

# Multiple Feature Kernel Sparse Representation Classifier for Hyperspectral Imagery

Le Gan<sup>ID</sup>, Junshi Xia<sup>ID</sup>, Member, IEEE, Peijun Du<sup>ID</sup>, Senior Member, IEEE,  
and Jocelyn Chanussot, Fellow, IEEE

**Abstract**—Multiple types of features, e.g., spectral, filtering, texture, and shape features, are helpful for hyperspectral image (HSI) classification tasks. Combining multiple features can describe the characteristics of pixels from different perspectives, and always results in better classification performance. Recently, multifeature combination learning has been widely employed to the multitask-learning-based representation-based model to obtain a multifeature representation vector. However, the linear sparse representation-based classifier (SRC) cannot handle the HSI with highly nonlinear distribution, and kernel sparse representation-based classifier (KSRC) can remedy the drawback of linear SRC. By adopting nonlinear mapping, the samples in kernel space are often of high or even infinite dimensionality. In this paper, we integrate kernel principal component analysis into multifeature-based KSRC and propose a novel multiple feature kernel sparse representation-based classifier (namely, MFKSRC) for hyperspectral imagery. More specifically, spatial features, Gabor textures, local binary patterns, and difference morphological profiles are adopted and then each kind of feature is transformed nonlinearly into a new low-dimensional kernel space. The proposed framework can handle data with nonlinear distribution and add a dimensionality reduction stage in kernel space before optimizing the corresponding cost function. Experimental results on different HSIs demonstrate that the proposed MFKSRC algorithm outperforms the state-of-the-art classifiers.

**Index Terms**—Hyperspectral image (HSI) classification, kernel principal component analysis (KPCA), multiple feature learning, multitask learning, sparse representation.

## I. INTRODUCTION

HYPERSPECTRAL images (HSIs), gathered by hyperspectral sensors in hundreds of contiguous spectral

Manuscript received June 13, 2017; revised January 26, 2018; accepted March 7, 2018. Date of publication April 4, 2018; date of current version August 27, 2018. This work was supported by the Natural Science Foundation of China under Grant 41471275 and Grant 41631176. (Corresponding author: Peijun Du.)

L. Gan and P. Du are with the Key Laboratory for Satellite Mapping Technology and Applications of National Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, Nanjing 210023, China, and also with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China (e-mail: ganleatlas@gmail.com; dupjrs@gmail.com).

J. Xia is with the Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 113-8654, Japan (e-mail: xiajunshi@gmail.com).

J. Chanussot is with the Univ. Grenoble Alpes, CNRS, Grenoble INP\*, GIPSA-lab, 38000 Grenoble, France, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr).

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

Digital Object Identifier 10.1109/TGRS.2018.2814781

bands, composed of high-dimensional image cube, including discriminative spectral signatures, and have provided numerous applications in materials classification [1]–[4], target identification, and anomaly detection. Among the numerous applications, classification of HSIs is one of the most important tasks from the machine-learning perspective. Numerous widely used classifiers including support vector machines (SVMs) [5], [6], multinomial logistic regression [7], [8], neural network [9], random forest [10], [11], and rotation forest [12], [13] have been developed over the past few years. The key idea of classification is to accurately categorize each pixel into certain classes based on meaningful and discrimination information learned from the original image.

To improve the classification performance, we often adapt different types of feature for spectral–spatial classification. Many attempts have been made to explore various types of feature descriptors (e.g., spectral, filtering, texture, and shape features) to depict a pixel of HSI [14]–[20]. In [14], extended morphological profiles (EMPs) have been extracted for constructing spectral–spatial features from HSI. In [15], 2-D Gabor filtering feature extracted from selected bands was investigated for HSI classification. In [16], local binary pattern (LBP) and local phase quantization feature were investigated for texture characterization of land cover classification of optical remote sensing image data. In [18], two fusion strategies are applied to the LBP-based histogram feature for HSI classification. Although those single-feature-based classifiers have good performance, one kind of feature can only describe the pixels from one perspective, and none of the common feature descriptors have the same discriminative power for all classes [21]. Combining multiple types of features is a recent trend for HSI classification. It is apparent that among many kinds of feature descriptors, each feature has various discriminative power for the given classes. It is natural to combine different types of descriptions to depict a pixel. Li *et al.* [22] proposed a new framework for multiple feature learning by integrating different types of (linear and nonlinear) features. Zhong and Wang [23] proposed a multiple conditional random field ensemble model, including gray level co-occurrence matrix, Gabor texture features, and gradient orientation feature. Zhang *et al.* [24] introduced a patch alignment framework to linearly combine multiple features (e.g., spectral, texture, and shape) and obtain a unified low-dimensional representation of these features for subsequent HSI classification.

Recently, the sparse representation-based classifier (SRC) [25], [26] has been widely used in various HSI

applications [27]–[35]. In SRC, a given test sample is represented as a sparse linear combination of all training samples (namely, dictionary atoms), and the recovered nonzero sparse coefficients are supposed to concentrate on the dictionary atoms with the sample class label as the test sample. To exploit the spectral and contextual information of HSI, Chen *et al.* [28] proposed a joint sparse representation classifier based on the joint sparsity model (JSM) with neighboring pixel information. Zhang *et al.* [32] extended the JSM by assigning different weights for the neighboring pixels around the central pixel via a nonlocal spatial prior. Fang *et al.* [36] proposed a multiscale adaptive sparse representation (MASR) method by integrating multiscale spatial information with the JSM model via an adaptive sparse strategy. Li *et al.* [29] introduced the collaborative representation strategy by adopting  $\ell_2$ -norm regularization instead of  $\ell_1$ -norm for HSI classification. Aiming at improving the performance of the SR-based classifier, we apply multiple types of features to depict a pixel from different perspectives to enhance the discriminative ability. Motivated by multitask learning theory [37], [38], single-feature-based SR models (each type of feature is one task) can be extended to multiple feature learning for HSI classification tasks in [39]–[41]. These multifeature-based classifiers commonly improve the robustness of HSI classification.

It is well known that many types of feature descriptors of an HSI often tend to be linearly inseparable. Despite their robust performance, the linear nature of representation-based classifier (SRC and CRC) is almost always inadequate for representing samples distributed linearly in one direction (with nonlinear distribution). Based on the standard SR model, some kernel SR-based classifiers [kernel sparse representation-based classifier (KSRC)] have been developed as robust classifiers that are more adaptable to cope with samples with the distribution in [42]–[45]. In kernel SR-based methods, original samples in input space are mapped into high-dimensional feature space to capture the nonlinear structure of these samples via kernel tricks. Thus, the same distribution in samples are easily grouped together and are linearly separable. Samples in kernel space are often of high or even infinite dimensionality, and  $\ell_1$ -minimization in this space is impractical. One of the commonly used strategies is adding a kernel-based dimensionality reduction (DR) step in the kernel space before implementing SR-based classifiers. Toward that end, kernel principal component analysis (namely, KPCA, as a nonlinear version of PCA) [46] is capable of capturing the higher order statistics by mapping nonlinearly the input space to a feature space where DR is implemented simultaneously.

Motivated by the above concerns, in this paper, we combine KPCA and linear SRC into a novel KSRC based on multiple types of feature descriptors of HSI. The new KSRC with multiple types of feature descriptors is integrated into a multitask learning framework, and we propose a multiple feature kernel sparse representation-based classifiers (MFKSRCS) for HSI classification. To obtain a sparse solution, the number of feature dimension should be smaller than the number of dictionary atoms. Wright *et al.* [25] considered that selecting

feature dimension is not that critical, as long as enough information is contained to recover the representation coefficients. Moreover, a small number of features can reduce the storage requirements and computational complexity. To cope with nonlinear HSI, we apply the kernel-based-DR method into the proposed multifeature-kernel SR-based framework to obtain a small number of representation features. The flowchart of the proposed MFKSRC method is shown in Fig. 1.

- 1) To obtain the more discriminative features of HSI, multiple types of feature descriptors are applied to HSI for extracting spectral–spatial information from original spectral data.
- 2) KPCA is applied to original linear SRC for HSI classification and to generate new kernel SR-based methods, which can efficiently handle nonlinear descriptors in their reduced feature space.
- 3) Based on multitask learning theory, the single-feature-based KSRC is extended to multiple feature kernel learning framework for HSI classification.

The remainder of this paper is organized as follows. Section II briefly introduces three feature descriptors and two related kernel techniques (KSRC and KPCA). Section III proposes the MFKSRC algorithm for HSI classification. The effectiveness of the proposed algorithms is given in Section IV. Finally, Section V concludes this paper.

## II. MULTIPLE FEATURE LEARNING AND RELATED KERNEL METHODS

In this section, we review some related works on multiple feature descriptor for HSI and kernel tricks.

### A. Multiple Feature Descriptor

Given a pixel  $y$ , three types of feature descriptors, i.e., Gabor feature  $y_{\text{gabor}}$ , LBP feature  $y_{\text{lbp}}$ , and difference morphological profile (DMP) feature  $y_{\text{DMP}}$ , are introduced in this section.

1) *Gabor Filtering Feature*: Gabor wavelets were introduced to image analysis due to its biological relevance and computational properties. The commonly used 2-D Gabor function can be defined as follows:

$$\psi_{\mu,v}(z) = \frac{\|k_{\mu,v}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,v}\|^2 \|z\|^2}{2\sigma^2}} \left[ e^{ik_{\mu,v} z} - e^{-\frac{\sigma^2}{2}} \right] \quad (1)$$

where  $z = (a, b)$  denotes the spatial domain variable, and  $k_{\mu,v} = \pi/2f^v \cdot e^{i(\pi\mu/8)}$  denotes the frequency vector, in which  $v$  and  $\mu$  define the scale and orientation of the Gabor function. The number of oscillations under the Gaussian envelope is determined by  $\sigma = 2\pi$ . The Gabor wavelet representation of the  $p$ th principal component (PC) image  $I_p(a, b)$  for HSI is the magnitude part of the convolution of the image with a family of Gabor kernels defined by (1) in specific orientation and scale

$$F_{\mu,v}^p(x, y) = \psi_{\mu,v}(a, b) * I_p(a, b) \quad (2)$$

and the Gabor filtering feature of a pixel  $y$  located at  $(a, b)$  associated with the  $p$ th PC image is given by

$$y_{\text{gabor}}^p = [F_{1,1}^p(a, b), \dots, F_{\mu,v}^p(a, b)] \in \mathcal{R}^{\mu v}. \quad (3)$$

