Class-Oriented Weighted Kernel Sparse Representation With Region-Level Kernel for Hyperspectral Imagery Classification

Le Gan, Junshi Xia, *Member, IEEE*, Peijun Du, *Senior Member, IEEE*, and Jocelyn Chanussot, *Fellow, IEEE*

***Abstract*—As a nonlinear extension of traditional sparse representation-based classifier (SRC), kernel SRC (KSRC) has shown its excellent performance for hyperspectral image (HSI) classification, by mapping the nonlinearly separable samples into high-dimensional feature space. However, the rich locality structure of HSI contains more discriminative information, which should be considered in KSRC. We intend to incorporate the locality structure and kernel method into a unified SR-based framework by a local spatial kernel. As a powerful texture descriptor, local binary patterns (LBP) was used to extract local feature for remote sensing. Region-level kernels are applied to calculate the distance between two LBP histogram features. To discover nonlinear similarity information between test and training samples, we integrate the LBP feature into spatial region-level kernel for HSI classification. Then, we propose a weighted kernel sparse representation classifier optimized via class-oriented strategy, which combines local structure information and SRC in the kernel feature space based on spatial region-level kernel. Experimental results on three open HSIs demonstrate that the proposed method achieves better classification performance than other state-of-the-art classification methods.**

***Index Terms*—Classification, class-oriented strategy, hyperspectral image (HSI), kernel, local structure information, sparse representation.**

I. Introduction

HYPERSPECTRAL image (HSI), captured by hyperspectral sensors in hundreds of contiguous spectral bands, includes precious discriminative information about materials, and has provided numerous opportunity to improve the performance of materials classification [1]–[4], target identification, and anomaly detection. Among the numerous applications, HSI classification is a long-standing research topic where pixels are labeled to one of the several thematic classes. Various advanced classification methods, including support vector machines (SVMs) [5], [6], multinomial logistic regression [7], [8], neural network [9], random forest [10], and rotation forest [11] have been investigated in recent years.

Recently, sparse representation-based classifier (SRC), which assumes that a test sample can be compactly approximated by only a few atoms that carry the most important information in a structured dictionary, has attracted great concern to solve many computer vision tasks [12]–[15]. Different from previous classifiers which make decisions based on standard classifiers (e.g., *k*-nearest-neighbor [16] or SVMs [5]), the decision rule of SRC are based on the minimum reconstruction error of the test sample over a set of training samples or a dictionary learned from training samples. Likewise, SRC [17]–[24] have been applied to HSI classification, which relies on the fundamental assumption that samples belonging to the same class approximately lie in the same low-dimensional subspace. Chen *et al.* [18] proposed a joint sparse representation classifier based on the joint sparse model (JSM) with neighboring pixel information for HSI classification, and Zhang *et al.* [22] further extended the JSM with a nonlocal spatial prior.

Previous studies [6], [25] have shown the importance of incorporating the spatial neighborhood information into the classification. As a powerful spatial feature extraction method, local binary patterns (LBP) descriptor was proposed by Ojala *et al.* for texture classification [26], where a local circular neighborhood is binary thresholded by the gray of the center pixel. Various extensions were developed for the classical LBP descriptors [27]–[29]. As a feature extraction approach, LBP histogram features have been used to extract local region feature, and two fusion strategies are applied to the LBP-based feature with other two types of features to perform an HSI classification task [30]. The success of LBP feature is due to the discriminative ability and computational simplicity of the operator. LBP-based features with other classifiers have been proposed. For instance, Min *et al.* [31] combined LBP-based features with SRC to improve the robustness to various localized variations. The

Manuscript received May 17, 2017; revised July 31, 2017 and September 2, 2017; accepted September 22, 2017. Date of publication October 10, 2017; date of current version April 11, 2018. The work was supported in part by the Natural Science Foundation of China under Grant 41471275 and in part by the Key Scientific Instrument Development Foundation of China under Grant 2012YQ05025004. *(Corresponding auther: 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, 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: 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 University Grenoble Alpes, CNRS, Grenoble Institute of Engineering Univ. Grenoble Alpes, 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/JSTARS.2017.2757475

combination of LBP features and SRC strengthens the representation power of the test samples. However, LBP histogram features have only been integrated into the linear SRC algorithm, which is not able to discover nonlinear relationships of the LBP-based features. For better utilizing the advantages, region-level kernels were adopted to capture the nonlinear characteristics of two LBP histogram features [32]. Thanks to the region-level kernel, the SRC model can be extended as KSRC.

The typical SRC model does not consider the distance and similarity between the test sample and each dictionary atom. In general, training samples that are closer or more similar to the test sample play greater role in representing the test sample. To enhance the discriminative ability of original SRC model, Lu *et al.* [33] extended the original SRC model by integrating the locality structure information of samples in a unified formulation, and proposed weighted SRC (WSRC) model. In [34] and [35], WSRC adaptively exploits the similarity between the test sample and each dictionary atom in representing the test sample.

