

# Multikernel Adaptive Collaborative Representation for Hyperspectral Image Classification

Peijun Du<sup>1</sup>, Senior Member, IEEE, Le Gan<sup>1</sup>, Member, IEEE, Junshi Xia<sup>2</sup>, Member, IEEE, and Daming Wang

**Abstract**—To adequately represent the nonlinearities in the high-dimensional feature space for hyperspectral images (HSIs), we propose a multiple kernel collaborative representation-based classifier (CRC) in this paper. Extended morphological profiles are first extracted from the original HSIs, because they can efficiently capture the spatial and spectral information. In the proposed method, a novel multiple kernel learning (MKL) model is embedded into CRC. Multiple kernel patterns, e.g., Naive, Multimetric, and Multiscale are adopted for the optimal set of basic kernels, which are helpful to capture the useful information from different pixel distributions, kernel metric spaces, and kernel scales. To learn an optimal linear combination of the predefined basic kernels, we add an extra training stage to the typical CRC where kernel weights are jointly learned with the representation coefficients from the training samples by minimizing the representation error. Moreover, by considering different contributions of dictionary atoms, the adaptive representation strategy is applied to the MKL framework via a dissimilarity-weighted regularizer to obtain a more robust representation of test pixels in the fused kernel space. Experimental results on three real HSIs confirm that the proposed classifiers outperform the other state-of-the-art representation-based classifiers.

**Index Terms**—Collaborative representation (CR), extended morphological profiles (EMPs), hyperspectral image (HSI) classification, multiple kernel learning (MKL).

## I. INTRODUCTION

**H**YPERSPECTRAL images (HSIs), which contain rich spectral information in hundreds of narrow contiguous spectral bands, have been widely applied in a variety of applications, such as target detection [1], anomaly detection [2], [3], and classification [4]. Classification is an active topic in hyperspectral remote sensing. In recent years, different kinds of classifiers, such as support vector machines (SVMs) [4], [5], multinomial logistic regression [6],

extreme learning machine [7], and random forests [8], have been proposed for HSIs. Recently, representation-based classifiers [9], [10], which do not consider any sample prior distribution, gradually become a hot topic in HSI classification tasks [11]–[17]. The basic idea of such classifiers is that a test pixel can be linearly represented by some labeled samples, namely, dictionary. The class label of a given test pixel can be determined according to the class whose subdictionary provides the minimum representation error.

The representation coefficients can be solved by optimizers with different regularizations. The available methods can be divided into two main categories: sparse representation (SR) with  $\ell_0$ -norm or  $\ell_1$ -norm minimization and collaborative representation (CR) with  $\ell_2$ -norm minimization. In [9], the SR model, which assumes that a test pixel can be linearly represented by only a few atoms from the training dictionary, was initially proposed for face recognition. In CR-based classifier (CRC) [10], the test pixel is linearly represented by all the training atoms. Based on different representation mechanisms, all the atoms have an equal opportunity to participate in the representation of a given test pixel in the CR-based model, whereas only a few atoms have the chance to represent the test pixel in the SR-based model [18]. In [19], the SR-based model was adopted for HSI classification task and achieved a good classification result. In [20], *et al.* proposed a fused representation-based classification framework by integrating the SR- and CR-based model via a suitable balance parameter in residual domain.

Due to the fact that the number of dictionary atoms is very limited and the spectral features are strongly correlated, the traditional representation based-classifiers often achieve very noisy results. To achieve satisfactory classification performance, different strategies, such as spatial joint representation model [11], [21], adaptive representation [22], spectral-spatial features [15], [23], and kernel strategy [24], [25], are applied to the traditional representation-based classifiers. Integrating spatial neighborhood information to SRC, Chen *et al.* [11] proposed a joint SRC (JSRC) via a joint sparsity model to increase the accuracy of SRC because neighboring pixels usually belong to the same class. Inspired by JSRC, Li and Du [21] further proposed a joint CRC (JCRC) by jointly using the neighboring spatial information. Besides, adaptive representation via dissimilarity-weighted regularization is widely applied to representation-based classifiers. In [26] and [27], weighted SRC adaptively exploits the similarity between the test pixel and each dictionary atom in the SR

Manuscript received December 3, 2017; revised March 21, 2018; accepted April 30, 2018. Date of publication June 5, 2018; date of current version July 20, 2018. This work was supported by the National Natural Science Foundation of China under Grant 41471275 and Grant 41402293. (Corresponding authors: Peijun Du; Le Gan.)

P. Du and L. Gan 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, also with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China, and also with the Collaborative Innovation Center of South China Sea Studies, Nanjing 210093, China (e-mail: dupjrs@gmail.com; ganleatlas@gmail.com).

J. Xia is with the RIKEN Center for Advanced Intelligence Project, RIKEN, Tokyo 103-0027, Japan (e-mail: junshi.xia@riken.jp).

D. Wang is with the Survey Center for Oil and Gas Resource, China Geological Survey, Beijing 100029, China (e-mail: daming82@qq.com).

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.2018.2833882

0196-2892 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See [http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) for more information.

model. Gan *et al.* [28] proposed an adaptive SRC based on the locality constrained dictionary for HSI classification. In [22], a CRC, called nearest regularized subspace (NRS), was proposed for HSI classification, where a distance-weighted Tikhonov regularization was adopted to calculate the similarity between the test pixel and the within-class dictionary atoms. Moreover, the resulted coefficients carry a meaning that they reflect the relative importance of each dictionary atom.

In addition, many studies [29]–[36] demonstrate that spectral–spatial features can effectively enhance the classification performance of HSI. Kang *et al.* [29] proposed a novel principal component analysis (PCA)-based edge-preserving features method for HSI classification. In [30], a Gabor filtering-based NRS classifier was investigated. Jia *et al.* [31] proposed a 3-D Gabor feature-based CR model. In [32], an extended random walker-based method was proposed for spectral–spatial classification. Combining extended morphological profiles (EMPs) with classifiers is one of the widely used strategies for HSI classification. In [33] and [34], EMPs are used to exploit the spatial information, which can improve the classification performance. When samples are not linearly separable, linear representation is inadequate for representing the nonlinear structures of samples. To address this issue, Gao *et al.* [24] proposed a nonlinear SRC (KSRC) to project the samples into a high-dimensional feature space using the kernel trick, where class separability can be improved. In [25], the kernel version of a joint SRC was proposed and denoted as KJSRC. In [14], a kernel CR with Tikhonov regularization was introduced. In the kernel-induced space, nonlinear representation-based classifiers can achieve better accuracy than the linear representation-based classifiers. However, the type of kernel function and its corresponding parameters have to be selected appropriately, and it is difficult to determine. Moreover, the capability of a predefined kernel is insufficient in mining rich information from training dictionary [37]. A typical solution is to learn an appropriate combination of multiple kernel functions [38].

Recently, multiple kernel learning (MKL) methods, which exploit learning an optimal combination of multiple kernels instead of selecting a specific kernel function, have been proposed in the literature [38]–[42]. The basic idea of MKL methods focuses mainly on learning an optimal linear combination of a set of predefined kernels, known as basic kernels. Single kernel-based methods lack the generalization capability of dealing with samples with various distributions, multidimensional and multiclass data, and data containing heterogeneous information. Compared to the single-kernel case, MKL can provide a more flexible framework to combine multiple kernels with different capabilities. In the remote sensing literature, MKL has been successfully applied in HSI classification [37], [43]–[49]. In [43], a multiple kernel framework for combining and assessing the relevance of multiple source heterogeneous information in SVM-based classification was proposed. In [44], a representative MKL was proposed to find the max-variance kernel by learning the linear combination of basic kernels and minimizing the Frobenius norm error. In [45], an MKL model was proposed for urban classification to integrate heterogeneous information from the spectral

images and the light detection and ranging data. In [46], a superpixel-based classification framework via multiple kernels was proposed. In [47], a discriminant MKL (DMKL) was proposed to learn an optimal multiple kernel combination from predefined basic kernels by maximizing separability in reproduction kernel Hilbert space. Although those MKL-based methods can achieve better classification performance than single kernel-based methods, how to appropriately construct a kernel combination and tune its kernel weights is an important research topic that we need to solve.

To enhance the discrimination power of CRCs, by embedding MKL into the CRC, we propose a multiple kernel CRC (MKCRC) framework in this paper. In the framework, PCA is first applied on the original HSI, and then spectral–spatial features, i.e., EMPs, are extracted by performing on the reserved PCs with a series of morphological opening and closing operations. Then, multiple discriminate basic kernels are constructed from the EMPs feature via different types (or scales) of kernel mappings, and those basic kernels are appropriately integrated into the typical CRC model with different kernel weights via a novel MKL framework. Specifically, a training stage is added to the standard CRC to learn an optimal linear combination of multiple kernels. Based on a minimum reconstruction error criterion, we can alternate between solving for CR coefficients with fixed kernel weights and, then, solving for kernel weights while keeping CR coefficients fixed. Furthermore, an adaptive strategy is also incorporated into MKCRC to obtain a more discriminative CR coefficient, and a multiple kernel adaptive CRC (MKACRC) is then proposed. The schematic of the proposed multiple kernel CR-based classification framework is shown in Fig. 1. Compared with the existing representation-based classifiers, the main contributions of this paper are as follows:

- 1) To enhance the discrimination ability of CRCs, we propose a multiple kernel CRC to integrate multiple discriminative kernels for HSI. The kernel weights of multiple kernels are determined by minimizing the representation error of a set of training samples.
- 2) Multiple available kernel patterns are applied to the proposed MKL framework. The adopted predefined kernels are categorized into three different types, i.e., Naive, Multimetric, and Multiscale.
- 3) An adaptive strategy is integrated into the traditional CR model using dissimilarity-weighted regularization term in the multiple kernel-induced space, which is helpful to integrate locality structure information for MKCRC.

The outline of this paper is organized as follows. In Section II, the typical CRCs and MKL are briefly reviewed. In Section III, the proposed multiple kernel CR-based classification framework is presented. The effectiveness of the proposed multiple kernel CRCs is investigated in Section IV. Concluding remarks are given in Section V.