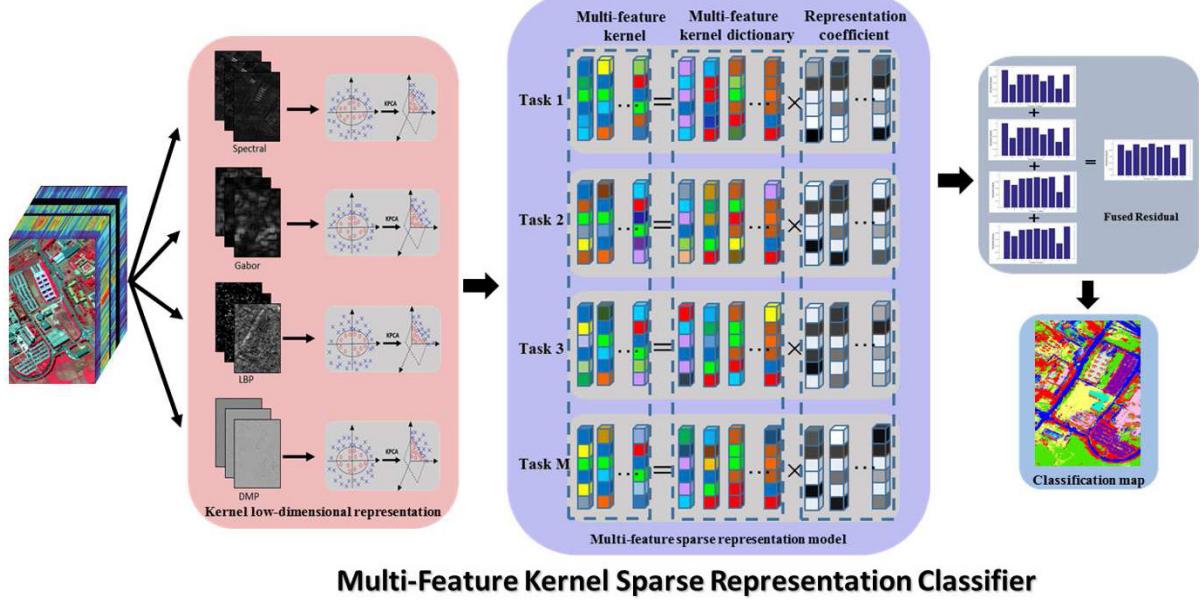


Fig. 1. Schematic illustration of the proposed MFKSRC algorithm with multiple feature descriptors in their reduced kernel subspace for HSI.

Then, the Gabor feature of a pixel  $y$  is obtained by stacked all the  $L$  PC-features  $\mathbf{y}_{\text{gabor}}^p$ , i.e.,  $\mathbf{y}_{\text{gabor}} = \{\mathbf{y}_{\text{gabor}}^p\}_{p=1,\dots,L} \in \mathcal{R}^{uvL}$ .

2) *LBP Texture Feature*: For each PC image  $I_p(a, b)$ , the LBP codes of a center pixel  $y$  in a local neighborhood can be obtained by comparing its value with its neighboring pixels in a single scan, which can be represented by

$$\text{LBP}_{u,r} = \sum_{u=0}^{u-1} s(g_t - g_c)2^t$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (4)$$

where  $g_c$  and  $g_t$  denote the gray value of the center pixel and its neighboring pixels (in a local circle),  $u$  denotes the number of neighboring pixels, and  $r$  denotes the radius of the neighborhood. The output of LBP operator in (4) is an  $u$ -bit binary string. After obtaining the LBP code image, the LBP histogram feature is computed over a local patch  $s_o \times s_o$ . As mentioned in [47], the dimensionality (i.e., number of bins) of the LBP features is  $u(u-1) + 3$  for each PC image. For HSI, the LBP feature of a pixel  $y$  of  $L$  PC image is defined as  $\mathbf{y}_{\text{lbp}} \in \mathcal{R}^{(u(u-1)+3)L}$ .

3) *DMP Shape Feature*: For shape feature extraction, Morphological profiles (MPs) [48], performing a series of morphological openings and closings with a family of structuring elements (SEs) of increasing size, are widely adopted to extract shape information for HSI. Let  $\gamma^{\text{SE}}(I_p)$  and  $\phi^{\text{SE}}(I_p)$  define the morphological opening and closing by reconstruction with the SEs for each PC image  $I_p(a, b)$ , and  $s_{\text{MP}}^p$  is defined using a series of SEs with increasing sizes on  $I_p$

$$s_{\text{MP}\gamma}^p = \{s_{\text{MP}\gamma}^{\tau}(I_p) = \gamma^{\tau}(I_p), \forall \tau \in [0, n]\}$$

$$s_{\text{MP}\phi}^p = \{s_{\text{MP}\phi}^{\tau}(I_p) = \phi^{\tau}(I_p), \forall \tau \in [0, n]\}$$

with

$$\gamma^0(I_p) = \phi^0(I_p) = I_p \quad (5)$$

where  $\tau$  denotes the radius of the disk-shaped-based SEs. Then,  $s_{\text{DMP}}^p$  are computed as the slopes of  $s_{\text{MP}}^p$  with an increasing SE series as follows:

$$s_{\text{DMP}\gamma}^p = \{s_{\text{DMP}\gamma}^{\tau}(I) = |s_{\text{MP}\gamma}^{\tau}(I_p) - s_{\text{MP}\gamma}^{\tau-1}(I_p)|\}$$

$$s_{\text{DMP}\phi}^p = \{s_{\text{DMP}\phi}^{\tau}(I) = |s_{\text{MP}\phi}^{\tau}(I_p) - s_{\text{MP}\phi}^{\tau-1}(I_p)|\}. \quad (6)$$

Then,  $s_{\text{DMP}\gamma}^p$  and  $s_{\text{DMP}\phi}^p$  are concatenated into a  $s_{\text{DMP}}^p$  vector for each PC image. For HSI, the DMP feature of a pixel  $y$  of  $L$  PC image is defined as  $\mathbf{y}_{\text{DMP}} \in \mathcal{R}^{2nL}$ .

### B. Kernel Tricks

For HSI, given original input space  $\mathcal{R}^h$  and the data sets  $\mathcal{D} = \{d_1, d_2, \dots, d_N\}$ ,  $N$  samples included, where  $d_i \in \mathcal{R}^h$  denotes a vector with  $h$ -dimensional feature,  $i = 1, 2, \dots, N$ . To make the samples separable, we give the nonlinear mapping  $\Phi$  from original feature space to high-dimensional (even infinite dimensional) kernel space as follows:

$$\Phi : \mathcal{R}^h \mapsto \mathcal{F} \quad \mathcal{D} \mapsto \Phi(\mathcal{D}) = [\phi(d)_1, \dots, \phi(d)_H]^T \quad (7)$$

where  $\mathcal{F}$  denotes the mapped feature space,  $\Phi$  denotes the corresponding kernel function, and  $H \gg h$  denotes the dimension of feature space  $\mathcal{F}$ . The main idea of kernel tricks is to employ a kernel function defined in the input space and indirectly operate the dot product in the higher dimensional feature space. The kernel trick can be formulated as

$$k(d_i, d_j) = \langle \phi(d_i), \phi(d_j) \rangle \quad (8)$$

where operator  $\langle \cdot \rangle$  denotes input operator and  $k(\cdot)$  denotes the defined kernel function. Generally, the kernels should satisfy the Mercer's condition [49] such as continuous, symmetric, and positive semidefinite properties.

### III. PROPOSED MULTIPLE FEATURE KERNEL SPARSE REPRESENTATION CLASSIFIER

#### A. Multiple-Feature-Based Sparse Representation Classifier

The sparse coding obtained by single-feature-based dictionary only can reflect the sparse pattern from one perspective. To fully exploit the discriminative coding patterns, multiple coding coefficients of different representation patterns are learned from multiple-feature-based dictionary set, which constructed by different kinds of feature descriptors. In the multiple-feature-based SR model, each pixel is described by  $M$  different modalities of discriminative features. Suppose that we are given  $C$  distinct classes and a set of  $N_c$  dictionary atoms per class. For each feature modality denoted by  $m$  with  $h_m$  dimensions, let  $\mathcal{D}_c^m \equiv \{d_{1,c}^m, \dots, d_{N_c,c}^m\} \in \mathcal{R}^{h_m \times N_c}$  ( $m = 1, \dots, M; c = 1, \dots, C$ ) denotes the subdictionary of the  $m$ th feature corresponding to the  $c$ th class and  $N = \sum_c N_c$ . Construct a new dictionary  $\mathcal{D}^m$  associated with the  $m$ th feature as the concatenation of subdictionary  $\mathcal{D}_c^m$  from all the classes as

$$\begin{aligned}\mathcal{D}^m &= \{\mathcal{D}_1^m, \dots, \mathcal{D}_C^m\} \\ &= \{d_{1,1}^m, \dots, d_{N_1,1}^m | \dots | d_{1,C}^m, \dots, d_{N_C,C}^m\}. \quad (9)\end{aligned}$$

For each test pixel  $\mathbf{y} = \{\mathbf{y}^m\}_{m=1,\dots,M}$  (with  $M$  features) of unknown class, we denote  $\mathbf{y}^m$  as the  $m$ th modality of the feature descriptor. Based on multitask learning [38] theory, the multifeature-based sparse representation model can be formulated as

$$\begin{aligned}\hat{\mathbf{A}} &= \arg \min_{\boldsymbol{\alpha}^m} \frac{1}{2} \sum_{m=1}^M \left\| \mathbf{y}^m - \sum_{c=1}^C \mathcal{D}_c^m \boldsymbol{\alpha}_c^m \right\|_2^2 + \lambda \sum_{c=1}^C \|\boldsymbol{\alpha}_c^m\|_1 \\ &= \arg \min_{\boldsymbol{\alpha}^m} \frac{1}{2} \sum_{m=1}^M \|\mathbf{y}^m - \mathcal{D}^m \boldsymbol{\alpha}^m\|_2^2 + \lambda \|\boldsymbol{\alpha}^m\|_1 \quad (10)\end{aligned}$$

where  $\lambda$  is a positive parameter,  $\mathbf{A} = \{\boldsymbol{\alpha}^m\}_{m=1,\dots,M}$  and  $\boldsymbol{\alpha}^m = [\alpha_1^m, \dots, \alpha_N^m]^T$  is the sparse coefficient of the  $m$ th feature  $\mathbf{y}^m$  over the  $m$ th dictionary  $\mathcal{D}^m$ . Once  $\{\boldsymbol{\alpha}^m\}_{m=1,\dots,M}$  is obtained, the label of pixel  $\{\mathbf{y}^m\}_{m=1,\dots,M}$  described by multiple features is determined according to the minimum residual between  $\{\mathbf{y}^m\}_{m=1,\dots,M}$  and its approximations obtained over their corresponding subdictionaries  $\{\mathcal{D}^m\}_{m=1,\dots,M}$ , i.e.,

$$\text{class}(\mathbf{y}) = \arg \min_{c=1,\dots,C} \sum_{m=1}^M \|\mathbf{y}^m - \mathcal{D}^m \delta^m(\boldsymbol{\alpha}^m)\|_2^2 \quad (11)$$

where  $\delta^m$  is the characteristic function [25] for the  $m$ th feature that chooses coefficients associated with class  $c$  and makes the rest to zero.

