 refereed papers. His professional interests include compressive sensing, sparse representation, image processing, and pattern recognition.

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

Wuhui Duan received the B.S. degree in electrical and information engineering from Hunan University, Changsha, China, in 2012. She is currently working toward the M.S. degree with the same university.

Her research interest includes spectral–spatial hyperspectral image classification.

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

Jinchang Ren he received the M.Eng. degree in image processing and pattern recognition and the B.Eng. degree in computer software from Northwestern Polytechnical University, Xi'an, China, in 1997 and 1992, respectively, and the Ph.D. degree in electronic imaging and media communication from the University of Bradford, Bradford, U.K., in 2009.

Since December 2010, he has been a (Senior) Lecturer with the Centre for excellence of Signal and Image Processing (CeSIP), Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K. Before he joined Strathclyde, he was a Research Staff in several universities in U.K., including the University of Bradford, the University of Surrey, Kingston University, and the University of Abertay, Dundee. He has published over 100 peer-reviewed research papers in prestigious international journals and conferences, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON SYSTEM MAN, AND CYBERNETICS, *Computer Vision and Image Understanding*, etc., as well as the British Machine Vision Conference, the International Conference on Image Processing, the International Conference on Multimedia and Expo, Visual Information Engineering, book series of Lecture Notes in Computer Science etc. His research interests include pattern recognition, human–computer interaction, visual surveillance, archive restoration, motion estimation, and hyperspectral imaging.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree from the University of Iceland, Reykjavík, 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, all in electrical engineering.

On July 1, 2015, he became the Rector of the University of Iceland. From 2009 to 2015, he was the Pro Rector of Science and Academic Affairs and a Professor of electrical and computer engineering at the University of Iceland. His research interests are in remote sensing, image analysis, pattern recognition, biomedical analysis of signals, and signal processing, and he has published extensively in those fields.

Prof. Benediktsson is a Fellow of SPIE. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica, and Tau Beta Pi. He was a member of the 2014 IEEE Fellow Committee. He was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS Administrative Committee since 2000. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008, and he has served as an Associate Editor for TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and IEEE ACCESS since 2013. He is on the Editorial Board of the PROCEEDINGS OF THE IEEE and the International Editorial Board of the *International Journal of Image and Data Fusion*, and he was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS) 2007–2010. He is a cofounder of the biomedical startup company Oxymap. He received the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award; in 2000, he was granted the IEEE Third Millennium Medal; in 2004, he was a corecipient of the University of Iceland's Technology Innovation Award; in 2006, he received the Yearly Research Award from the Engineering Research Institute of the University of Iceland; and in 2007, he received the Outstanding Service Award from the IEEE GRSS. He was a corecipient of the 2012 IEEE TGRS Paper Award, and in 2013, he was a corecipient of the IEEE GRSS Highest Impact Paper Award. In 2013, he received the IEEE/VFI Electrical Engineer of the Year Award. In 2014, he was a corecipient of the *International Journal of Image and Data Fusion* Best Paper Award.