## II. CR-BASED CLASSIFIERS AND MKL

Let  $\mathcal{D} = \{\mathcal{D}_c\}_{c=1}^C \in \mathbb{R}^{B \times N}$  be a class-specific dictionary set of  $C$  classes for HSI (with  $N$  training atoms), where  $B$  denotes the feature dimension. The subdictionary

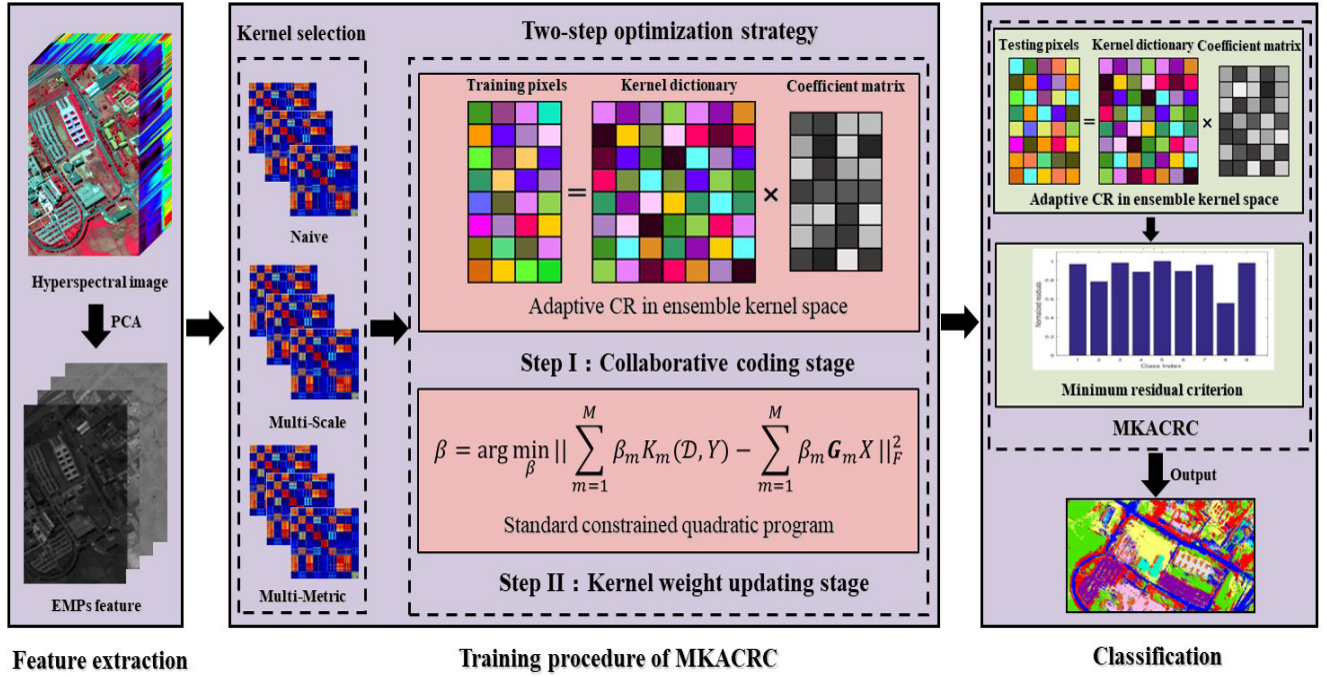


Fig. 1. Schematic of the proposed multiple kernel CR-based classification framework.

$\mathcal{D}_c = [\mathbf{d}_{c,1}, \mathbf{d}_{c,2}, \dots, \mathbf{d}_{c,N_c}]$  is associated with the  $c$ th class ( $\sum_{c=1}^C N_c = N$ ) and  $\mathbf{d}_{c,i}$  denotes the  $i$ th atom in  $\mathcal{D}_c$ .

#### A. CR-Based Classifiers

In CRC [10], a test pixel  $\mathbf{y}$  can be represented as a linear combination of all training atoms in dictionary  $\mathcal{D}$  with an  $\ell_2$ -regularizer on the coefficient vector. The coefficient vector  $\boldsymbol{\alpha} \in \mathbb{R}^N$  can be solved by an  $\ell_2$ -norm regularization, that is,

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathcal{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_2^2 \quad (1)$$

where  $\lambda$  denotes a global regularization parameter that balances the minimization between the approximation error and the  $\ell_2$ -norm term. Then, (1) can be solved with a closed-form solution as

$$\hat{\boldsymbol{\alpha}} = (\mathcal{D}^T \mathcal{D} + \lambda \mathbf{I})^{-1} \mathcal{D}^T \mathbf{y} \quad (2)$$

where  $\mathbf{I}$  denotes the identity matrix. Generally, samples of HSI are not linearly separable in original input space. To cope with the problem, kernel trick has been successfully applied to the HSI classification [50]. The linearly inseparable samples in original space  $\mathbb{R}^B$  are mapped into a high-dimensional reproducing kernel Hilbert space  $\mathcal{H}$  via nonlinear mapping, where samples become linearly separable. Let define the nonlinear mapping  $\Phi: \mathbb{R}^B \mapsto \mathcal{H}$  corresponding to a kernel function  $K(\mathbf{d}_i, \mathbf{d}_j) = \langle \Phi(\mathbf{d}_i), \Phi(\mathbf{d}_j) \rangle = \Phi(\mathbf{d}_i)^T \Phi(\mathbf{d}_j)$  [51]–[53]. Then, the CR-based representation in kernel space  $\mathcal{H}$  can be reformulated as follows:

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\mathbf{K}(\mathcal{D}, \mathbf{y}) - \mathbf{G}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_2^2 \quad (3)$$

where  $\mathbf{K}(\mathcal{D}, \mathbf{y}) = [K(\mathbf{d}_1, \mathbf{y}), \dots, K(\mathbf{d}_N, \mathbf{y})]^T \in \mathbb{R}^{N \times 1}$ , and  $\mathbf{G} = \Phi(\mathcal{D})^T \Phi(\mathcal{D}) \in \mathbb{R}^{N \times N}$  denotes the kernel Gram

matrix with  $G_{ij} = K(\mathbf{d}_i, \mathbf{d}_j)$  and  $\Phi(\mathcal{D}) = \{\Phi(\mathbf{d}_1), \Phi(\mathbf{d}_2), \dots, \Phi(\mathbf{d}_N)\}$ . The most widely used kernel is the Gaussian radial basis function (RBF) kernel, which is given as follows:

$$K(\mathbf{d}_i, \mathbf{d}_j) = \exp\left(-\frac{\|\mathbf{d}_i - \mathbf{d}_j\|^2}{2\sigma^2}\right), \quad \sigma \in \mathbb{R}^+ \quad (4)$$

where  $\sigma$  denotes the bandwidth parameter of Gaussian RBF kernel, which controls the smoothness of kernel measure. Similar to (2), the coefficient vector  $\boldsymbol{\alpha}$  can be obtained in a closed-form solution

$$\hat{\boldsymbol{\alpha}} = (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{G}^T \mathbf{K}(\mathcal{D}, \mathbf{y}) \quad (5)$$

After obtaining the coefficient vector  $\boldsymbol{\alpha}$  in kernel space  $\mathcal{H}$ , we can classify the test pixel  $\mathbf{y}$  based on the minimum residual criterion in the kernel-induced space as follows:

$$\text{class}(\mathbf{y}) = \arg \min_{c=1, \dots, C} \|\mathbf{K}(\mathcal{D}, \mathbf{y}) - \mathbf{G}_c \boldsymbol{\alpha}_c\|_2^2 \quad (6)$$

where  $\mathbf{G}_c$  denotes the  $c$ th subdictionary in kernel space, and  $\boldsymbol{\alpha}_c$  denotes the class-specific coefficients associated with the  $c$ th class.

#### B. Multiple Kernel Learning

Selecting a suitable kernel function and tuning its parameters are important issues in the kernel-based methods. A feasible method, known as MKL, is to adopt a linear combination function to integrate a series of predefined kernels into an ensemble kernel. The predefined kernel can be constructed from different sample distributions, metric spaces, or parameters (e.g., scales). The MKL framework aims at finding appropriate weight coefficients for best fusing the given basic kernels. After constructing a series of basic kernel



functions  $\{K_m\}_{m=1}^M$ , an ensemble kernel is then formulated as follows:

$$K(\mathbf{d}_i, \mathbf{d}_j) = \sum_{m=1}^M \beta_m K_m(\mathbf{d}_i, \mathbf{d}_j) \quad (7)$$

$$\text{s.t. } \beta_m \geq 0, \quad \text{and} \quad \sum_{m=1}^M \beta_m = 1$$

where  $M$  denotes the total number of the predefined basic kernels, and  $\beta_m$  denotes the weight of the  $m$ th basic kernel function  $K_m(\mathbf{d}_i, \mathbf{d}_j)$ . The ensemble kernel can extract richer discriminative information using multiple basic kernels from data samples than single kernel-based methods. In order to ensure the ensemble kernel satisfies the positive semidefinite condition, all the weight coefficients are nonnegative and sum to one [49].

### III. MULTIPLE KERNEL ADAPTIVE COLLABORATIVE REPRESENTATION-BASED CLASSIFIER

#### A. Problem Formulation

As the kernel CR model in Eq. (3) presents the adopted kernel plays a vital role for the problem. Determining a suitable kernel function and adjusting its parameters for the kernel CR model are challenging tasks. The MKL methods, which suggest adopting an appropriate combination of a series of basic kernel to instead of the single kernel function, are adopted to tackle this problem. The MKL framework not only allows the kernel-based methods from selecting an optimal basic kernel as the best kernel within a time-consuming procedure but also presents a fused framework by combining multiple heterogeneous information based on different discriminate kernel function. To make full use of the discriminative information contained in samples or features, by embedding the MKL model in a CR-based method, we propose a multiple kernel CR-based classification framework (namely, MKCRC) based on EMPs feature. In this framework, MKL is applied to capture information from different sample distributions, kernel metric spaces, and scale levels, and a joint multiple kernel CRC with different kinds of kernel patterns is achieved. To further exploit different contributions of the dictionary atoms in representing a test pixel, the adaptive version MKCRC, namely, MKACRC, is also proposed in this paper. The goal of the proposed multiple kernel CRCs is to learn an ensemble kernel (i.e., an optimal linear combination of a series of basic kernels) from a set of training samples. Specifically, a training stage is added to the traditional CRC to learn the kernel weights of basic kernels by minimizing the total reconstruction residual of the training samples, boosting the representation power of the training dictionary in an ensemble kernel space.

Let  $\mathcal{D} = \{\mathcal{D}_c\}_{c=1}^C \in \mathbb{R}^{B \times N}$  be a set of available training dictionary, where  $\mathcal{D}_c$  denotes the class-specific dictionary associated with the  $c$ th class. As stated previously, suppose we have a set of  $M$  candidate basic kernels  $\{K_m\}_{m=1}^M$ , which are produced by either  $M$  different kernel types such as linear, polynomial, and Gaussian kernels, by  $M$  different metric-based kernels, or by the sample kernel function, but with

$M$  different parameters (or scales). Given a set of training samples  $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_P\} \in \mathbb{R}^{B \times P}$  with  $P$  samples, and  $\mathbf{X} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_P\} \in \mathbb{R}^{N \times P}$  denotes the coding coefficient matrix of training samples  $\mathbf{Y}$  in the ensemble kernel space. Similar to (7), the corresponding multiple kernel Gram matrix  $\mathbf{G}$  can also be generated by a series of basic kernel Gram matrices  $\{\mathbf{G}_m\}_{m=1}^M$ , formulated as follows:

$$\mathbf{G} = \sum_{m=1}^M \beta_m \mathbf{G}_m \quad \text{s.t. } \beta_m \geq 0, \quad \text{and} \quad \sum_{m=1}^M \beta_m = 1. \quad (8)$$

Embedding the ensemble kernel into (3), and the CR of  $\mathbf{Y}$  over dictionary  $\mathcal{D}$  in the ensemble kernel space can be formulated as follows:

$$\langle \hat{\mathbf{X}}, \hat{\boldsymbol{\beta}} \rangle = \arg \min_{\mathbf{X}, \boldsymbol{\beta}} \left\| \sum_{m=1}^M \beta_m K_m(\mathcal{D}, \mathbf{Y}) - \sum_{m=1}^M \beta_m \mathbf{G}_m \mathbf{X} \right\|_F^2 + \lambda \|\mathbf{X}\|_F^2$$

$$\text{s.t. } \beta_m \geq 0, \quad \text{and} \quad \sum_{m=1}^M \beta_m = 1 \quad (9)$$

where  $K_m(\mathcal{D}, \mathbf{Y}) = \{K_m(\mathcal{D}, \mathbf{y}_1), \dots, K_m(\mathcal{D}, \mathbf{y}_P)\}$ , and  $\mathbf{G}_m$  denotes the  $m$ th basic kernel Gram matrix. The proposed multiple kernel CR model can alternately learn the discriminative CR coefficient matrix  $\mathbf{X}$  and the kernel weight vector  $\boldsymbol{\beta}$  for multiple kernels.