#### B. Dimensionality Reduction in Kernel Space

As we all know, samples of HSI are not linearly separable. Linear SR model is inadequate to represent the nonlinear structures of samples. To capture the nonlinear characteristic of samples, nonlinear SR model was adopted by projecting the samples into a high-dimensional feature space using kernel tricks. By introducing nonlinear mapping  $\Phi$ , original linear

SRC is extended to nonlinear SRC (i.e., KSRC)

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \arg \min_{\hat{\boldsymbol{\alpha}}} \|\Phi(\mathbf{y}) - \Phi(\mathcal{D})\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \\ &= \arg \min_{\hat{\boldsymbol{\alpha}}} \|\mathbf{K}(\mathcal{D}, \mathbf{y}) - \mathbf{G}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (12)\end{aligned}$$

where  $\mathbf{K}(\mathcal{D}, \mathbf{y}) = [k(d_1, \mathbf{y}), \dots, k(d_N, \mathbf{y})]^T \in \mathcal{R}^{N \times 1}$  and  $\mathbf{G} = \Phi\Phi^T \in \mathcal{R}^{N \times N}$  denote the kernel Gram matrix with  $G_{i,j} = k(d_i, d_j)$ . However, samples in kernel space are often of very high or possibly infinite dimensionality, and the sparse representation with  $\ell_1$ -regularizer in the space is impractical. It is essential for a method to be able to capture the nonlinear characteristic of samples in a low-dimensional (reduced) subspace. Note that KPCA [50] applies kernel methods by nonlinearly mapping the data samples into a new kernel-induced space, where DR is ultimately implemented simultaneously. Hence, by integrating KPCA into linear SRC, a novel kernel SRC can be formulated as

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\hat{\boldsymbol{\alpha}}} \|\mathcal{P}^T \Phi(\mathbf{y}) - \mathcal{P}^T \Phi(\mathcal{D})\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (13)$$

where  $\mathcal{P} = [\mathcal{P}_1, \dots, \mathcal{P}_s] \in \mathcal{R}^{H \times s}$  ( $H \gg s$ ,  $s$  is the dimension of projected subspace) is the projection matrix associated with KPCA. As we know, it is difficult to select an appropriate kernel function for kernel-based methods. Some widely adopted kernel functions are based on Euclidean inner product and Euclidean distance, including linear kernel

$$k(d_i, d_j) = d_i^T d_j \quad (14)$$

polynomial kernel

$$k(d_i, d_j) = (a + d_i^T d_j)^b \quad (15)$$

and Gaussian kernel

$$k(d_i, d_j) = \exp\left(-\frac{\|d_i - d_j\|^2}{2\sigma^2}\right), \quad \sigma \in \mathbb{R}^+ \quad (16)$$

where  $a, b$ , and  $\sigma$  are the kernel parameters. Each of them transfers samples to a specific higher dimensional space which is determined by the form of the corresponding kernel function.

#### C. Multiple Feature Kernel Sparse Representation Classifier

Again, the sparse coding in different kinds of feature space can capture different discriminative characteristics, and the joint exploitation of sparse coding in multiple-feature-based space can lead to an improved classification performance for HSI. Meanwhile, KPCA is adopted to capture the intrinsic nonlinear characteristic of different kinds of features in the low-dimensional subspace. Inspired by the above ideas, we develop a classification framework for the integration of different sparse patterns in multiple-feature-based kernel subspace, and propose an MFKSRC for HSI classification. It can be formulated as

$$\hat{\mathbf{A}} = \arg \min_{\boldsymbol{\alpha}^m} \sum_{m=1}^M \|\mathcal{P}_m^T \Phi^m(\mathbf{y}^m) - \mathcal{P}_m^T \Phi^m(\mathcal{D}^m)\boldsymbol{\alpha}^m\|_2^2 + \lambda \|\boldsymbol{\alpha}^m\|_1 \quad (17)$$

where  $\Phi^m(\mathbf{y}^m)$  denotes the unlabeled pixel in the kernel space associated with the  $m$ th modality of feature,  $\Phi^m(\mathcal{D}^m)$  denotes

the training kernel dictionary associated with feature  $m$ , and  $\mathcal{P} = \{\mathcal{P}_m\}_{m=1,\dots,M}$  ( $\mathcal{P}_m \in \mathbb{R}^{H_m \times s_m}$ ,  $H_m \gg s_m$ ,  $s_m$  denotes the subspace dimension associated with feature  $m$ ) denote the transformation matrices via KPCA for each feature. By introducing the kernel Gram matrix  $\mathbf{K}^m = \Phi^m(\mathcal{D}^m)^T \Phi^m(\mathcal{D}^m) \in \mathbb{R}^{N \times N}$  and kernel vector  $\mathbf{k}^m(\cdot, \mathbf{y}^m) = [k^m(d_1^m, \mathbf{y}^m), \dots, k^m(d_N^m, \mathbf{y}^m)]$  (associated with feature  $m$ ), and  $\mathbf{K}_{ij}^m = k^m(d_i, d_j)$ , (17) can be rewritten as

$$\hat{\mathbf{A}} = \arg \min_{\boldsymbol{\alpha}^m} \sum_{m=1}^M \|\mathcal{P}_m^T \mathbf{k}^m(\cdot, \mathbf{y}^m) - \mathcal{P}_m^T \mathbf{K}^m \boldsymbol{\alpha}^m\|_2^2 + \lambda \|\boldsymbol{\alpha}^m\|_1. \quad (18)$$

The problem of (18) is known as the multitask joint covariate selection model in sparse learning, and can be efficiently solved by several  $\ell_1$ -minimization problem. Once obtaining the multifeature sparse coefficients  $\hat{\mathbf{A}} = \{\boldsymbol{\alpha}^m\}_{m=1,\dots,M}$ , we classify  $\{\mathbf{y}^m\}_{m=1,\dots,M}$  in term of  $\hat{\mathbf{A}}$ . We use the minimum residual between  $\mathbf{y}$  and its  $c$  approximations in the reduced subspace to determine the label of  $\{\mathbf{y}^m\}_{m=1,\dots,M}$

$$\text{class}(\mathbf{y}) = \arg \min_c \sum_{m=1}^M \|\mathcal{P}_m^T \mathbf{k}^m(\cdot, \mathbf{y}^m) - \mathcal{P}_m^T \mathbf{K}^m \delta_c^m(\boldsymbol{\alpha}^m)\|_2^2. \quad (19)$$

The classification procedure of MFKSRC is described in Algorithm 1.

---

**Algorithm 1** Proposed MFKSRC for HSI

---

- 1: **Input:** 1) A HSI containing training dictionary;  
2) Parameters:  $\lambda$ , subspace dimension  $s^m$
  - 2: **Initialization:**
    - 1) Extract multiple features for each pixel of HSI;
    - 2) Construct multi-feature dictionary  $\{\mathcal{D}^m\}_{m=1,\dots,M}$  of  $M$  features, where  $\mathcal{D}_c^m = \{d_{cj}\}_j^{N_c} \in \mathcal{R}^{h_m \times N_c}$  is subdictionary associated with class  $c$ ;
    - 3) Select suitable Mercer kernel  $k^m(\cdot, \cdot)$  and compute the kernel Gram matrix associated with the  $m$ -th feature  $\mathbf{K}^m$ ;
    - 4) Obtain the projected matrix  $\mathcal{P} = \{\mathcal{P}_m\}_{m=1,\dots,M}$  via KPCA, and normalize the columns of  $\mathcal{P}^m T \mathbf{K}^m$  and to have unit  $\ell_2$ -norm;
  - 3: **for** each unlabeled pixel  $\{\mathbf{y}^m\}_{m=1,\dots,M}$  in HSI **do**
  - 4:   1) Compute the kernel vector  $\mathbf{k}^m(\cdot, \mathbf{y})$ , and normalize each column of  $\mathcal{P}^m T \mathbf{k}^m(\cdot, \mathbf{y}^m)$  to have unit  $\ell_2$ -norm;
  - 5:   2) Code  $\{\mathbf{y}^m\}_{m=1,\dots,M}$  over multi-feature dictionary  $\{\mathcal{D}^m\}_{m=1,\dots,M}$ , obtain the multifeature sparse representation  $\hat{\mathbf{A}} = \{\boldsymbol{\alpha}^m\}_{m=1,\dots,M}$  by optimization Eq. (18);
  - 6:   3) Decide the final label of  $\{\mathbf{y}^m\}_{m=1,\dots,M}$  based on (19).
  - 7: **end for**
  - 8: **Output:** A 2-D map which records the labels of the HSI.
- 

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

To validate the efficacy of the proposed MFKSRC algorithm, we adopt three publicly available hyperspectral data sets<sup>1</sup>: the Airborne Visible/Infrared Imaging Spectrometer

(AVIRIS) (Indian Pines) image, the Reflective Optics System Imaging Spectrometer (ROSIS) University of Pavia image, and the AVIRIS Kennedy Space Center (KSC) image. For comparison purpose, we compare the proposed MFKSRC method with several widely used state-of-the-art classifiers, including the spatial SVM classifier by composite kernel (SVM-CK [54]), the nearest regularized subspace (NRS) classifier [55], the joint sparse representation-based classifier (namely, SOMP [28]), the MASR classifier [36], the extended multi-attribute profiles (EMAPs)-based multiple nonlinear feature learning classifier (MNFL [22]), and the pixel-wise multiple feature adaptive sparse representation classifier (MFASR [56]). When the selected feature number reduce to 1, MFKSRC converts to be single-feature-based method (namely, SFKSRC). Based on the aforementioned multiple features, the optimal features for the three HSIs are different. For SFKSRC, we adopt DMP feature for Indian Pines image and Gabor feature for the other two HSIs. In our experiments, the average individual class accuracy [%], overall accuracy (OA) [%], average accuracy (AA) [%], and kappa coefficient ( $\kappa$ ) [%]) for different classifiers are adopted by the average results over 10 independent Monte Carlo (MC) runs, and the corresponding standard deviation is considered to evaluate the statistical significance of the results.

For the multiple feature learning stage, original spectral feature and three feature descriptors (as described in Section II-A, Gabor filtering feature [52], [57], LBP texture feature [51], and DMP shape feature [53]) are adopted to describe each pixel of the three HSI sets in our experiments. DR is applied to obtain the first  $L$ -PC image for each feature descriptor. PCA is chosen as the DR method. The detailed parameter values used in this paper for the three feature descriptors are listed in Table I.

For MFKSRC, three different kernels (i.e., linear, polynomial, and gaussian kernel) are adopted. As for the parameter setting, we have two key parameters: regularizer parameter  $\lambda$  and kernel parameter  $\sigma$ . Regularizer parameter  $\lambda$  is learned by cross-validation (CV) in the range from  $1e-7$  to  $1e-1$ . For gaussian kernel,  $\sigma$  is set by the median of  $(1/\|d_i - \bar{d}\|^2)$ ,  $i = 1, \dots, N$ , where  $\bar{d}$  is the mean of all training dictionary. The stacked kernel is adopted for SVM-CK (with spatial mean feature), which is implemented with the help of the LIBSVM [58] package. For SOMP, the spatial scale is learned by CV in the range from  $3 \times 3$  to  $11 \times 11$  for three HSIs. For MASR, a multiscale combination of  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ , and  $9 \times 9$  is adopted for three HSIs. For MASAR, four features, (i.e., spectral feature, EMP feature [6], Gabor feature [15], and DMP [59]) are adopted for three HSIs, and the parameter values for different features can be found in [56]. The sparsity level for joint sparse representation methods (e.g., SOMP, MASR, and MFASR) is set to 3. All the experiments are performed using MATLAB R2016a on Intel Core i7-4790 CPU PC machine with 16 GB of RAM.