However, the multiple kernel CR model in Eq. (9) does not consider the dissimilarity (or distance) information between the unlabeled test pixel and each atom of the dictionary. The atoms that are far from the test pixel generally play less important role in reconstructing the test pixel. Considering that the contribution of the atoms in reconstructing the test pixel are different, the CR-based model in (9) is extended to an adaptive CR model where a different measurement between the test pixel and each dictionary atom is added to the  $\ell_2$ -norm regularization. By replacing the  $\ell_2$ -norm regularization with a dissimilar-weighted regularization, (9) can be reformulated as

$$\langle \hat{\mathbf{X}}, \hat{\boldsymbol{\beta}} \rangle = \arg \min_{\mathbf{X}, \boldsymbol{\beta}} \left\| \sum_{m=1}^M \beta_m K_m(\mathcal{D}, \mathbf{Y}) - \sum_{m=1}^M \beta_m \mathbf{G}_m \mathbf{X} \right\|_F^2 + \lambda \|\boldsymbol{\Gamma} \odot \mathbf{X}\|_F^2$$

$$\text{s.t. } \beta_m \geq 0, \quad \text{and} \quad \sum_{m=1}^M \beta_m = 1 \quad (10)$$

where  $\odot$  denotes elementwise multiplication and  $\boldsymbol{\Gamma} \in \mathbb{R}^{N \times P}$  denotes the biasing Tikhonov matrix with  $\Gamma_{i,j} = \text{dist}(\Phi(\mathbf{d}_i), \Phi(\mathbf{y}_j))$ .  $\text{dist}(\Phi(\mathbf{d}_i), \Phi(\mathbf{y}_j))$  denotes the Euclidean distance (ED) between the atom  $\mathbf{d}_i$  and the sample  $\mathbf{y}_j$  in the ensemble kernel space, which adaptively controls the contribution of each atom  $\mathbf{d}_i$  in representing  $\mathbf{y}_j$ . A larger dist indicates a greater dissimilarity between  $\mathbf{d}_i$  and  $\mathbf{y}_j$  in the ensemble kernel space, it is less possible to yield a large CR coefficient. Hence, the CR coefficients obtained by MKACRC tend to integrate the locality structure information into the multiple kernel CR-based methods. Compared with MKCRC, MKACRC can obtain more discriminative CR coefficients to represent the test pixel more robustly in the ensemble kernel space.



### B. Optimization

Generally, the objective function in Eq. (10) is difficult to optimize directly, a feasible solution is to alternatively solve the coefficient matrix  $X$  and the kernel weight vector  $\beta$  with an iterative two-step optimization strategy. At each iteration, one of  $X$  and  $\beta$  is fixed and the other is solved, then the roles of  $X$  and  $\beta$  are transformed. Once the convergence criterion (or a maximum number of iteration) reaches, the iteration procedure stops.

1) *On Optimizing  $X$* : When update  $X$ ,  $\beta$  can be fixed, the optimization in (10) is reduced to an adaptive CR-based coding problem

$$\hat{X} = \arg \min_X \left\| \sum_{m=1}^M \beta_m K_m(\mathcal{D}, Y) - \sum_{m=1}^M \beta_m G_m X \right\|_F^2 + \lambda \|\Gamma \odot X\|_F^2 \quad (11)$$

Since  $X$  denotes the combination of each CR coefficient vector  $\alpha_i (1 \leq i \leq P)$ , we can separately optimize  $\alpha_i$  and combine them into  $X$ . For solving  $\alpha_i$ , the optimization problem in (11) can be simplified as an equivalent optimization

$$\hat{\alpha}_i = \arg \min_{\alpha_i} \left\| \sum_{m=1}^M \beta_m K_m(\mathcal{D}, y_i) - \sum_{m=1}^M \beta_m G_m \alpha_i \right\|_2^2 + \lambda \|\Gamma_i \alpha_i\|_2^2 \quad (12)$$

where  $\Gamma_i = \text{diag}(\text{dist}(\Phi(d_1), \Phi(y_i)), \dots, \text{dist}(\Phi(d_N), \Phi(y_i)))$  denotes the dissimilar-weighted regularization matrix, which can adaptively assign large coefficients to the atoms that are close to the training pixel  $y_j$  and diminish the adverse effect of dissimilar pixels. Similar to (5), the coefficient vector  $\alpha_i$  in the ensemble kernel space can be computed in a closed form

$$\hat{\alpha}_i = (G^T G + \lambda \Gamma_i^T \Gamma_i)^{-1} G^T K(\mathcal{D}, y_i) \quad (13)$$

where  $K(\mathcal{D}, y_i) = \sum_{m=1}^M \beta_m K_m(\mathcal{D}, y_i)$ .

2) *On Optimizing  $\beta$* : When update  $\beta$ ,  $X$  can be fixed. The optimization in (10) is reduced to

$$\begin{aligned} \hat{\beta} = \arg \min_{\beta} & \left\| \sum_{m=1}^M \beta_m K_m(\mathcal{D}, Y) - \sum_{m=1}^M \beta_m G_m X \right\|_F^2 \\ \text{s.t. } & \beta_m \geq 0, \quad \text{and} \quad \sum_{m=1}^M \beta_m = 1. \end{aligned} \quad (14)$$

Since  $K_m(\mathcal{D}, Y)$  and  $G_m$  are known as a prior, one can solve the optimization in (14) via a standard constrained quadratic program. Hence, we further reformulate (14) to a concise form

$$\begin{aligned} \hat{\beta} = \arg \min_{\beta} & \left\| \sum_{m=1}^M \beta_m e_m \right\|_F^2 \\ \text{s.t. } & \beta_m \geq 0, \quad \text{and} \quad \sum_{m=1}^M \beta_m = 1 \end{aligned} \quad (15)$$

where  $e_m = K_m(\mathcal{D}, Y) - G_m X$ . Then, we stack each column of  $e_m$  into one column, turning it to a column vector  $\tilde{e}_m$ .

Defining  $\tilde{E} = \{\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_M\}$ , we can reformulate (15) to a constrained quadratic program problem

$$\hat{\beta} = \arg \min_{\beta} \|\tilde{E} \beta\|_2^2 \quad \text{s.t. } \beta_m \geq 0, \quad \text{and} \quad C \cdot \beta = 1 \quad (16)$$

where  $C = [1, 1, \dots, 1] \in \mathbb{R}^{1 \times M}$ . The optimization problem in (16) can be efficiently solved by various standard convex optimization solvers. The training procedure of MKACRC for learning the kernel weights for multiple kernels is given in Algorithm 1.

---

#### Algorithm 1 Training Procedure of MKACRC

---

- 1: **Input**: Dictionary set  $\mathcal{D} = \{\mathcal{D}_c\}_{c=1}^C \in \mathbb{R}^{B \times N}$ , training set  $Y = \{y_1, y_2, \dots, y_P\} \in \mathbb{R}^{B \times P}$ , parameters  $\lambda$ , maximum iteration count  $T$ .
  - 2: **Initialization**:
    - 1) Selecting a suitable kernel combination: Naive, multi-metric, or multiscale form;
    - 2) Initialize the kernel weight vector  $\beta$  and compute  $\{K(\mathcal{D}, y_i)\}_{i=1}^M$  and multiple kernel Gram matrix  $G_m$  in the ensemble kernel space.
  - 3:  $t \leftarrow 0$
  - 4: **while**  $t \leq T$  **do**
  - 5:   **Collaborative coding stage**:
  - 6:   **for** each training pixel  $y_i$  **do**
  - 7:     1) Compute the locality adaptor  $\Gamma_i$  with entries  $\text{dist}(\Phi(d_i), \Phi(y_j))$  for  $i = 1, 2, \dots, N$ ;
  - 8:     2) Compute the coding coefficients  $\alpha_i$  of training pixel  $y_i$  via an adaptive CR model in the ensemble kernel space according to (12);
  - 9:   **end for**
  - 10:   Obtain the representation  $X = \{\alpha_1, \alpha_2, \dots, \alpha_P\}$
  - 11:   **Kernel weights updating stage**:
  - 12:   Solve the kernel weight vector  $\beta$  with fixed coding coefficient matrix  $X$ ;
  - 13:   Transform the problem in (14) into a constrained quadratic program;
  - 14:    $t \leftarrow t + 1$
  - 15: **end while**
  - 16: **Output**: Kernel weight vector  $\beta$ .
- 

### C. Classification Procedure for Test Pixels

In order to determine the class label of an unknown pixel  $y$ , we calculate the CR coefficient vector  $\alpha$  over  $\mathcal{D}$  in the ensemble kernel space via adaptive CR model. After accomplishing the training procedure of the proposed multiple kernel CR classification framework, both the ensemble kernel  $K(\mathcal{D}, y)$  and the multikernel Gram matrix  $G$  can be computed based on the learned kernel weights. In the multiple kernel-induced space, the coefficient vector  $\alpha$  of  $y$  can be solved by a dissimilar-weighted  $\ell_2$ -norm regularization

$$\hat{\alpha} = \arg \min_{\alpha} \left\| \sum_{m=1}^M \beta_m K_m(\mathcal{D}, y) - \sum_{m=1}^M \beta_m G_m \alpha \right\|_2^2 + \lambda \|\Gamma \alpha\|_2^2 \quad (17)$$

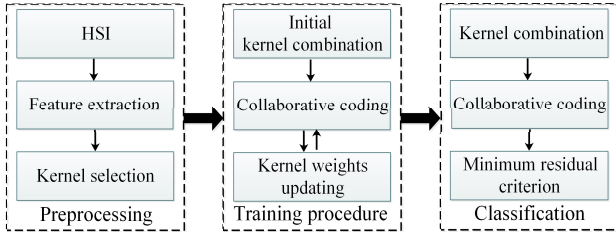


Fig. 2. Flowchart of the proposed multiple kernel CR-based classification framework.

where  $\Gamma = \text{diag}(\text{dist}(\Phi(d_1), \Phi(y)), \dots, \text{dist}(\Phi(d_N), \Phi(y)))$  denotes the dissimilar-weighted regularization matrix. Similar to (5), the coefficient vector  $\alpha$  can be computed in a closed form

$$\hat{\alpha} = (G^T G + \lambda \Gamma^T \Gamma)^{-1} G^T K(\mathcal{D}, y). \quad (18)$$

Finally, the class label of the test pixel  $y$  can be assigned by the minimum residual between  $y$  and its  $C$  approximations in the ensemble kernel space

$$\text{class}(y) = \arg \min_{c=1, \dots, C} \|K(\mathcal{D}, y) - G_c \delta_c(\alpha)\|_2^2 \quad (19)$$

where  $G_c$  denotes the multiple kernel Gram matrix of the atoms in the  $c$ th class, and  $\delta_c$  denotes the indicator function [9] that extracts coefficients related to the elements of the  $c$ th class. The flowchart of the proposed multiple kernel CR-based classification framework is shown in Fig. 2.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

We investigate the effectiveness of the proposed multiple kernel CR-based classification framework, compared against the existing state-of-the-art representation-based classifiers on three public HSIs.

##### A. Data Set Description

In our experiments, three popular HSIs are used to evaluate the performance of the proposed multiple kernel CR-based methods.

- 1) *Indian Pines*: The first HSI was gathered by the AVIRIS sensor over an agricultural test site in Northwestern Indiana. The scene originally contains 220 spectral channels, and the image size is  $145 \times 145$  pixels. Twenty water absorption bands are removed, and the remaining 200 channels are used for experiments. This scene contains 16 land-cover classes originally, and nine classes are considered in our experiments (it is allowed for more training dictionary from a statistical viewpoint [4]). A total of 9345 labeled pixels are contained in the ground truth map. Fig. 3(a) presents the false color map and the ground truth map of the Indian Pines image.
- 2) *Pavia University*: The second HSI was acquired by the ROSIS-03 optical sensor [54] over the campus in the University of Pavia, Italy. This scene originally contains 115 spectral channels with an image size of  $610 \times 340$  pixels. Twelve noisy and water absorption channels are discarded, the remaining 103 spectral channels are retained. There are 43 923 labeled pixels in total and

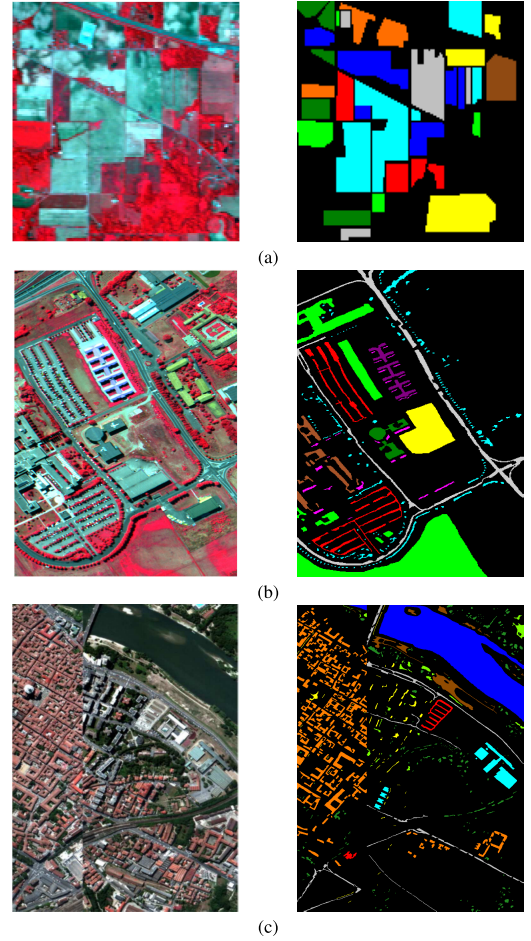


Fig. 3. False color images and ground truth maps for the three HSIs. (a) Indian Pines. (b) Pavia University. (c) Pavia Center.

nine ground-truth classes. Fig. 3(b) shows the three-band false color map and the reference map of this image.

- 3) *Pavia Center*: The third HSI was also acquired by the ROSIS-03 optical sensor in the city center of Pavia, Italy. The scene contains 102 spectral channels (with a range from  $0.43$  to  $0.86 \mu\text{m}$ ) and the image size is  $1096 \times 715$  pixels with a spatial resolution of  $1.3$  m per pixel. The ground reference map contains 148 152 labeled pixels and nine land-cover classes.

More details about these three HSIs are listed in Tables I and II.

##### B. Experimental Setting

In the proposed multiple kernel CR-based classification framework, EMPs [55] are used to extract spectral spatial information, which can provide an extra discrimination capability for remote sensing classification [33], [43]. Individually, PCA is performed on the original HSI, and a series of morphological openings and closings with structuring elements of different sizes is carried out on the extracted PCs to obtain the EMP features. The parameters adopted in our experiments for EMPs feature on three HSIs are listed in Table III.

TABLE I  
TOTAL NUMBERS OF LABELED PIXELS FOR THE THREE HISs

Indian Pines			University of Pavia			Pavia Center		
Class	Name	Number	Class	Name	Number	Class	Name	Number
C1	Corn-notill	1434	C1	Alfalfa	6852	C1	Water	65971
C2	Corn-mintill	834	C2	Meadows	18686	C2	Tree	7598
C3	Grass-pasture	497	C3	Gravel	2207	C3	Meadow	3090
C4	Grass-trees	747	C4	Trees	3436	C4	Brick	2685
C5	Hay-windrowed	489	C5	Metal sheet	1378	C5	Soil	6584
C6	Sybean-notill	968	C6	Bare soil	5104	C6	Asphalt	9248
C7	Sybean-mintill	2468	C7	Bitumen	1356	C7	Bitumen	7287
C8	Sybean-clean	614	C8	Brick	3878	C8	Tile	42826
C9	Woods	1294	C9	Shadows	1026	C9	Shadows	1265
Total		9345	Total		43923	Total		148152

TABLE II  
MAIN CHARACTERISTICS OF THE THREE HISs

HSI	Indian Pines	Pavia University	Pavia Center
Sensor	AVIRIS	ROSIS-03	ROSIS-03
Spectral coverage	0.4 $\mu$ m-2.5 $\mu$ m	0.43 $\mu$ m-0.86 $\mu$ m	0.43 $\mu$ m-0.86 $\mu$ m
Image size	145 $\times$ 145	610 $\times$ 340	1096 $\times$ 715
Spatial resolution	20m	1.3m	1.3m
No. of classes	9	9	9
No. of bands	200	103	102
Labeled pixels	9345	43923	148152

TABLE III  
PARAMETERS OF EMPs FEATURE FOR THE THREE HISs

Datasets	Parameters	No. of Dimension
Indian pines	Base image: PC1, PC2, PC3	39
	No. of openings and closings:6	
	Increasing step:2	
Pavia University	Base image: PC1, PC2, PC3	39
	No. of openings and closings:6	
	Increasing step:2	
Pavia Center	Base image: PC1, PC2, PC3	27
	No. of openings and closings:4	
	Increasing step:2	

After obtaining EMPs, multiple discriminative basic kernels associated with EMPs feature are constructed. Three kinds of kernel patterns are adopted in our experiments. Here, we briefly present the adopted kernel construction patterns for the proposed multiple kernel CR-based classification framework.

- 1) *Naive Combination*: Three widely used kernels, i.e., linear kernel ( $K(\mathbf{d}_i, \mathbf{d}_j) = \mathbf{d}_i^T \mathbf{d}_j$ ), polynomial kernel ( $K(\mathbf{d}_i, \mathbf{d}_j) = (\mathbf{d}_i^T \mathbf{d}_j + 1)^a$ ), and Gaussian RBF kernel ( $K(\mathbf{d}_i, \mathbf{d}_j) = \exp(-\text{dist}(\mathbf{d}_i - \mathbf{d}_j)/2\sigma^2)$ ), are selected as the predefined basic kernels. The corresponding variants for the proposed methods denote as a Naive and a Naive MKACRC.
- 2) *Multimetric Combination*: The Gaussian RBF kernel ( $K(\mathbf{d}_i, \mathbf{d}_j) = \exp(-\text{dist}(\mathbf{d}_i - \mathbf{d}_j)/2\sigma^2)$ ) with three well-performing discriminative kernel distance metrics ( $\text{dist}(\mathbf{d}_i - \mathbf{d}_j)$ ), including ED, Mahalanobis distance (MD), and spectral angle distance (SAD), is adopted in Multimetric MKCRC and Multimetric MKACRC.
- 3) *Multiscale Combination*: The Gaussian RBF kernel ( $K(\mathbf{d}_i, \mathbf{d}_j) = \exp(-\text{dist}(\mathbf{d}_i - \mathbf{d}_j)/2\sigma^2)$ ) with different scales  $\sigma$  correspond to different estimates of the

pixel sample similarity. Integrating the basic kernels of different scales can improve the classification ability of single-scale-based kernel methods. The proposed multiple kernel methods by combining different kernel scales are denoted as Multiscale MKCRC and Multiscale MKACRC.

To validate the performance of the proposed method, we compare it with several state-of-the-art representation-based methods, including EMPs-based CRC (EMPs-CRC) [10] and its kernel version (EMPs-KCRC) [56] and EMPs-based NRS [22] (EMPs-NRS) and its kernel version (EMPs-KNRS). The two joint representation-based methods (JSRC [11] and JCRC [57]) and its kernel version (KJSRC [25] and KJCRC) are also included. Note that EMPs-CRC, EMPs-KCRC, EMPs-NRS, and EMPs-KNRS are pixelwise classifiers based on the spectral-spatial feature, and JSRC, KJSRC, JCRC, and KJCRC denote spectral-spatial methods.

For the proposed multiple kernel CR-based methods, the regularization parameter  $\lambda$  is learned by  $k$ -fold cross validation in the range of  $[10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}]$ , and  $k$  is set to 10. For the polynomial kernel,  $a$  is set as 2 for the three HSIs. For the Naive and Multimetric variants, the parameter  $\sigma$  of Gaussian RBF kernel is set by the median value of  $1/(\|\mathbf{d}_i - \hat{\mathbf{d}}\|_2^2)$ ,  $i = 1, 2, \dots, N$ , where  $\hat{\mathbf{d}} = (1/N) \sum_{i=1}^N \mathbf{d}_i$  is the mean of all available dictionary atoms [52]. For Multiscale variants, the parameter  $\sigma$  is set in the range  $[0.2 \ 0.5 \ 0.8 \ 1.1 \ 1.4 \ 1.7 \ 2.0]$ . For Multimetric variants, the Gaussian RBF kernel with multiple distance metrics is chosen to construct the predefined kernels, the parameter  $\sigma$  for each metric-based kernel is set as  $1/(\|\mathbf{d}_i - \hat{\mathbf{d}}\|_2^2)$ ,  $i = 1, 2, \dots, N$ . For spectral-spatial methods, the optimal window size is selected differently for varied HSIs ( $7 \times 7$  for Indian Pines,  $15 \times 15$  for Pavia University, and  $5 \times 5$  for Pavia Center) to achieve the best classification performance.

For quantitative assessment, three widely used metrics (i.e., individual class accuracy (CA), overall accuracy (OA), and kappa coefficient ( $\kappa$ )) are applied to the proposed methods by calculating the averaged results with standard deviations over ten independent Monte Carlo (MC) runs.