##### A. Hyperspectral Data Sets

Three HSIs are considered to validate the performance of the proposed MFKSRC in several challenging scenarios.

- 1) *Indian Pines Image:* The first HSI in our experiment was collected by the AVIRIS sensor over the Indian Pines

<sup>1</sup> Available online: <http://www.ehu.es/ccwintco/index.php>.

TABLE I  
PARAMETER SETTING FOR MULTIPLE FEATURE DESCRIPTORS

Feature	Indian pines	University of Pavia	KSC
LBP [51]	Base image: PC1, PC2, PC3 Neighborhood and radius: (8, 2) Local patch size: $21 \times 21$	No. of Base image: PC1, ..., PC7 Neighborhood and radius: (8, 2) Local patch size: $17 \times 17$	No. of Base image: PC1, ..., PC7 Neighborhood and radius: (8, 1) Local patch size: $21 \times 21$
Gabor [52]	Base image: PC1, ..., PC10 Scale: 1,2,3,4,5,6,7,8,9,10 Direction: $30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ$	Base image: PC1, ..., PC10 Scale: 1,2,3,4,5,6,7,8,9,10 Direction: $30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ$	Base image: PC1, ..., PC10 Scale: 1,2,3,4,5,6,7,8,9,10 Direction: $30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ$
DMP [53]	Base image: PC1, ..., PC10 Size of SEs: 1,2,3,4,5,6,7,8,9,10 Morphological operators: opening and closing	Base image: PC1, ..., PC10 Size of SEs: 1,2,3,4,5,6,7,8,9,10 Morphological operators: opening and closing	Base image: PC1, ..., PC10 Size of SEs: 1,2,3,4,5,6,7,8,9,10 Morphological operators: opening and closing

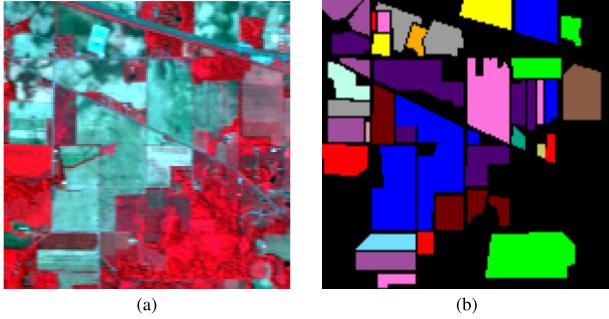


Fig. 2. Indian Pines data set. (a) RGB false color image (R:50, G:27, and B:17). (b) Ground truth map containing 16 land cover classes.

region in Northwestern Indiana, USA, on June 12, 1992. This original data set covers a mixed agricultural/forest area, and consists of  $145 \times 145$  pixels and 224 spectral bands range from 0.4 to  $2.5 \mu\text{m}$  with a spatial resolution of 20 m per pixel. In our experiments, the number of bands is reduced to 200 by removing four bands full of zero and 20 water absorption bands. The scene originally contains 16 different land cover classes (most of which are different kinds of crops) and constitutes a very challenging classification scenarios due to the presence of unbalanced labeled classes and mixed pixels. The false color image of three bands and the ground truth map of 16 land cover classes are visually shown in Fig. 2(a) and (b). Two kinds of sample selection methods, i.e., all 16 classes and nine classes with the maximum number (Corn-no till, Corn-min till, Grass/Pasture, Grass/Trees, Hay-windrowed, Soybean-no till, Soybean min till, Soybean-clean till, and Woods), are adopted for the data set.

- 2) *University of Pavia Image:* The second HSI was acquired by the ROSIS-03 (ROSIS-03) optical sensor, which covers an urban area surrounding the University of Pavia, Pavia, Italy, on July 8, 2002, with a pixel response in 115 spectral channels that range from 0.43 to  $0.86 \mu\text{m}$  covering the visible and infrared spectrum [1]. After removing the 12 noisy bands, the remaining 103 spectral bands are reserved for experiments. The spatial size of the image is  $610 \times 340$  pixels, and the spatial resolution is 1.3 m per pixel. The false color composite image and the ground truth map of nine land cover classes are shown in Fig. 3(a) and (b), respectively.
- 3) *KSC Image:* The third HSI is gathered by AVIRIS over the KSC, Brevard County, FL, USA, on March 23, 1996.

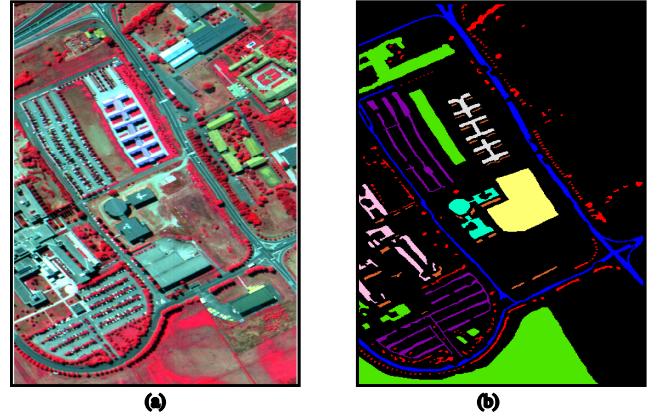


Fig. 3. University of Pavia data set. (a) RGB false color image (R:102, G:56, and B:31). (b) Ground truth map containing nine land cover classes.

The data set contains 224 bands whose wavelength covers the spectral range from 0.4 to  $2.5 \mu\text{m}$  with a spectral resolution of 10 nm, and the image size is  $512 \times 614$  pixels with a spatial resolution of 18 m. After removing water absorption and low signal-to-noise bands, a total of 176 bands remained for experiments. This scene contains 13 labeled classes of land cover.

### B. Experimental Results of AVIRIS Indian Pines Image

1) *Mechanism of the Proposed MFKSRC Classifier:* To exhibit the detail information of the proposed MFKSRC method, we investigate the distribution of sparse coefficients and the decision rule of the classifier in a visualization form. As a comparison, single-feature-based method (SFKSRC) with the aforementioned feature descriptors, i.e., spectral-feature-based method (SFKSRC-Spec), Gabor-filtering-feature-based method (SFKSRC-Gabor), LBP-texture-feature-based method (SFKSRC-LBP), and DMP-shape-feature-based method (SFKSRC-DMP), is also investigated in the experiment to demonstrate the necessity of combining multiple features. Fig. 4 shows the normalized sparse coefficients of these single- and multifeature methods and the corresponding normalized residuals for decision. In particular, nine classes are adopted in this experiment for the Indian Pines data set. Based on the features mentioned above, 50 atoms per class are randomly chosen to construct the single- and multifeature dictionary set. Then, a test pixel belonging to class 6 is represented as a sparse linear combination of all the single- and multifeature

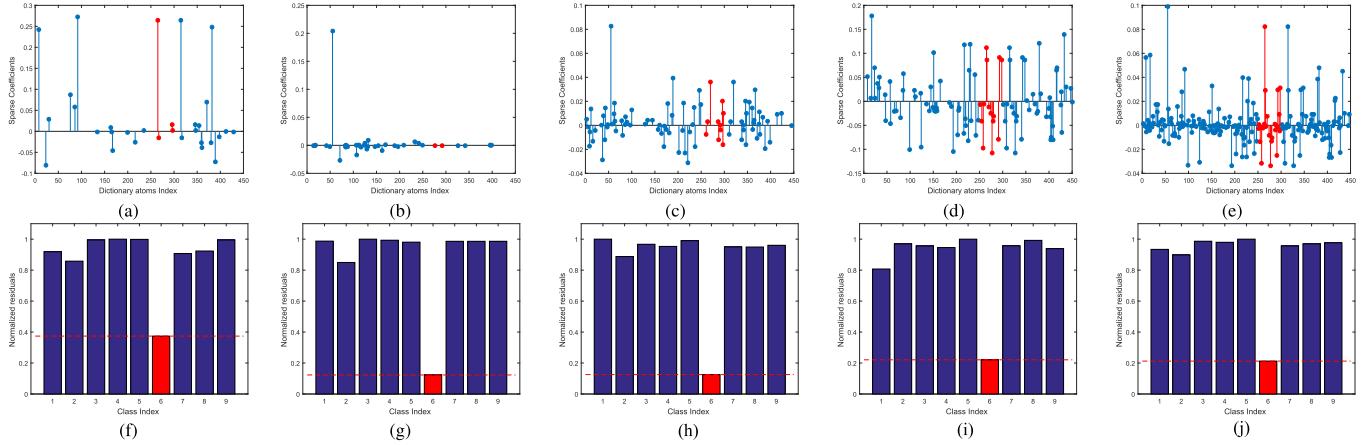


Fig. 4. Representation coefficients obtained by (a) SFKSRC-Spec, (b) SFKSRC-Gabor, (c) SFKSRC-LBP, (d) SFKSRC-DMP, and (e) MFKSRC. Normalized residuals obtained by (f) SFKSRC-Spec, (g) SFKSRC-Gabor, (h) SFKSRC-LBP, (i) SFKSRC-DMP, and (j) MFKSRC.

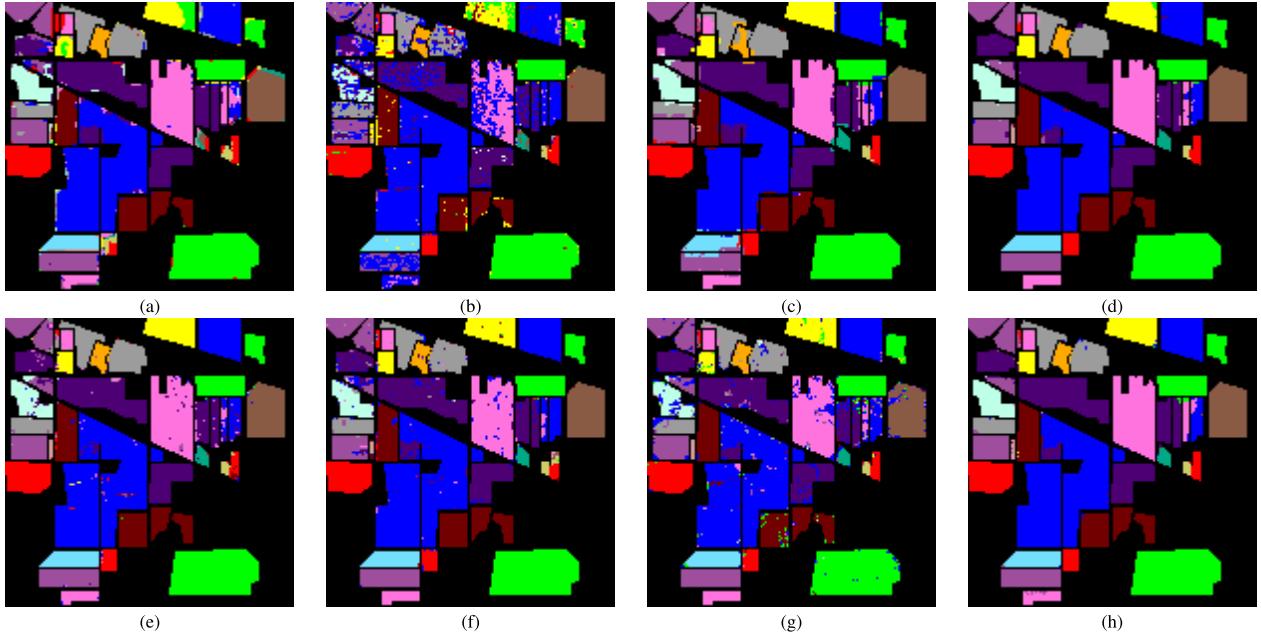


Fig. 5. Classification maps obtained by different classifiers for the AVIRIS Indian Pines image with 16 classes. (a) SVM-CK [54]. (b) NRS [55]. (c) SOMP [28]. (d) MASR [36]. (e) MNFL [22]. (f) MFASR [56]. (g) SFKSRC. (h) MFKSRC.

dictionary atoms. We plot recovered sparse coefficients when a test pixel is represented as a sparse linear combination of dictionary atoms constructed by different features. From Fig. 4, most of the recovered coefficients are clustered toward the atom index from 251 to 300 corresponding to class-specific subdictionary from class 6. Moreover, the labels of the unlabeled pixel for those methods are determined by the minimum residual rule in their feature space, respectively.

2) *Classification Results*: The averaged classification results of the proposed methods (SFKSRC and MFKSRC) and its competitors with the 16 classes for the Indian Pines image over 10 times MC runs are listed in Table II, and the classification maps are shown in Fig. 5. In the experiment, 10% of all labeled pixels per class are selected as a balanced dictionary set (or training samples); the remaining labeled pixels act as the test set. The regularizer parameters for SFKSRC and

MFKSRC are learned with a value of  $1e-4$  and  $1e-5$  for Indian pines image. In the proposed methods, the subspace dimension is set as 100 for the Indian pines image. To investigate the necessity of combining multiple features, we consider SFKSRC with the optimal feature in the experiment. As can be observed, the proposed MFKSRC method achieves an OA result of 98.33%, which is superior to the pixel-wise MFASRC and MNFL (with EMAPs) with a relative OA gain of 3.47% and 3.36%. By integrating different aspects of multiple modalities of the feature, MFKSRC outperforms than the single-feature-based method (SFKSRC). Compared to SOMP, the proposed single-feature method (SFKSRC) with optimal feature has a less OA result than that of the single-scale contextual information-based method. However, combining multiple features with SFKSRC, the multifeature method (MFKSRC) yields a better OA result than does the

TABLE II  
CLASSIFICATION ACCURACY (%) OBTAINED BY DIFFERENT CLASSIFIERS ON THE INDIAN PINES IMAGE  
USING 10% DICTIONARY ATOMS (OR TRAINING SAMPLES) FOR EACH CLASS

Class	#Sample		SVM-CK [54]	NRS [55]	SOMP [28]	MASR [36]	MNFL [22]	MFASR [56]	SFKSRC	MFKSRC
	Train	Test								
Alfalfa	6	48	17.65 $\pm$ 10.32	72.50 $\pm$ 4.97	92.08 $\pm$ 9.66	<b>95.42<math>\pm</math>2.55</b>	89.58 $\pm$ 4.17	88.75 $\pm$ 3.16	83.75 $\pm$ 5.19	92.71 $\pm$ 4.84
Corn-no till	144	1290	92.51 $\pm$ 1.08	68.82 $\pm$ 1.33	94.45 $\pm$ 1.02	97.44 $\pm$ 0.54	90.98 $\pm$ 2.34	92.57 $\pm$ 0.78	87.44 $\pm$ 2.12	<b>97.92<math>\pm</math>0.53</b>
Corn-min till	87	750	88.13 $\pm$ 1.75	52.93 $\pm$ 3.64	92.00 $\pm$ 1.90	95.95 $\pm$ 0.60	89.52 $\pm$ 3.73	90.99 $\pm$ 1.98	87.73 $\pm$ 3.04	<b>98.28<math>\pm</math>0.78</b>
Corn	24	210	95.05 $\pm$ 0.88	39.14 $\pm$ 7.88	95.33 $\pm$ 3.37	96.38 $\pm$ 0.62	89.71 $\pm$ 2.71	86.38 $\pm$ 5.30	69.71 $\pm$ 3.48	<b>99.19<math>\pm</math>0.98</b>
Grass/Pasture	50	447	83.90 $\pm$ 2.71	90.43 $\pm$ 2.88	92.68 $\pm$ 1.35	92.66 $\pm$ 0.58	92.17 $\pm$ 1.08	92.48 $\pm$ 1.35	94.18 $\pm$ 1.35	<b>96.64<math>\pm</math>1.56</b>
Grass/Trees	75	672	96.05 $\pm$ 1.27	95.00 $\pm$ 1.53	94.87 $\pm$ 1.42	97.38 $\pm$ 0.13	97.92 $\pm$ 0.85	98.04 $\pm$ 0.96	95.33 $\pm$ 0.65	<b>99.21<math>\pm</math>0.56</b>
Grass/Pasture-mowed	3	23	75.00 $\pm$ 4.26	49.57 $\pm$ 22.51	64.35 $\pm$ 28.31	95.65 $\pm$ 8.13	94.78 $\pm$ 1.94	<b>95.67<math>\pm</math>3.07</b>	83.48 $\pm$ 10.83	90.00 $\pm$ 10.46
Hay-windrowed	49	440	90.95 $\pm$ 1.07	98.64 $\pm$ 0.48	99.86 $\pm$ 0.24	<b>100<math>\pm</math>0.00</b>	99.27 $\pm$ 0.25	99.55 $\pm$ 0.28	86.14 $\pm$ 2.75	99.61 $\pm$ 0.30
Oats	2	18	31.58 $\pm$ 17.02	34.44 $\pm$ 7.24	26.11 $\pm$ 19.07	81.11 $\pm$ 14.91	83.33 $\pm$ 16.20	63.33 $\pm$ 20.64	<b>83.33<math>\pm</math>13.03</b>	71.67 $\pm$ 23.49
Soybeans-no till	97	871	91.95 $\pm$ 0.89	51.76 $\pm$ 2.54	92.90 $\pm$ 0.52	96.53 $\pm$ 1.19	93.04 $\pm$ 2.05	89.58 $\pm$ 1.32	83.42 $\pm$ 1.64	<b>96.61<math>\pm</math>2.28</b>
Soybeans-min till	247	2221	92.32 $\pm$ 2.40	94.65 $\pm$ 0.74	95.93 $\pm$ 0.85	98.34 $\pm$ 0.29	95.97 $\pm$ 0.67	97.25 $\pm$ 0.44	96.19 $\pm$ 0.72	<b>98.92<math>\pm</math>0.75</b>
Soybeans-clean till	62	552	91.25 $\pm$ 1.99	63.55 $\pm$ 3.76	91.72 $\pm$ 3.83	<b>96.41<math>\pm</math>0.83</b>	95.14 $\pm$ 2.37	91.12 $\pm$ 2.04	79.64 $\pm$ 3.05	96.36 $\pm$ 1.34
Wheat	22	190	85.57 $\pm$ 2.04	95.79 $\pm$ 2.07	88.84 $\pm$ 4.75	93.16 $\pm$ 0.24	<b>99.58<math>\pm</math>0.24</b>	98.84 $\pm$ 0.44	96.11 $\pm$ 0.47	98.58 $\pm$ 1.74
Woods	130	1164	98.13 $\pm$ 0.77	96.60 $\pm$ 0.84	99.05 $\pm$ 0.66	99.95 $\pm$ 0.11	<b>100<math>\pm</math>0.00</b>	99.76 $\pm$ 0.31	97.75 $\pm$ 0.44	99.87 $\pm$ 0.16
Bldg-grass-tree-drives	38	342	68.14 $\pm$ 1.72	61.99 $\pm$ 5.41	94.59 $\pm$ 2.47	98.42 $\pm$ 2.74	93.98 $\pm$ 2.21	95.56 $\pm$ 0.63	91.11 $\pm$ 3.14	<b>99.33<math>\pm</math>0.78</b>
Stone-steel towers	10	85	<b>98.89<math>\pm</math>0.08</b>	84.71 $\pm$ 5.88	74.71 $\pm$ 10.94	90.12 $\pm$ 1.93	96.24 $\pm$ 0.98	95.29 $\pm$ 4.16	86.82 $\pm$ 5.85	97.88 $\pm$ 1.08
OA			90.99 $\pm$ 1.81	79.27 $\pm$ 0.63	94.59 $\pm$ 0.46	97.40 $\pm$ 0.24	94.97 $\pm$ 0.98	94.86 $\pm$ 0.38	90.70 $\pm$ 0.82	<b>98.33<math>\pm</math>0.32</b>
AA			81.07 $\pm$ 1.09	71.91 $\pm$ 2.13	86.84 $\pm$ 2.67	95.31 $\pm$ 1.29	93.83 $\pm$ 1.32	92.20 $\pm$ 1.66	87.63 $\pm$ 2.00	<b>95.80<math>\pm</math>1.44</b>
$\kappa$			89.76 $\pm$ 0.83	75.88 $\pm$ 0.75	93.83 $\pm$ 0.52	97.04 $\pm$ 0.28	94.14 $\pm$ 0.40	94.13 $\pm$ 0.44	89.32 $\pm$ 0.94	<b>98.09<math>\pm</math>0.36</b>

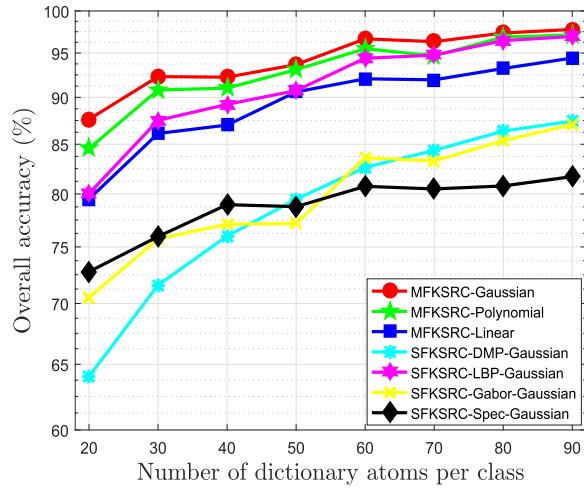


Fig. 6. OA results of the proposed MFKSRC with three kernels over a different number of dictionary atoms on the Indian pines image.

multiple-scale contextual information-based method (MASR). Moreover, the proposed methods also show a better performance than do the other two state-of-the-art methods, i.e., NRS and SVM-CK.

3) *Influences of Dictionary Size*: To evaluate the effect of dictionary size of the proposed multikernel-based method, we examine the OA results (shown in Fig. 6) for MFKSRC with three kernels (i.e., linear, polynomial, and Gaussian) over a different number of dictionary atoms on the Indian Pines image (nine classes). SFKSRC with different features based on the Gaussian kernel is selected as baseline methods. We randomly select 20–90 labeled pixels per class as dictionary sets, and the remaining samples act as a test set. From Fig. 6, among three multikernel methods based on multiple-feature space, the Gaussian-kernel method yields the best performance; polynomial-kernel method is better than linear-kernel method. Moreover, multifeature-kernel-based methods outperform single-feature-kernel methods.