### C. Effectiveness of MKCRC and MKACRC

1) *Comparisons of Different Classifiers*: The first experiment is implemented on three HSIs to verify the effectiveness



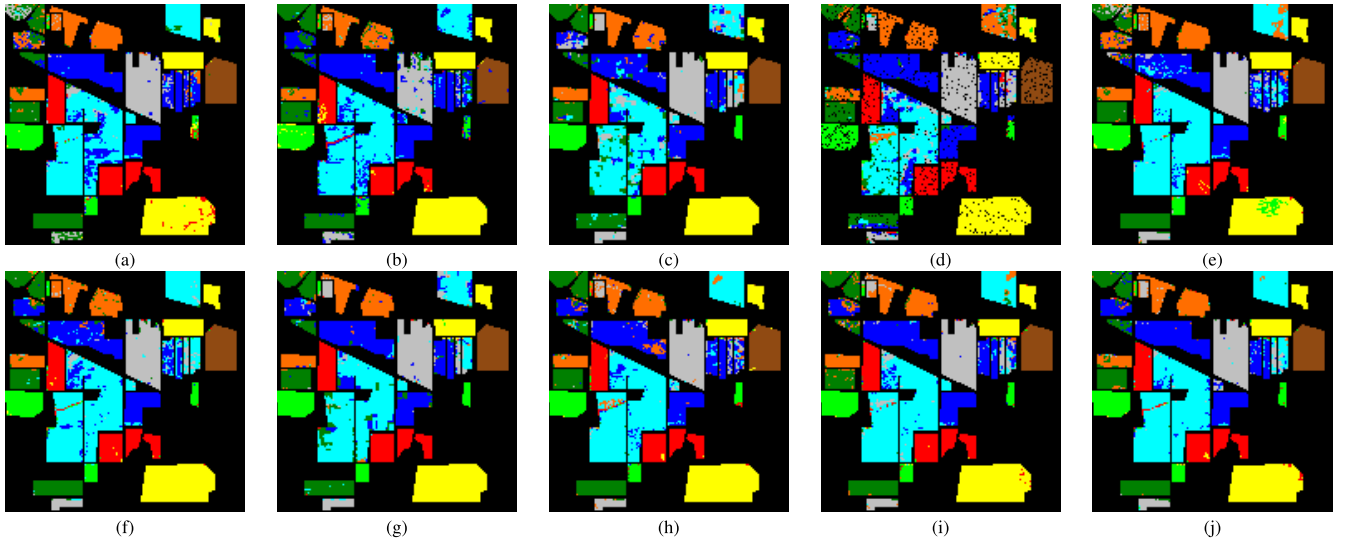


Fig. 4. Classification maps obtained by different representation-based classifiers on the Indian Pine image. (a) EMPs-CRC [10]. (b) EMPs-NRS [22]. (c) JSRC [11]. (d) JCRC [57]. (e) EMPs-KCRC [56]. (f) EMPs-KNRS. (g) KJSRC [25]. (h) KJCRC. (i) Multimetric MKCRC. (j) Multimetric MKACRC.

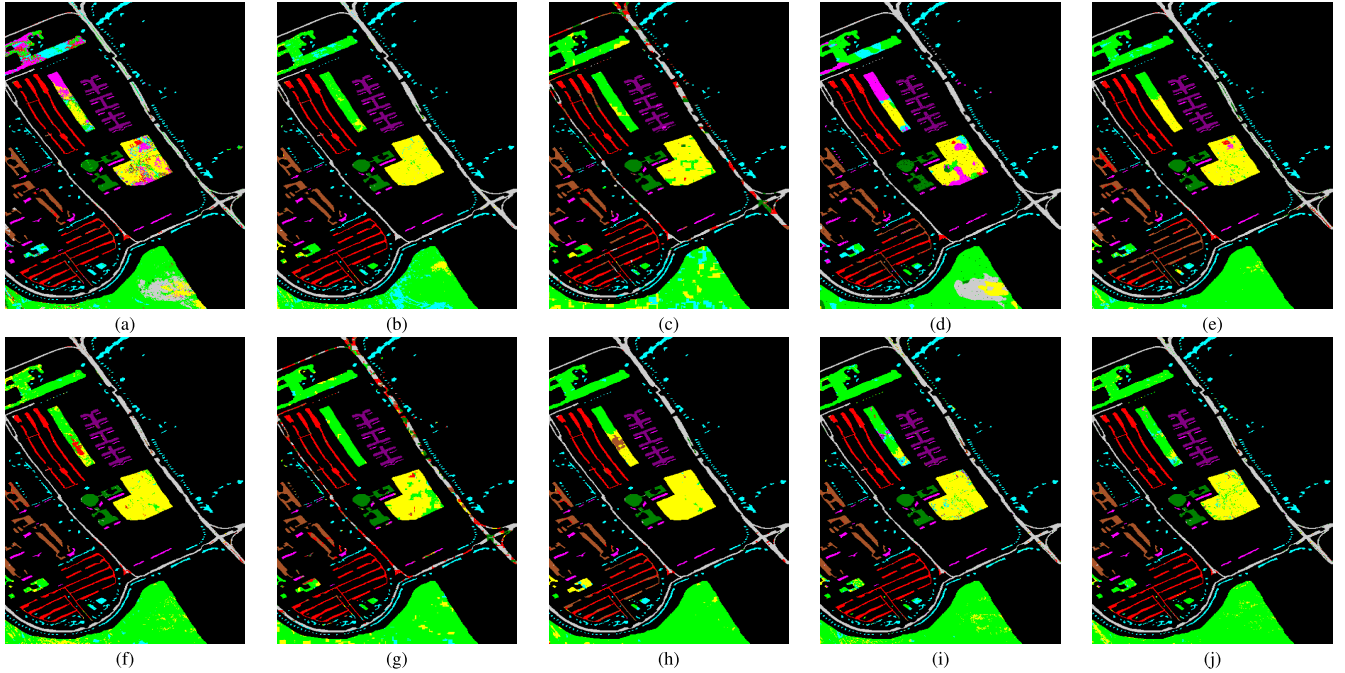


Fig. 5. Classification maps obtained by different representation-based classifiers on the Pavia University image. (a) EMPs-CRC [10]. (b) EMPs-NRS [22]. (c) JSRC [11]. (d) JCRC [57]. (e) EMPs-KCRC [56]. (f) EMPs-KNRS. (g) KJSRC [25]. (h) KJCRC. (i) Multimetric MKCRC. (j) Multimetric MKACRC.

of the proposed multiple kernel CR-based classification framework. In the test, we randomly choose 80, 60, and 60 labeled pixels per class as a training dictionary for Indian Pine, Pavia University, and Pavia Center, choose 120 labeled pixels per class as a training set, and the remaining labeled pixels are used for the test set. The multimetric pattern is adopted in this experiment for the proposed multiple kernel CR-based classification framework. The OAs and individual classification accuracies (in percentage) and the  $\kappa$  statistic with the standard deviation obtained by the proposed multiple kernel CRCs, and its state-of-the-art representation-based competitors on the Indian Pine, Pavia University, and Pavia Center are

listed in Tables IV–VI, respectively. The classification maps for various representation-based classifiers on the three HSIs are shown in Figs. 4–6, respectively.

In Tables IV–VI, it is apparent that the proposed multiple kernel CRCs, i.e., MKCRC and MKACRC, show powerful classification results among those state-of-the-art representation-based classifiers on three HSIs. In MKCRC and MKACRC, three distance metric-based (i.e., ED, MD, and SAD based) kernels are adopted to construct the unified multiple kernel CR-based classification framework, and the resulted methods yield a better performance than single kernel-based classifiers. Taking Pavia University image

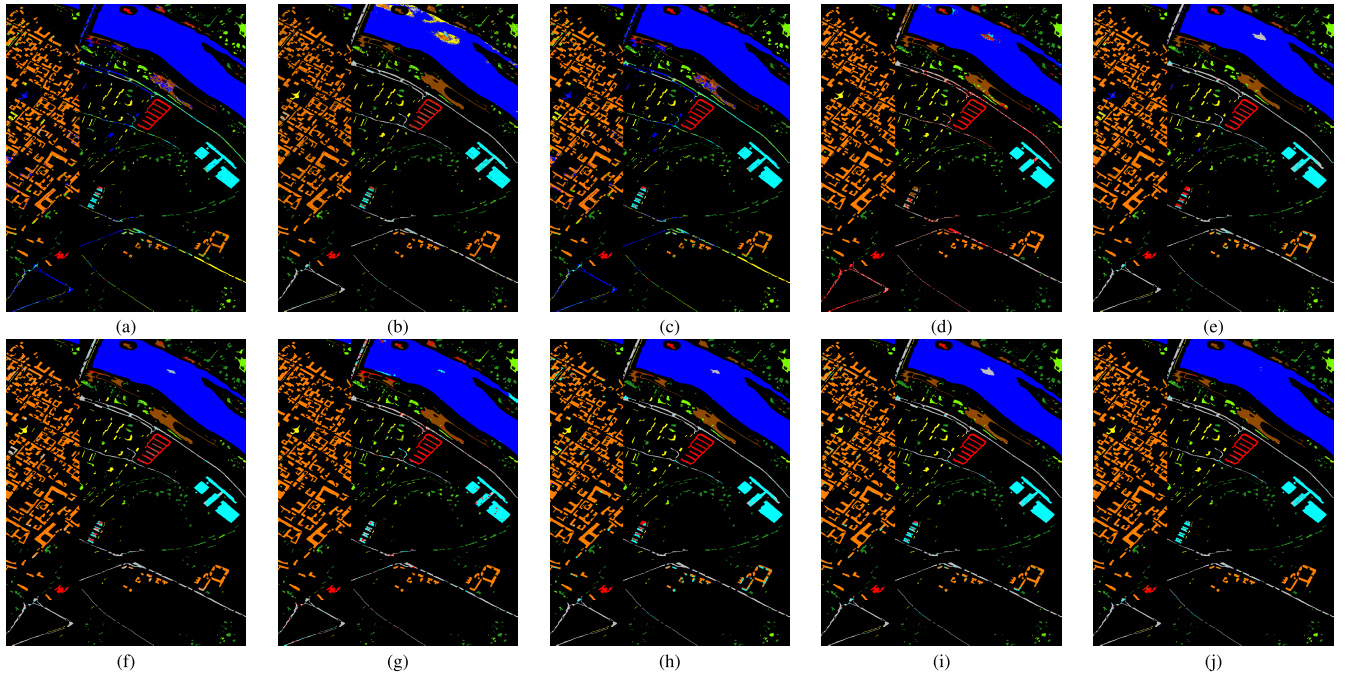


Fig. 6. Classification maps obtained by different representation-based classifiers on the Pavia Center image. (a) EMPs-CRC [10]. (b) EMPs-NRS [22]. (c) JSRC [11]. (d) JCRC [57]. (e) EMPs-KCRC [56]. (f) EMPs-KNRS. (g) KJSRC [25]. (h) KJCRC. (i) Multimetric MKCRC. (j) Multimetric MKACRC.