4) *Evaluation of Kernel Projected Subspace*: To find an optimal number of kernel subspace by using KPCA, the

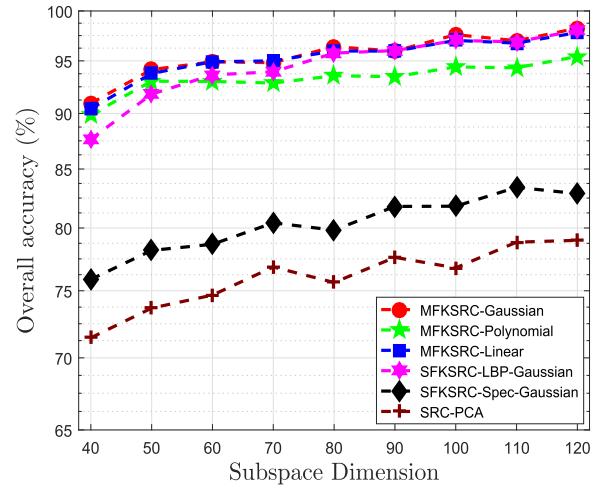


Fig. 7. OA results of the proposed MFKSRC method over different dimensionalities of subspaces on the AVIRIS Indian pines image.

averaged OA results of the MFKSRC with different reduced space dimensions of 40, 50, 60, 70, 80, 90, 100, 110, and 120 are reported in Fig. 7. For comparison purpose, a standard SR-based method with PCA as a preprocessing step (namely, SRC-PCA) and two single-feature-based SR methods (SFKSRC-Spec-Gaussian and SFKSRC-LBP-Gaussian) are considered in the experiment. As observed in Fig. 7, the OA performance of SRC-PCA, two SFKSRC, and the MFKSRC with different kernels slightly improve along with the number of projected dimension increase. Moreover, the proposed multifeature kernel-based classifier has a higher OA performance than that of single-feature and nonkernel-based classifiers on all projected subspaces.

### C. Experimental Results of ROSIS University of Pavia

The classification results (OA, AA, and  $\kappa$ ) averaged over 10 MC runs of the proposed method and its competitors are presented in Table III for the University of Pavia image. Sixty labeled pixels per class are randomly chosen as dictionary

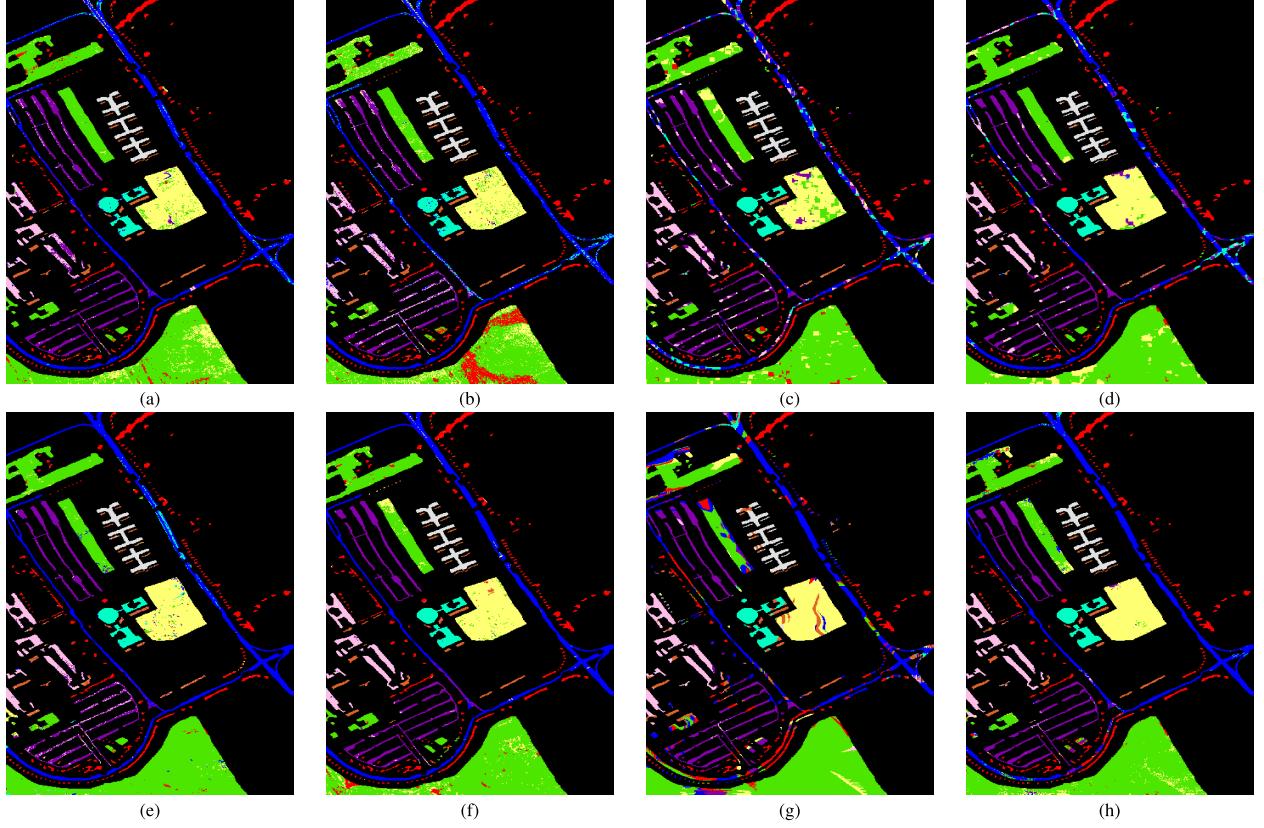


Fig. 8. Classification maps obtained by different classifiers for the University of Pavia image with nine classes. (a) SVM-CK [54]. (b) NRS [55]. (c) SOMP [28]. (d) MASR [36]. (e) MNFL [22]. (f) MFASR [56]. (g) SFKSRC. (h) MFKSRC.

TABLE III  
CLASSIFICATION ACCURACY (%) OBTAINED BY DIFFERENT CLASSIFIERS ON THE UNIVERSITY OF PAVIA IMAGE  
WITH 60 DICTIONARY ATOMS (OR TRAINING SAMPLES) FOR EACH CLASS

Class	#Sample	SVM-CK [54]	NRS [55]	SOMP [28]	MASR [36]	MNFL [22]	MFASR [56]	SFKSRC	MFKSRC	
	Train	Test								
Alfalfa	60	6571	<b>97.67±1.95</b>	83.55±0.79	54.26±3.42	62.53±4.56	92.77±2.24	96.88±1.37	61.31±0.95	92.88±1.18
Meadows	60	18589	89.39±2.66	78.49±4.68	90.82±2.35	91.75±1.41	<b>94.86±2.18</b>	89.29±1.58	82.29±0.85	92.62±0.97
Gravel	60	2039	80.61±3.59	75.64±3.74	94.69±1.39	95.34±2.45	91.90±2.51	98.14±0.61	86.91±1.28	<b>99.35±0.37</b>
Trees	60	3004	95.76±1.18	95.50±2.62	83.33±2.50	94.57±2.80	95.23±2.48	97.64±1.07	61.47±2.60	<b>98.35±0.24</b>
Metal sheet	60	1285	99.49±0.30	99.60±0.15	96.34±1.10	<b>100±0.00</b>	99.95±0.02	99.92±0.03	87.99±5.30	99.97±0.04
Bara soil	60	4969	83.10±3.88	88.16±3.41	91.36±2.66	92.35±2.06	94.51±1.20	95.86±0.95	92.84±1.61	<b>98.82±0.61</b>
Bitumen	60	1270	92.61±1.18	91.42±2.67	97.56±1.71	99.69±0.16	98.24±0.33	<b>99.80±0.04</b>	98.32±0.48	99.53±0.69
Brick	60	3622	87.47±3.40	84.11±5.80	91.08±4.36	89.73±1.44	87.52±3.14	96.77±1.48	84.96±6.17	<b>97.23±0.90</b>
Shadows	60	887	99.75±0.22	99.41±0.20	70.57±6.26	65.39±2.71	83.65±3.93	<b>100±0.00</b>	66.78±1.05	97.93±1.21
OA			88.87±0.81	83.44±1.98	84.82±1.30	87.41±0.86	93.77±1.13	93.96±0.64	80.85±1.38	<b>95.06±0.70</b>
AA			90.65±0.57	88.43±0.22	85.56±0.28	87.93±0.84	93.18±0.92	97.14±0.27	80.32±1.50	<b>97.41±0.41</b>
$\kappa$			85.38±1.02	78.79±2.33	80.20±1.65	83.57±1.13	91.78±1.48	91.87±0.82	73.74±1.70	<b>93.52±0.90</b>

atoms (or training samples) for SR-based methods (or SVM), and the remaining labeled pixels as the test set. The regularizer parameters for MFKSRC and SFKSRC are both set as  $1e - 5$  for the University of Pavia image. The subspace dimension is set as 100 for MFKSRC and SFKSRC. As can be observed, by combining multiple features in kernel projected subspace, the proposed MFKSRC yields a better classification performance (an OA of 95.06%) than do the two typical multiple feature methods (MFASR and MNFL). Although the proposed SFKSRC has a lower classification performance than that of single-scale contextual-feature-based method (SOMP), the proposed MFKSRC significantly outperforms the single- and multiple-scale contextual-feature-based methods (SOMP)

and MASR), and also outperforms other compared classifiers (including NRS and SVM-CK). Moreover, the corresponding classification maps obtained by the classifiers as mentioned earlier are shown in Fig. 8. Then, the averaged OA results of the proposed MFKSRC method with three kernels and SFKSRC with different features based on the Gaussian kernel along with a different number of dictionary atoms per class are presented in Fig. 9. MFKSRC with different kernels shows a better and more stable performance than does the single-feature-based methods along with the number of dictionary atoms per class from 20 to 90, in which 500 pixels per class are randomly chosen to be test samples. Fig. 10 shows the OA results of MFKSRC with multiple kernels and

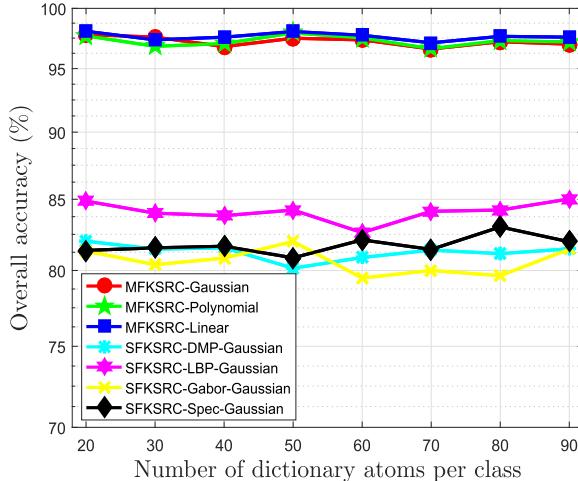


Fig. 9. OA results of the proposed MFKSRC with three kernels over a different number of dictionary atoms on the University of Pavia image.

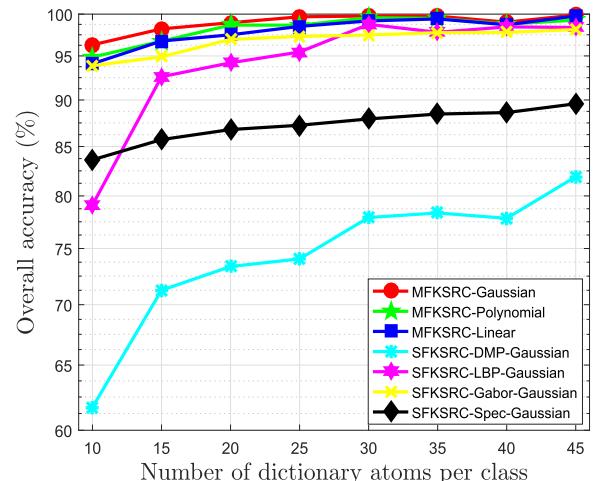


Fig. 11. OA results of the proposed MFKSRC with three kernels over a different number of dictionary atoms on the KSC image.

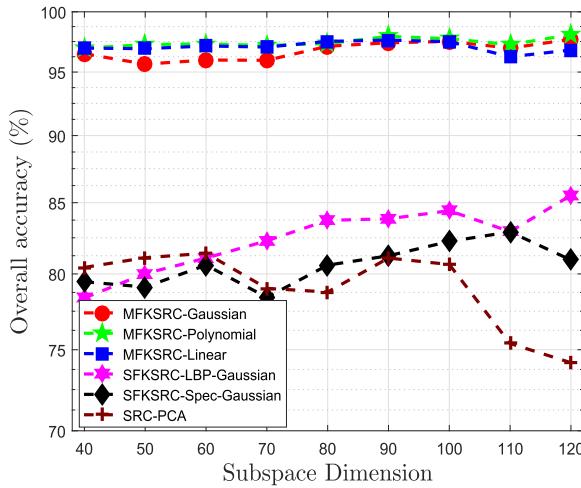


Fig. 10. OA results of the proposed MFKSRC method over different dimensionalities of subspaces on the University of Pavia image.

its compared methods (two SFKSRC and SRC-PCA) under different dimensionalities of reduced subspace in the range from 40 to 120. In this case, 60 labeled pixels per class are randomly chosen as a dictionary set, and 500 pixels per class are randomly chosen to a test set. It can be seen that MFKSRC consistently outperforms single-feature-based method and the standard SR-based method with PCA as a preprocessing step.

#### D. Experimental Results of Kennedy Space Center Image

The global accuracies (OA, AA, and  $\kappa$ ) obtained by averaging 10 MC runs of the proposed MFKSRC, and its competitors are reported in Table IV for the KSC image. In the experiments, we randomly choose 10 labeled pixels per class as dictionary atoms and the remaining as the test set. The regularizer parameters for SFKSRC and MFKSRC are set to be  $1e-5$  and  $1e-4$  for the KSC image. The subspace dimension is set to 80 for the proposed methods. It can be observed from Table IV that the proposed methods (SFKSRC and MFKSRC) outperform other state-of-the-art classifiers [including two multiple feature methods (MFASR and MNFL), two spatial contextual-based classifiers (SOMP and MASR), NRS, and

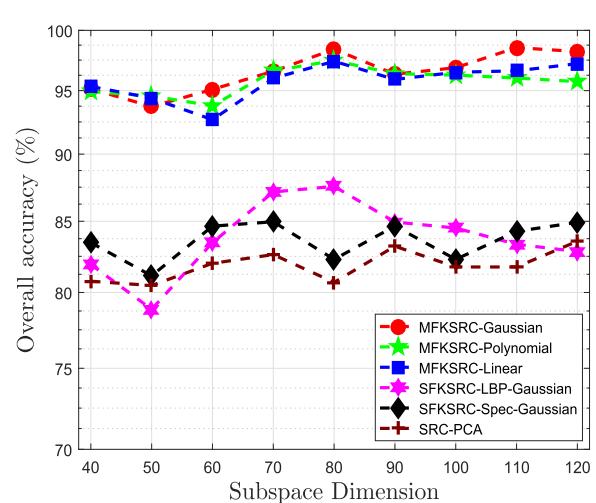


Fig. 12. OA results of the proposed MFKSRC method over different dimensionalities of subspaces on the KSC image.

SVM-CK] in terms of classification performance for KSC image. In Fig. 11, we plot the averaged OA results of the MFKSRC with three widely used kernel methods under a different number of dictionary atoms for the KSC image. It can be seen that MFKSRC with various kernels have similar OA performance, and outperform single-feature-based method. Fig. 12 illustrates the influence of subspace dimension on the OA results of the proposed MFKSRC. From Fig. 12, MFKSRC with different kernels has a stable OA results over different dimensionalities of subspace in the range from 40 to 120 for the KSC image and outperforms than single-feature-based method and the standard SR-based classifier with PCA as a preprocessing step.

#### E. Running Time and Statistical Difference

In this section, we first investigate the running time of the SFKSRC and MFKSRC, and the averaged CPU times over 10 MC runs of the two methods with a different number of dictionary atoms per class and subspace dimension on two HSIs are presented in Tables V and VI, respectively. For the

TABLE IV  
CLASSIFICATION ACCURACY (%) OBTAINED BY DIFFERENT CLASSIFIERS ON THE KSC IMAGE WITH 10 DICTIONARY ATOMS (OR TRAINING SAMPLES) FOR EACH CLASS

Class	#Sample		SVM-CK [54]	NRS [55]	SOMP [28]	MASR [36]	MNFL [22]	MFASR [56]	SFKSRC	MFKSRC
	Train	Test								
Scrub	10	751	94.54±10.50	82.77±10.00	85.27±1.48	94.35±4.98	88.39±3.05	89.56±4.31	89.56±3.87	<b>96.51±3.21</b>
Willow swamp	10	233	71.24±9.68	85.41±5.68	86.87±8.19	84.38±13.83	78.03±8.56	76.91±10.90	<b>86.70±1.80</b>	83.95±7.95
Cabbage palm	10	246	93.50±2.99	89.11±4.01	97.64±3.01	98.54±1.96	84.39±7.14	93.66±2.29	94.39±8.16	<b>99.35±0.94</b>
Cabbage palm/oak	10	242	57.85±12.83	45.54±14.98	86.28±9.60	85.37±8.80	45.37±4.36	66.28±9.87	<b>97.52±2.50</b>	95.95±4.14
Slash pine	10	151	73.51±7.83	67.68±9.77	91.13±8.35	93.77±6.84	74.97±5.47	93.38±4.54	93.91±8.80	<b>94.83±8.24</b>
Oak/broadleaf	10	219	57.53±9.50	57.53±11.45	92.05±4.20	86.58±6.97	75.62±7.60	82.56±5.16	84.02±7.79	<b>96.26±2.22</b>
Hardwood swamp	10	95	90.53±4.53	91.16±5.70	99.79±0.47	<b>100±0.00</b>	90.95±10.28	96.42±3.69	<b>100 ± 0.00</b>	<b>100 ± 0.00</b>
Graminoid marsh	10	421	85.27±4.58	74.44±6.02	64.66±4.74	74.92±14.76	77.72±5.27	83.18±9.17	93.49±4.27	<b>99.05±0.99</b>
Spartina marsh	10	510	95.49±2.66	89.29±13.86	96.00±3.79	<b>99.53±0.77</b>	91.25±1.92	88.08±4.14	94.86±10.08	92.75±9.68
Cattail marsh	10	394	71.57±9.93	87.46±2.85	90.66±7.46	95.94±2.23	88.02±2.55	70.46±9.31	<b>100±0.00</b>	95.28 ± 3.88
Salt marsh	10	409	97.31±3.06	95.01±3.17	98.29±1.52	97.51±3.99	89.44±7.48	94.13±4.26	95.94±5.23	<b>99.71±0.14</b>
Mud flats	10	493	86.21±15.34	79.55±3.39	67.18±9.35	83.94±7.66	56.63±4.45	72.13±6.80	90.95±6.60	<b>96.55±2.67</b>
Water	10	917	94.11±4.27	97.93±0.00	94.46±8.89	<b>100±0.00</b>	79.08±4.99	98.06±0.34	96.92 ± 3.21	99.91±0.14
OA			86.26±3.17	83.78±2.19	87.47±2.17	92.72±2.89	79.47±0.99	86.06±0.86	93.77±1.16	<b>96.66±0.99</b>
AA			82.21±2.36	80.22±2.09	88.48±2.17	91.91±3.07	78.45±0.83	84.98±0.82	93.71±1.04	<b>96.16±0.95</b>
$\kappa$			84.68±3.47	81.96±2.42	86.08±2.39	91.89±3.21	77.29±1.07	84.49±0.95	93.07±1.29	<b>96.28±1.11</b>

TABLE V

RUNNING TIME (SECONDS) OBTAINED BY THE PROPOSED METHOD WITH A DIFFERENT NUMBER OF DICTIONARY ATOMS PER CLASS AND DIFFERENT DIMENSIONS OF PROJECTED SPACE FOR THE INDIAN PINES IMAGE (NINE CLASSES)

	Number of dictionary atoms per class					
	20	30	40	50	60	70
SFKSRC	15.09	27.24	35.01	51.10	69.18	78.35
MFKSRC	120.78	202.38	270.18	356.44	440.25	518.71
Number of projected dimensionality						
	60	70	80	90	100	110
SFKSRC	59.25	74.24	89.05	100.83	118.03	137.61
MFKSRC	290.92	391.99	499.36	620.94	739.37	882.96

TABLE VI

RUNNING TIME (SECONDS) OBTAINED BY THE PROPOSED METHOD WITH A DIFFERENT NUMBER OF DICTIONARY ATOMS PER CLASS AND DIFFERENT DIMENSIONS OF PROJECTED SPACE FOR THE UNIVERSITY OF PAVIA IMAGE

	Number of dictionary atoms per class					
	20	30	40	50	60	70
SFKSRC	66.26	67.62	65.65	66.26	62.17	65.08
MFKSRC	209.91	210.00	210.58	209.47	205.17	207.21
Number of projected dimensionality						
	60	70	80	90	100	110
SFKSRC	223.23	241.05	249.68	271.84	268.59	284.69
MFKSRC	1337.24	1711.19	1831.74	1961.28	2103.40	2183.40

proposed methods, the dominant computational cost comes from the learning of sparse coding via  $\ell_1$ -norm regularization problem in multiple-feature-based kernel subspace, which is implemented by the Least Angle Regressions method via the SPArse Modeling Software package [60], [61]. In the experiments, 20–70 labeled pixels are chosen as training dictionary per class, the remaining labeled pixels are used as test samples for the Indian Pines image, and 500 labeled pixels are used as test samples for the University of Pavia image. From Tables V and VI, MFKSRC costs more time than does the SFKSRC owing to multitask learning. When dealing with an increasing number of dictionary atoms, much computing time is required for SFKSRC and MFKSRC. Likewise, as the projected dimensionality increases, the computation time for SFKSRC and MFKSRC increases.