TABLE IV

CLASSIFICATION RESULTS (%) OF THE INDIAN PINES IMAGE OBTAINED BY EMPs-CRC [10], EMPs-NRS [22], JSRC [11], JCRC [57], EMPs-KCRC [56], EMPs-KNRS, KJSRC [25], KJCRC, MULTIMETRIC MKCRC, AND MULTIMETRIC MKACRC

Classifier	EMPs-CRC	EMPs-NRS	JSRC	JCRC	EMPs-KCRC	EMPs-KNRS	KJSRC	KJCRC	Multi-Metric MKCRC	Multi-Metric MKACRC
C1	73.42±3.78	87.52±3.26	75.69±6.41	87.17±5.44	61.93±3.32	83.42±1.32	81.76±3.19	85.79±3.39	86.95±2.14	89.30±1.77
C2	73.53±0.68	92.27±2.46	92.11±3.72	71.92±2.93	84.54±4.08	92.37±2.55	92.43±1.38	91.80±3.42	92.95±2.08	93.80±0.55
C3	77.98±1.82	75.76±4.33	98.99±4.16	98.65±0.18	89.56±2.49	93.06±1.66	99.10±0.58	99.33±0.85	94.28±1.18	94.21±0.16
C4	94.73±2.87	96.71±1.34	96.82±2.36	96.63±0.69	95.72±0.82	96.53±2.48	96.45±0.48	96.33±0.46	95.06±1.20	97.07±0.59
C5	99.31±0.24	96.54±1.20	100 ±0.00	100 ±0.00	99.58±0.29	99.10±0.63	100 ±0.00	100 ±0.00	99.31±0.60	100 ±0.00
C6	75.52±2.26	78.39±3.00	90.89±4.53	85.94±1.69	77.79±1.42	88.28±4.08	94.57±1.90	94.23±0.66	87.11±1.99	90.89±2.55
C7	76.82±1.45	78.84±6.16	77.34±4.53	55.11±1.55	87.34±1.88	87.07±3.35	81.10±2.28	79.50±2.69	91.14±0.94	92.99±1.56
C8	92.27±1.73	92.03±3.40	93.48±4.53	95.89±1.78	92.85±2.03	90.05±2.27	97.02±2.44	96.30±0.50	93.00±1.75	87.68±3.56
C9	93.99±0.83	99.54±0.43	99.27±4.53	99.82±1.06	90.55±0.88	99.52±0.21	99.63±0.14	99.48±0.23	99.36±0.14	98.54±0.28
OA	81.40±0.49	86.92±1.33	87.10±4.53	80.94±0.79	83.91±0.55	90.40±0.53	90.68±1.08	89.87±1.16	92.38±0.49	93.02±0.54
$\kappa$	0.779±0.006	0.855±0.016	0.847±0.045	0.778±0.009	0.809±0.007	0.885±0.006	0.889±0.013	0.880±0.013	0.909±0.006	0.918±0.021

TABLE V

CLASSIFICATION RESULTS (%) OF THE PAVIA UNIVERSITY IMAGE OBTAINED BY EMPs-CRC [10], EMPs-NRS [22], JSRC [11], JCRC [57], EMPs-KCRC [56], EMPs-KNRS, KJSRC [25], KJCRC, MULTIMETRIC MKCRC, AND MULTIMETRIC MKACRC

Classifier	EMPs-CRC	EMPs-NRS	JSRC	JCRC	EMPs-KCRC	EMPs-KNRS	KJSRC	KJCRC	Multi-Metric MKCRC	Multi-Metric MKACRC
C1	87.83±1.54	93.88±2.80	61.15±2.64	98.62±1.61	81.44±2.56	95.10±1.07	54.26±3.42	97.05±1.49	96.03±0.80	95.88±1.01
C2	58.32±1.89	76.22±3.65	79.50±4.97	64.35±3.42	86.59±1.09	87.58±2.16	90.82±2.35	82.26±3.64	89.44±3.24	93.23±1.49
C3	96.25±2.31	98.07±1.24	90.78±2.79	95.94±3.42	87.91±3.22	97.34±0.96	94.69±1.39	94.74±2.09	98.11±0.63	98.59±0.18
C4	98.23±1.25	99.24±0.61	81.73±1.95	91.99±3.42	98.30±0.86	97.36±1.14	83.33±2.50	99.76±0.71	97.24±2.21	96.67±0.34
C5	99.83±0.04	99.74±0.18	95.94±1.73	100 ±0.00	99.91±0.06	99.31±0.31	96.34±1.10	100 ±0.00	99.46±0.35	100±0.00
C6	64.59±4.10	94.53±2.28	89.71±2.26	70.24±3.42	91.96±1.67	87.76±2.68	91.36±2.66	97.98±3.35	88.92±1.23	88.70±2.74
C7	97.39±0.42	99.04±0.58	98.26±0.46	100 ±0.00	98.61±2.16	98.62±1.53	97.56±1.71	97.17±0.66	98.46±1.08	98.43±1.06
C8	91.78±5.26	96.40±1.53	83.89±1.03	80.53±3.42	83.58±2.88	93.80±3.09	91.08±4.36	77.01±5.74	96.69±0.83	97.37±0.50
C9	94.39±3.38	99.67 ±0.02	73.14±4.65	98.70±3.42	99.74±0.21	99.66 ±0.21	70.57±6.26	99.87±0.13	99.87±0.13	99.88±0.30
OA	74.04±1.31	87.33±1.65	79.75±2.38	77.85±3.42	88.06±1.36	91.28±0.76	84.82±1.30	91.76±1.63	92.71±1.51	94.40±0.61
$\kappa$	0.679±0.004	0.838±0.019	0.740±0.028	0.723±0.034	0.843±0.016	0.879±0.007	0.802±0.016	0.892±0.021	0.904±0.019	0.925±0.008

as an example, the OAs of MKCRC and MKACRC are 92.71% and 94.40% for the giving training dictionary and test pixels, which are higher than those of EMPs-KCRC, EMPs-KNRS, KJSRC, and KJCRC (88.06%, 91.28%, 84.82%, and 91.76%, respectively). Also, the single kernel-based classifiers perform better than its corresponding linear representation-based classifiers, i.e., EMPs-CRC, EMPs-NRS, JSRC, and JCRC. Moreover, by integrating the spectral-spatial

TABLE VI

CLASSIFICATION RESULTS (%) OF THE PAVIA CENTER IMAGE OBTAINED BY EMPs-CRC [10], EMPs-NRS [22], JSRC [11], JCRC [57], EMPs-KCRC [56], EMPs-KNRS, KJSRC [25], KJCRC, MULTIMETRIC MKCRC, AND MULTIMETRIC MKACRC

Classifier	EMPs-CRC	EMPs-NRS	JSRC	JCRC	EMPs-KCRC	EMPs-KNRS	KJSRC	KJCRC	Multi-Metric MKCRC	Multi-Metric MKACRC
C1	97.36±0.78	92.82±1.36	98.97±0.57	98.23±0.96	99.14±0.49	99.15±0.74	98.95±0.33	99.64±0.26	99.05±0.61	99.98±0.01
C2	70.02±3.87	80.21±4.04	83.07±2.12	86.53±1.99	77.22±2.51	86.27±2.62	75.82±5.72	91.14±1.62	91.68±1.76	88.58±4.32
C3	68.56±2.61	73.46±8.94	95.17±1.89	93.73±3.76	92.06±0.81	86.78±3.97	91.52±3.49	92.47±1.27	91.62±2.31	93.75±1.73
C4	94.97±1.31	79.03±5.26	93.70±2.36	88.43±2.22	96.93±1.11	88.46±2.82	87.98±0.94	99.62±0.17	99.20±0.26	94.41±1.61
C5	45.41±6.91	96.44±1.80	89.00±1.28	86.30±3.12	92.35±2.16	91.15±2.95	82.93±1.09	92.54±2.77	94.33±2.16	97.52±0.79
C6	43.15±5.27	82.35±9.14	58.47±1.99	44.50±3.90	89.94±3.17	94.66±1.21	77.36±5.35	95.42±2.44	92.05±2.31	92.80±1.67
C7	93.84±1.62	90.40±4.08	88.87±2.68	87.80±1.30	80.08±2.97	87.57±1.71	85.75±0.84	87.32±4.33	91.96±1.52	94.34±2.21
C8	86.66±2.45	81.29±7.26	98.90±0.36	99.90±0.17	96.07±0.56	93.43±1.76	99.22±0.24	91.67±2.63	95.90±1.21	98.15±0.71
C9	67.80±5.79	97.25±1.33	71.60±1.25	72.80±0.69	76.63±2.71	99.76±0.32	61.25±5.78	99.52±0.37	100±0.00	100±0.00
OA	85.62±1.98	87.68±2.50	86.41±0.67	84.25±0.73	94.75±1.32	95.22±0.74	94.16±0.12	95.21±0.61	96.19±0.81	97.82±0.45
$\kappa$	0.778±0.017	0.831±0.033	0.847±0.007	0.823±0.008	0.926±0.011	0.933±0.010	0.917±0.002	0.952±0.016	0.956±0.006	0.969±0.008

TABLE VII

CLASSIFICATION RESULTS OF BOTH MKCRC AND MKACRC USING DIFFERENT KERNEL COMBINATION PATTERNS ON THREE HSIs

Variants			MKCRC			MKACRC		
			Naive MKCRC	Multi-Metric MKCRC	Multi-Scale MKCRC	Naive MKACRC	Multi-Metric MKACRC	Multi-Scale MKACRC
Indian Pines	40	OA	83.26±1.02	87.50±1.27	85.88±0.98	90.21±1.34	90.17±0.73	86.60±0.73
		$\kappa$	0.803±0.013	0.849±0.011	0.833±0.010	0.889±0.015	0.888±0.008	0.841±0.008
	60	OA	87.09±1.28	90.52±2.03	90.53±1.76	92.24±0.92	92.33±0.31	91.36±1.55
		$\kappa$	0.847±0.011	0.887±0.018	0.887±0.017	0.911±0.010	0.912±0.004	0.901±0.018
	80	OA	90.02±0.78	92.38±0.49	91.13±1.39	92.86±0.69	93.02±0.54	92.66±0.68
		$\kappa$	0.881±0.007	0.909±0.006	0.906±0.013	0.913±0.008	0.913±0.010	0.911±0.007
Pavia University	40	OA	91.17±1.59	91.40±1.05	91.36±1.76	91.41±2.46	92.59±1.06	92.21±0.87
		$\kappa$	0.893±0.014	0.887±0.008	0.889±0.016	0.890±0.021	0.903±0.013	0.898±0.011
	60	OA	92.19±1.51	92.71±1.74	92.28±1.50	94.11±1.15	94.40±0.61	93.23±0.75
		$\kappa$	0.912±0.013	0.904±0.015	0.911±0.012	0.927±0.009	0.925±0.008	0.911±0.005
	80	OA	94.81±1.27	94.67±1.21	94.02±0.79	95.06±1.01	96.59±1.79	94.67±1.83
		$\kappa$	0.931±0.011	0.929±0.010	0.934±0.006	0.934±0.009	0.954±0.016	0.929±0.015
Pavia Center	40	OA	95.05±1.36	95.78±1.43	96.37±1.87	95.65±1.02	96.34±2.43	96.75±1.71
		$\kappa$	0.948±0.017	0.945±0.023	0.949±0.018	0.939±0.013	0.948±0.023	0.954±0.018
	60	OA	95.42±1.52	96.19±0.81	95.59±1.39	96.43±1.54	97.82±0.45	96.04±1.87
		$\kappa$	0.947±0.014	0.956±0.006	0.952±0.015	0.949±0.017	0.969±0.008	0.944±0.018
	80	OA	96.22±1.36	96.97±2.57	96.15±1.41	95.83±1.36	97.97±2.43	97.02±1.62
		$\kappa$	0.947±0.012	0.961±0.024	0.959±0.013	0.955±0.011	0.971±0.023	0.958±0.015

information of HSI, the proposed multiple kernel CR-based methods based on EMPs feature outperform the spectral-spatial techniques (i.e., JSRC and JCRC) and its kernel variants (i.e., KJSRC and KJCRC). Besides, by incorporating locality constraint information into the typical CR model in the ensemble kernel space, the proposed adaptive version (i.e., MKACRC) yields a more satisfying OA results of 0.64% for Indian Pines, 1.69% for Pavia University, and 1.63% for Pavia Center than MKCRC. This is because the test pixel is more prone to establish relations with the atoms that are more similar to the test pixel via the adaptive strategy.

2) *Analysis of the Influence of Dictionary Size:* To investigate the influence of dictionary size on the proposed multiple kernel CR-based classification framework, we experiment by adopting a series of dictionary sizes and collecting the corresponding classification results (OA and  $\kappa$ ). The dictionary size is set to 40, 60, and 80, the number of training samples is set to 120, and the remaining labeled pixels are used for the test. All the kernel construction patterns, including Naive, Multi-metric, and Multiscale, are investigated in the experiment for the proposed MKL framework. The classification results with respect to different dictionary sizes on three HSIs are summarized in Table VII. As can be observed, small-sized dictionary leads to a low OA result for the proposed multiple kernel

CR-based methods, and the reason is that a small dictionary set contains insufficient information to represent some test pixels, i.e., the training dictionary is incomplete. As the dictionary size increases, the training dictionary gradually becomes complete and yields a better OA result. Note that Multi-metric variants (i.e., Multimetric MKCRC and Multimetric MKACRC) outperform its Naive and Multiscale variants with respect to different dictionary sizes in most cases. Besides, by incorporating spatial information via adaptive weights to control the CR coefficients, MKACRC yields an improved OA result compared with MKCRC under its different kernel patterns.

#### D. Effect of Ensemble Kernel for CR-Based Classifiers

To investigate the effectiveness of an ensemble kernel strategy for CRCs, we show the OA results of the proposed multiple kernel CRCs (including three kernel patterns) and its single kernel-based classifiers (i.e., baselines) with respect to different dictionary sizes on the Pavia Center image in Fig. 7. In this experiment, various numbers (from 5 to 50 pixels per class) of labeled pixels are randomly chosen as the training dictionary, 120 labeled pixels per class are randomly chosen as the training set, and 1000 labeled pixels per class are randomly chosen as the test set. For Naive variants, the CR



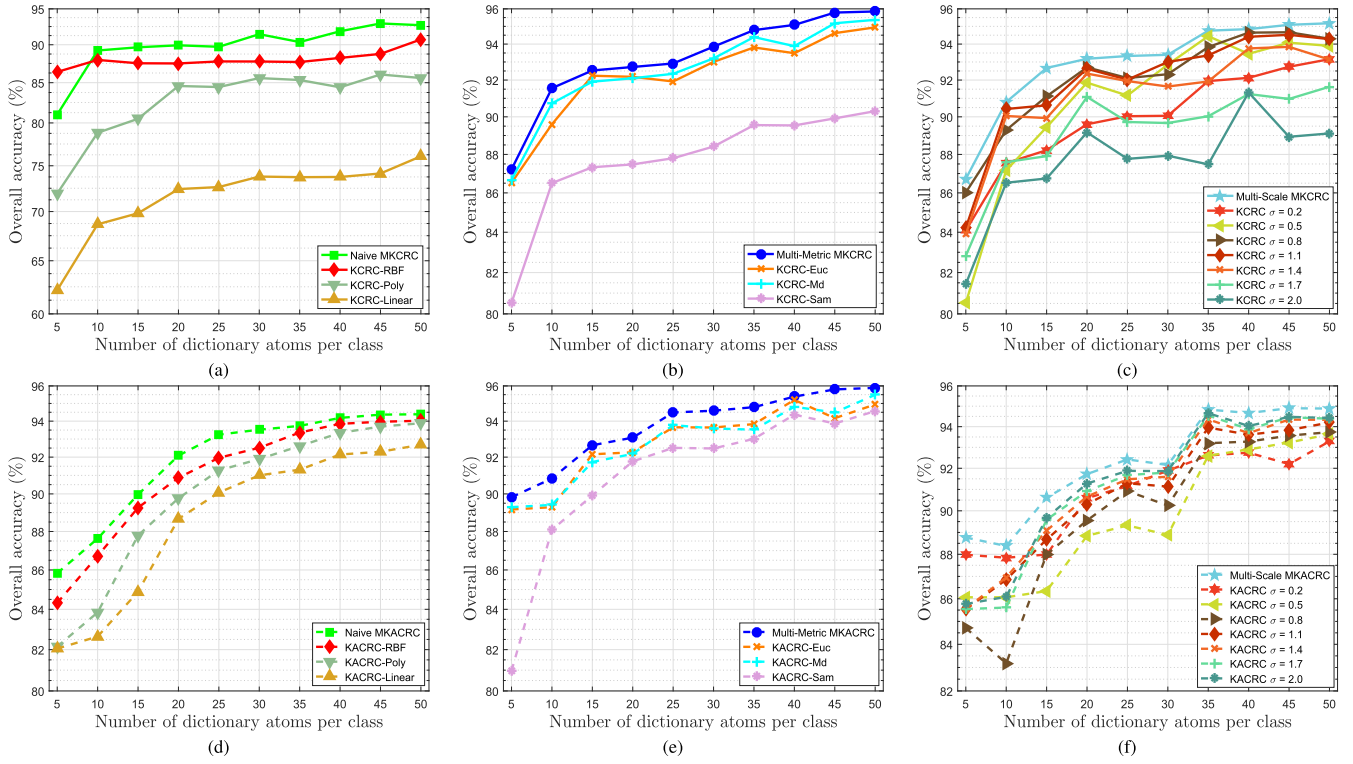


Fig. 7. OA results obtained by the proposed multiple kernel CRCs and its corresponding baselines on the Pavia Center image. (a) Naive MKCRC and baselines. (b) Multimetric MKCRC and baselines. (c) Multiscale MKCRC and baselines. (d) Naive MKACRC and baselines. (e) Multimetric MKACRC and baselines. (f) Multiscale MKACRC and baselines.

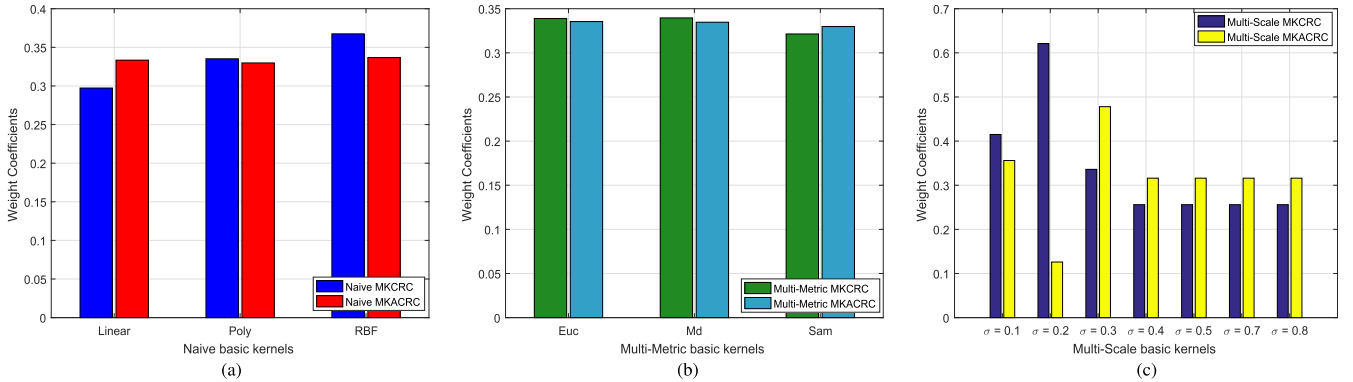


Fig. 8. Kernel weights for basic kernels obtained by the proposed multiple kernel CRCs on the Pavia Center image. (a) Naive variants. (b) Multimetric variants. (c) Multiscale variants.

and adaptive CR in RBF kernel space (i.e., KCRC-RBF and KACRC-RBF) generally achieve better OA results than the OA results in polynomial and linear kernel space, and the corresponding fused kernel methods (i.e., Naive MKCRC and Naive MKACRC) yield the best OA results by combining the three heterogeneous kernels (i.e., linear, poly, and RBF) together. For Multimetric variants, MD-kernel-based classifiers (i.e., KCRC-MD and KACRC-MD) achieve similar OA results to Ed-kernel-based classifiers along with different dictionary sizes, and yield better OA results than Sam-kernel-based classifiers. The fused multiple metric-based kernel methods (i.e., Multimetric MKCRC and Multimetric MKACRC) yield the best OA results by integrating multiple

heterogeneous space information into a unified MKL framework. Besides, Multiscale MKCRC and Multiscale MKACRC outperform each single-scale kernel methods by learning an optimal linear combination of multiple kernels with different scales (corresponding to varying estimates of sample similarity). Overall, the proposed multiple kernel CRCs consistently provide superior classification results over its single kernel-based classifiers concerning different dictionary sizes.

#### E. Analysis for the Weights of Different Kernels

Three kernel construction patterns are adapted for the proposed MKL framework, how to determine their kernel weights is another critical problem. To enhance the

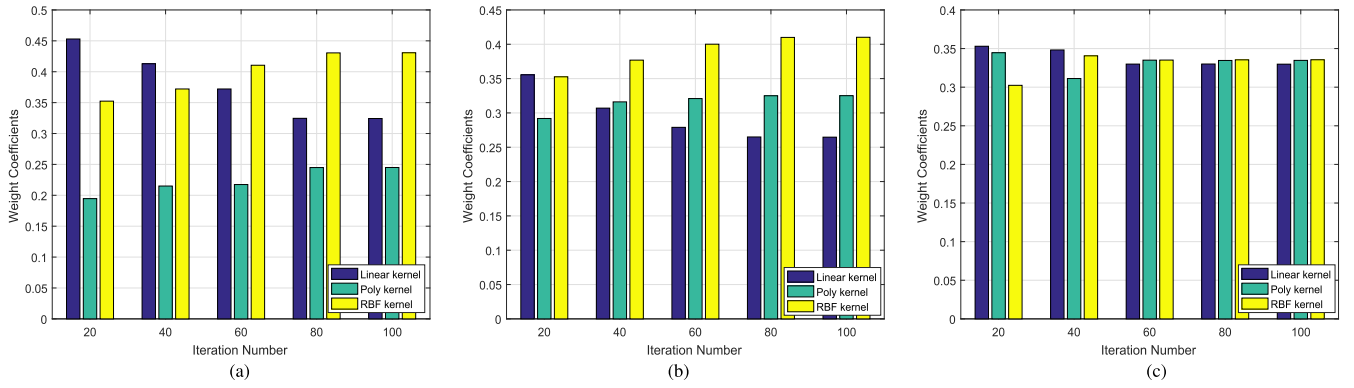


Fig. 9. Kernel weights of basic kernels obtained by Naive MKACRC along with different iteration numbers. (a) Indian Pines. (b) Pavia University. (c) Pavia Center.