Considering the classification results obtained by the proposed MFKSRC with different kernels are apparently similar,

TABLE VII  
MCNEMAR'S TEST OF THE PROPOSED MFKSRC WITH MULTIPLE KERNEL METHODS

	20	30	40	50	60	
	Gaussian vs. Poly	29.35	25.04	21.21	20.05	18.41
Indian	Gaussian vs. Linear	19.39	11.83	9.90	1.41	4.47
	Linear vs. Poly	22.05	22.07	18.76	20.00	17.86
	Gaussian vs. Poly	-4.24	1.73	4.24	-3.00	-6.63
Pavia	Gaussian vs. Linear	7.07	4.36	5.20	2.24	-4.47
	Linear vs. Poly	-8.24	-4.00	-3.00	-3.74	-4.89
	Gaussian vs. Linear	6.63	6.48	5.38	2.45	2.83
KSC	Gaussian vs. Poly	8.49	4.69	6.00	1.73	2.24
	Linear vs. Poly	-5.29	4.47	-2.65	1.73	1.73

we analyze the statistical differences among all the considered kernel methods using McNemar's test. If the test statistic  $|Z| > 1.96$ , the difference in accuracy between two classifiers is regarded as statistically significant at the 5% level of significance [62]. As it can be observed from Table VII, for the Indian Pines image, the Gaussian-kernel-based method is superior to ploy- and linear-kernel-based methods, and linear-kernel-based method is superior to the ploy-kernel-based method. For the University of Pavia image, the Gaussian method is superior to the polynomial method in a few cases and is worse than comparable in most cases. The Gaussian method is superior to the linear method in most cases, and linear method is worse than polynomial method ( $|Z| < -1.96$ ). For the KSC image, the Gaussian-kernel-based method is superior to ploy- and linear-kernel-based methods in most cases; the linear method is worse than the polynomial method in a few cases, and is superior to the polynomial method in most cases.

## V. CONCLUSION

In this paper, we propose a novel MFKSRC for HSI classification, which seeks a low-dimensional representation in multifeature-based kernel space, in which MFKSRC achieves better classification and becomes more efficient. To cope with the nonlinear distribution of multiple features, kernel tricks are applied to carry original input space of different descriptors to high-dimensional kernel feature space, which leads to the time-consuming procedure or even infeasible optimization for

the succeeding SR-based model ( $\ell_1$ -minimization problem). To overcome this problem, we adopt KPCA as a preprocessing stage in the multifeature-based kernel space to convert the problem into a feasible optimization problem. Subsequently, these learned multiple features of HSI in its reduced kernel space are then extended to multiple feature SR-based model to obtain the multifeature-kernel-based representation coefficients. The decision rule of the proposed MFKSRC builds on the minimal residual between the test pixel and its approximations obtained over each class subdictionaries in their reduced kernel space. Experimental results on three commonly used HSI data sets confirm that the MFKSRC methods can provide a satisfying classification performance, compared with the other state-of-the-art SR-based algorithms and SVMs. The future work is to study the extended MFKSRC based on multiple DR techniques.

#### ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the two anonymous reviewers for providing truly outstanding comments and suggestions that significantly helped in improving the technical quality and presentation of this manuscript.

#### REFERENCES

- [1] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. S110–S122, Sep. 2009.
- [2] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, 2007.
- [3] M. Borengasser, W. S. Hungate, R. Watkins, and O. Beer, *Hyperspectral Remote Sensing: Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2008.
- [4] Q. Tong, Y. Xue, and L. Zhang, "Progress in hyperspectral remote sensing science and technology in China over the past three decades," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 70–91, Jan. 2014.
- [5] 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.
- [6] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2007.
- [7] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 318–322, Mar. 2013.
- [8] J. Li, J. M. Bioucas-Dias, and A. J. 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.
- [9] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2271–2282, May 2010.
- [10] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.
- [11] J. Xia, N. Falco, J. A. Benediktsson, P. Du, and J. Chanussot, "Hyperspectral image classification with rotation random forest via KPCA," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1601–1609, Apr. 2017.
- [12] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, May 2015.
- [13] J. Xia, M. D. Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, Sep. 2015.
- [14] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
- [15] S. Jia, L. Shen, and Q. Li, "Gabor feature-based collaborative representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1118–1129, Feb. 2015.
- [16] M. Musci, R. Q. Feitosa, G. A. O. P. Costa, and M. L. F. Velloso, "Assessment of binary coding techniques for texture characterization in remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1607–1611, Nov. 2013.
- [17] L. Gan, J. Xia, P. Du, and J. Chanussot, "Class-oriented weighted kernel sparse representation with region-level kernel for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published.
- [18] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.
- [19] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, Dec. 2015.
- [20] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.
- [21] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 221–228.
- [22] 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.
- [23] P. Zhong and R. Wang, "A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3978–3988, Dec. 2007.
- [24] 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.
- [25] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [26] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proc. CVPR*, Jun. 2011, pp. 625–632.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] J. Li, H. Zhang, and L. Zhang, "Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification," *ISPRS J. Photogram. Remote Sens.*, vol. 94, pp. 25–36, Aug. 2014.
- [31] E. Zhang, X. Zhang, H. Liu, and L. Jiao, "Fast multifeature joint sparse representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1397–1401, Jul. 2015.
- [32] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.
- [33] W. Li, Q. Du, F. Zhang, and W. Hu, "Hyperspectral image classification by fusing collaborative and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4178–4187, Sep. 2016.
- [34] L. Gan, P. Du, J. Xia, and Y. Meng, "Kernel fused representation-based classifier for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 684–688, May 2017.
- [35] L. Gan, J. Xia, P. Du, and Z. Xu, "Dissimilarity-weighted sparse representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1968–1972, Nov. 2017.
- [36] 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.

- [37] S. Ozawa, A. Roy, and D. Roussinov, "A multitask learning model for online pattern recognition," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, pp. 430–445, Mar. 2009.
- [38] X.-T. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, Oct. 2010.
- [39] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.
- [40] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923–5936, Sep. 2014.
- [41] J. Li, H. Zhang, and L. Zhang, "A nonlinear multiple feature learning classifier for hyperspectral images with limited training samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2728–2738, Jun. 2015.
- [42] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition," in *Proc. ECCV*, 2010, pp. 1–14.
- [43] L. Zhang *et al.*, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.
- [44] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Sparse representation with kernels," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 423–434, Feb. 2013.
- [45] J. Yin, Z. Liu, Z. Jin, and W. Yang, "Kernel sparse representation based classification," *Neurocomputing*, vol. 77, no. 1, pp. 120–128, Feb. 2012.
- [46] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Advances in Kernel Methods—Support Vector Learning*. Cambridge, MA, USA: MIT Press, 1999, pp. 327–352.
- [47] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognit.*, vol. 43, no. 3, pp. 706–719, 2010.
- [48] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
- [49] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations," *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 209, pp. 415–446, May 1909.
- [50] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
- [51] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [52] D. A. Clausi and M. E. Jernigan, "Designing Gabor filters for optimal texture separability," *Pattern Recognit.*, vol. 33, no. 11, pp. 1835–1849, Nov. 2000.
- [53] D. Tuia, F. Pacifici, M. Kanevski, and W. J. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, Nov. 2009.
- [54] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [55] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2014.
- [56] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017.
- [57] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 2, no. 7, pp. 1160–1169, Jul. 1985.
- [58] 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.
- [59] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.
- [60] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.
- [61] J. Mairal, F. R. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. ICML*, 2009, pp. 689–696.
- [62] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.



**Le Gan** received the M.S. degree in cartography and geographic information system from Yangtze University, Wuhan, China, in 2015. He is currently pursuing the Ph.D. degree with the Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing, China.

His research interests include hyperspectral image classification, sparse representation, dictionary learning, deep learning, and time-series analyses.

Dr. Gan was a recipient of the National Scholarship for Doctoral Graduate Students granted by the Ministry of Education of the People's Republic of China in 2017, and the Scholarship granted by the Collaborative Innovation Center of South China Sea Studies.



**Junshi Xia** (S'11–M'16) received the B.S. degree in geographic information systems and the Ph.D. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2008 and 2013, respectively, and the Ph.D. degree in image processing from the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology, Grenoble, France, in 2014.

From 2014 to 2015, he was a Visiting Scientist with the Department of Geographic Information Sciences, Nanjing University, Nanjing, China. From 2015 to 2016, he was a Post-Doctoral Research Fellow with the University of Bordeaux, Talence, France. Since 2016, he has been the JSPS Post-Doctoral Overseas Research Fellow with the University of Tokyo, Tokyo, Japan. His research interests include multiple classifier system in remote sensing, hyperspectral remote sensing image processing, and urban remote sensing.

Dr. Xia won the first place in the IEEE Geoscience and Remote Sensing Society Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee in 2017.



**Peijun Du** (M'07–SM'12) received the Ph.D. degree from the China University of Mining and Technology, Xuzhou, China, in 2001.

He is currently a Professor of remote sensing and geographical information science with Nanjing University, Nanjing, China. He has authored over 70 articles in international peer-reviewed journals, and over 100 papers in international conferences and Chinese journals. His research interests include remote sensing image processing and pattern recognition, hyperspectral remote sensing, and applications of geospatial information technologies.

Dr. Du served as the Co-Chair of the Technical Committee of URBAN in 2009, IAPR-PRRS in 2012, EORSA in 2014, IEOAs in 2015, and CCGC in 2015, the Co-Chair of the Local Organizing Committee of JURSE in 2009, WHISPERS in 2012, and EORSA in 2012, and a member of the Scientific Committee or the Technical Committee of other international conferences, including WHISPERS during 2010–2016, URBAN in 2011, 2013, and 2015, MultiTemp in 2011, 2013, and 2015, ISIDF in 2011, and SPIE European Conference on Image and Signal Processing for Remote Sensing during 2012–2016. He is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.



**Jocelyn Chanussot** (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Générale de l'Armement (DGA–French National Defense Department), Toulon, France. Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble. He has been a Visiting Scholar at Stanford University, Stanford, CA, USA, the KTH Royal Institute of Technology, Stockholm, Sweden, and the National University of Singapore, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavík, Iceland. From 2015 to 2017, he was a Visiting Professor at the University of California at Los Angeles, Los Angeles, CA, USA. His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing.

Dr. Chanussot was a member of the IEEE Geoscience and Remote Sensing Society AdCom from 2009 to 2010, where he was in charge of membership

development and the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008. He is a member of the Institut Universitaire de France, Paris, France, from 2012 to 2017. He is the Founding President of the IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010 which received the 2010 IEEE GRS-S Chapter Excellence Award. He was a co-recipient of the NORSIG 2006 Best Student Paper Award, the IEEE Geoscience and Remote Sensing Society (GRSS) 2011 and 2015 Symposium Best Paper Award, the IEEE GRSS 2012 Transactions Prize Paper Award, and the IEEE GRSS 2013 Highest Impact Paper Award. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing. He was the Chair from 2009 to 2011 and the Co-Chair of the GRS Data Fusion Technical Committee from 2005 to 2008. He was the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He was an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS from 2005 to 2007 and *Pattern Recognition* from 2006 to 2008. Since 2007, he has been an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. He was a Guest Editor for the PROCEEDINGS OF THE IEEE in 2013 and the *IEEE Signal Processing Magazine* in 2014.