TABLE VIII  
COMPUTATIONAL TIME (IN SECONDS) OBTAINED BY THE PROPOSED MULTIPLE KERNEL CRCs ALONG WITH DIFFERENT NUMBERS OF TRAINING DICTIONARY

Variants		MKCRC			MKACRC		
		Naive MKCRC	Multi-Metric MKCRC	Multi-Scale MKCRC	Naive MKACRC	Multi-Metric MKACRC	Multi-Scale MKACRC
Indian Pines	40	36.74±2.31	36.76±2.71	37.37±1.92	81.67±2.65	81.07±1.81	82.31±2.26
	60	36.64±1.87	36.79±2.57	55.88±3.02	81.80±2.61	80.59±2.39	208.21±4.35
	80	36.53±2.76	36.54±2.88	79.39±2.61	81.81±2.76	80.73±2.35	257.36±3.79
Pavia University	40	113.75±2.96	120.30±3.07	121.07±3.26	139.85±2.72	134.66±2.66	133.77±2.71
	60	322.86±4.17	290.72±3.22	311.86±2.99	430.20±3.69	420.03±3.81	526.32±4.28
	80	634.87±5.32	613.83±4.91	622.68±4.86	735.29±5.71	732.94±5.66	880.32±8.29
Pavia Center	40	328.42±3.17	322.95±2.64	449.99±3.72	357.82±3.86	353.12±3.07	557.49±4.57
	60	1052.89±10.07	1015.47±9.82	1121.02±8.87	1192.15±8.27	1435.08±13.24	1676.52±14.72
	80	2023.71±17.32	2004.36±18.09	2051.36±16.76	2189.32±18.61	2572.81±16.25	2851.26±15.38

reconstruction power of the kernel dictionary, we learn an ensemble kernel (i.e., an optimal linear combination of a series of basic kernels) from the training samples based on the minimum residual criterion via a two-step optimization strategy. Fig. 8 plots the learned kernel weights of each basic kernels for the proposed multiple kernel CRCs (including three kernel patterns) on three HSIs when 120 labeled pixels for each class are used for training. As can be observed, the kernel weights obtained by linear, polynomial, and RBF kernels are close to each other via the adaptive CR model, whereas the RBF kernel gets a more substantial weight than the other two basic kernels via the CR model. The kernel weights obtained by each single metric-based kernels (i.e., ED-, MD-, and Sam-based kernels) are slightly different from each other for the Multiscale variants. Besides, the basic kernel with  $\sigma = 0.2$  gets the most considerable weight in Multiscale MKCRC, and the basic kernel with  $\sigma = 0.3$  provides the key role in Multiscale MKACRC. In a word, the proposed multiple kernel CRCs can assign kernel weights for different basic kernels according to their roles in representing the training samples, and further boost the discrimination power of the traditional CRCs.

To investigate the influence of iterations on the kernel weights, we carry out experiments by collecting the kernel weights of the proposed methods using different iterations. Taking the Naive MKACRC method as examples, the kernel weights of three basic kernels (i.e., linear, poly, and RBF kernels) in terms of different iterations are presented in Fig. 9. In this experiment, the iterations are varied from 20 to 100 with

a step of 20. As can be observed, the kernel weights of the Naive MKACRC method constantly change to a stable value as the iteration number increases from 20 to 80. Then, the kernel weights of the three basic kernels nearly keep unchanged as the range from 80 to 100.

#### F. Computational Time

Finally, we evaluate the computational time of the proposed multiple kernel CR-based methods. All the algorithms are implemented using MATLAB R2016a environment on a 3.60 GHz Core i7-4790 CPU PC with 16 GB of RAM. The computational time for the proposed methods consists of two parts: an iterative training procedure (i.e., the learning of kernel weights) via a two-step iteration strategy and a testing procedure. The iteration alternation of collaborative coding and kernel weights update is time-consuming. Table VIII presents the computational time (in seconds) obtained by the proposed multiple kernel CR-based methods along with different numbers of training dictionary on three HSIs. As the number of training dictionary increases, the running time increases. By considering the adaptive CR mechanism, the MKACRC methods have longer running time than its corresponding MKCRC methods.

#### V. CONCLUSION

In this paper, we propose a novel multiple kernel CR-based classification framework by embedding the MKL model into CRCs, which aims at enhancing the representation ability of

training dictionary in the fused kernel space. To efficiently learn the optimal combination of multiple kernels, we add a training stage to the typical CRCs via a two-step optimization strategy, which can alternatively learn the representation coefficients and kernel weights by minimizing the representation error of training set. Three kernel patterns, including Naive, Multimetric, and Multiscale, are applied to the proposed MKL framework. Experimental results conducted on the three HSIs demonstrate the superiority of the proposed multiple kernel CRCs over several state-of-the-art classifiers and its single kernel-based classifiers regarding classification results. Besides, the adaptive CRCs outperform the CRCs in the ensemble kernel space by considering different contributions of the dictionary atoms in representing the test pixels. Moreover, to further enhance the performance of the proposed multiple kernel CRCs, dictionary learning methods can be applied to our MKL framework. Our future paper will focus on exploiting discriminative dictionary learning-based MKL methods for HSI to improve the classification performance.

#### ACKNOWLEDGMENT

The authors would like to thank the Associated Editor and the three Anonymous Reviewers for providing insightful comments and suggestions that significantly helped in improving the quality and presentation of this paper.

#### REFERENCES

- [1] S. Yang, Z. Shi, and W. Tang, "Robust hyperspectral image target detection using an inequality constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3389–3404, Jun. 2015.
- [2] 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.
- [3] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. ICCV*, 2011, pp. 471–478.
- [11] 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.
- [12] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sep. 2013.
- [13] J. Li, H. Zhang, and L. Zhang, "Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 94, pp. 25–36, Aug. 2014.
- [14] W. Li, Q. Du, and M. Xiong, "Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 48–52, Jan. 2015.
- [15] E. Zhang, L. Jiao, X. Zhang, H. Liu, and S. Wang, "Class-level joint sparse representation for multifeature-based hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4160–4177, Sep. 2016.
- [16] S. Jia, J. Hu, Y. Xie, L. Shen, X. Jia, and Q. Li, "Gabor cube selection based multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3174–3187, Jun. 2016.
- [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.*, vol. 11, no. 4, pp. 1118–1130, Apr. 2017.
- [18] 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.
- [19] Q. S. ul Haq, L. Tao, F. Sun, and S. Yang, "A fast and robust sparse approach for hyperspectral data classification using a few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2287–2302, Jun. 2012.
- [20] 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.
- [21] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2200–2208, Jun. 2014.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] W. Li and Q. Du, "Adaptive sparse representation for hyperspectral image classification," in *Proc. IGARSS*, 2015, pp. 4955–4958.
- [27] W. Li and Q. Du, "A survey on representation-based classification and detection in hyperspectral remote sensing imagery," *Pattern Recognit. Lett.*, vol. 83, no. 2, pp. 115–123, Nov. 2015.
- [28] 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.
- [29] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017.
- [30] W. Li and Q. Du, "Gabor-filtering-based nearest regularized subspace for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1012–1022, Apr. 2014.
- [31] 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.
- [32] B. Sun, X. Kang, S. Li, and J. A. Benediktsson, "Random-walker-based collaborative learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 212–222, Jan. 2017.
- [33] 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. 42, no. 3, pp. 480–491, Mar. 2005.
- [34] 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.
- [35] 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.



- [36] L. Gan, J. Xia, P. Du, and J. Chanussot, "Multiple feature kernel sparse representation classifier for hyperspectral imagery," *IEEE Geosci. Remote Sens.*, to be published.
- [37] Y. Gu, T. Liu, X. Jia, and J. A. Benediktsson, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.
- [38] M. Gönen and E. Alpaydin, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, Jul. 2011.
- [39] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," in *Proc. Int. Conf. Mach. Learn.*, 2004, pp. 41–48.
- [40] Y.-Y. Lin, T.-L. Liu, and C.-S. Fuh, "Multiple kernel learning for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1147–1160, Jun. 2011.
- [41] Y.-R. Yeh, T.-C. Lin, Y.-Y. Chung, and Y.-C. F. Wang, "A novel multiple kernel learning framework for heterogeneous feature fusion and variable selection," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 563–574, Jun. 2012.
- [42] A. Shrivastava, V. M. Patel, and R. Chellappa, "Multiple kernel learning for sparse representation-based classification," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3013–3024, Jul. 2014.
- [43] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.
- [44] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.
- [45] Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson, "A novel MKL model of integrating LiDAR data and MSI for urban area classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5312–5326, Oct. 2015.
- [46] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
- [47] Q. Wang, Y. Gu, and D. Tuia, "Discriminative multiple kernel learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3912–3927, Jul. 2016.
- [48] Y. Gu, Q. Wang, and B. Xie, "Multiple kernel sparse representation for airborne LiDAR data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 1085–1105, Feb. 2017.
- [49] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.
- [50] B. Demir and S. Erturk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4071–4084, Nov. 2010.
- [51] J. Yin, Z. Liu, Z. Jin, and W. Yang, "Kernel sparse representation based classification," *Neurocomputing*, vol. 77, no. 1, pp. 120–128, 2012.
- [52] L. Zhang *et al.*, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.
- [53] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2004.
- [54] P. Gamba, "A collection of data for urban area characterization," in *Proc. IGARSS*, Anchorage, AK, USA, Sep. 2004, pp. 69–72.
- [55] Z. Y. Lv, P. Zhang, J. A. Benediktsson, and W. Z. Shi, "Morphological profiles based on differently shaped structuring elements for classification of images with very high spatial resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4644–4652, Dec. 2014.
- [56] B. Wang, W. Li, N. Poh, and Q. Liao, "Kernel collaborative representation-based classifier for face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 2877–2881.
- [57] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang. (Apr. 2012). "Collaborative representation based classification for face recognition." [Online]. Available: <https://arxiv.org/abs/1204.2358>



**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 more than 70 articles in international peer-reviewed journals and more than 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.



**Le Gan** (M'18) 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 analysis.

Dr. Gan was a recipient of the National Scholarship for Doctoral Graduate Students granted by the Ministry of Education of the Peoples 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 a JSPS Post-Doctoral Oversea 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 was a recipient of the IEEE Geoscience and Remote Sensing Society Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee in 2017.



**Daming Wang** received the M.S. degree in geographic information system from the University of New South Wales, Sydney, NSW, Australia, in 2008, and the Ph.D. degree in geographic information system from Peking University, Beijing, China, in 2009.

He is currently the Division Director of China Geological Survey, Beijing, China, where he is the Assistant Director with Shenyang Geological Survey. He has authored or co-authored more than 20 papers in Chinese and international journals.

His research interests include remote sensing applications in oil and gas exploration and black soil investigation.

Dr. Wang serves as a peer reviewer of the *International Journal of Geographical Information Science and Sensors*.