In HSI, neighboring pixels usually consist of similar material and their spectral signatures are highly correlated. Nevertheless, the original SRC only treats HSI as unordered pixels without considering its local characteristics. In this paper, region-level kernels based on LBP histogram feature are applied to SRC for each pixel over a local patch to improve the classification performance of HSI. However, as discussed in [33], dictionary atoms far away from the test pixel tend to have less significance in representing it than other dictionary atoms. Therefore, we integrate the locality adaptive structure information and region-level spatial kernel-based SRC into a unified framework, and propose a weighted kernel sparse representation classifier (WKSRC) for HSI classification. Previous work [36] has indicated that componentwise updating procedure is an effect optimization approach. Component- or class-oriented optimization strategy leads to monotonically decrease of the objective function and finds the global minimum without any risk of falling into the local minimum, which has been applied to dictionary update for discriminative sparse representation [37]. So the proposed framework is solved by class-specific (or class-dependent) weighted ℓ_1-minimization problem for more stable and effective representation coefficient. Finally, the test sample is assigned into the class label corresponding to the total minimum residual in reconstruction errors. The contributions of this paper are:

1) A region-level spatial kernel (HI kernel) based on LBP features is applied to the typical kernel SRC for HSI classification.
2) To improve the discriminative ability of spatial kernel (i.e., region-level kernel) based SRC, we integrate locality structure information calculated by Gaussian kernel distance into spatial KSRC and formulate as a weighted ℓ_1-minimization problem in region-level kernel space.
3) The weighted ℓ_1-minimization problem is solved via a class-oriented optimization strategy, which learns sparse coding based on class-specific subdictionary.

The remainder of this paper is organized as follows. In Section II, we briefly review the formulations of KSRC and WSRC. Section III presents the proposed WKSRC and class-oriented WKSRC methods. The effectiveness of the proposed algorithms is demonstrated in Section IV. Finally, Section V concludes this work.

II. SR-Based Classifier and Local Spatial Feature

A. Kernel Sparse Representation Classifier

Generally, the linear SRC model cannot cope with the nonlinear classification problem. To alleviate this problem, the original spectral feature should be mapped into a high dimensionality kernel space. Suppose the training dictionary $\mathcal{D}$ is given, the weighting coefficients $\boldsymbol{\alpha}$ in the kernel feature space can be obtained via KSRC [38], [39] by solving the optimization problem with a ℓ_1-norm regularization

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\phi(\boldsymbol{y}) - \Phi(\mathcal{D})\boldsymbol{\alpha}\|_2 + \lambda\|\boldsymbol{\alpha}\|_1 \tag{1}$$

where ϕ: $\mathcal{R}^m \mapsto \mathcal{R}^M$, $(m < M)$ defines a mapping function, which maps the samples (i.e., test sample) and training dictionary into high-dimensional space: $\boldsymbol{y} \mapsto \phi(\boldsymbol{y}) \in \mathcal{R}^{M\times 1}$, $\mathcal{D} = [\boldsymbol{d}_1, ..., \boldsymbol{d}_n] \mapsto \Phi(\mathcal{D}) = [\phi(\boldsymbol{d}_1), ..., \phi(\boldsymbol{d}_n)] \in \mathcal{R}^{M\times n}$. By introducing a kernel function $\mathcal{K}(\boldsymbol{a}, \boldsymbol{b}) = \langle\phi(\boldsymbol{a}), \phi(\boldsymbol{b})\rangle$, (1) can be modified as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{k}(\mathcal{D}, \boldsymbol{y}) - \boldsymbol{K}\boldsymbol{\alpha}\|_2 + \lambda\|\boldsymbol{\alpha}\|_1 \tag{2}$$

where $\boldsymbol{k}(\mathcal{D}, \boldsymbol{y}) = [\mathcal{K}(\boldsymbol{d}_1, \boldsymbol{y}), ..., \mathcal{K}(\boldsymbol{d}_n, \boldsymbol{y})]^T \in \mathcal{R}^{n\times 1}$, and $\boldsymbol{K} = \Phi^T\Phi \in \mathcal{R}^{n\times n}$ denotes the kernel Gram matrix with $\boldsymbol{K}_{i,j} = \mathcal{K}(\boldsymbol{d}_i, \boldsymbol{d}_j)$. Once obtaining the weighting coefficient $\boldsymbol{\alpha}$ in the kernel space, the class label of $\boldsymbol{y}$ is determined by the minimal total residual of each class

$$\begin{aligned} class(\boldsymbol{y}) &= \arg\min_{i=1,...,c} \|\phi(\boldsymbol{y}) - \Phi(\mathcal{D})\delta_i(\boldsymbol{\alpha})\|_2 \\ &= \arg\min_{i=1,...,c} \|\boldsymbol{k}(\mathcal{D}, \boldsymbol{y}) - \boldsymbol{K}\delta_i(\boldsymbol{\alpha})\|_2 \end{aligned} \tag{3}$$

where $\delta_i(\boldsymbol{\alpha})$ is the characteristic function that selects the coefficients $\boldsymbol{\alpha}$ associated with the ith class.

B. Weighted Sparse Representation Classifier

Considering the significance of different dictionary atoms of the HSI, local structure information of samples can be applied to represent the test pixel. The significance can be measured by the distance between each dictionary atom and the test pixel. Similarly, a distance-weighted CRC method, referred to nearest regularized subspace (NRS) [40], exploits the spectral similarity (Euclidean distance) to control the regularization term. To preserve the locality structure information, weighted SRC (WSRC) imposes locality constraint on the sparsity regularized reconstruction problem, which can be solved by the weighted ℓ_1-minimization problem

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} \|\boldsymbol{y} - \mathcal{D}\boldsymbol{\alpha}\|_2 + \lambda_1\|\boldsymbol{\Gamma}\boldsymbol{\alpha}\|_1 \tag{4}$$

$$\boldsymbol{\Gamma}_{\boldsymbol{y}} = \begin{bmatrix} \|\boldsymbol{y} - \boldsymbol{d}_1\|_2 & & 0 \\ & \ddots & \\ 0 & & \|\boldsymbol{y} - \boldsymbol{d}_n\|_2 \end{bmatrix} \tag{5}$$

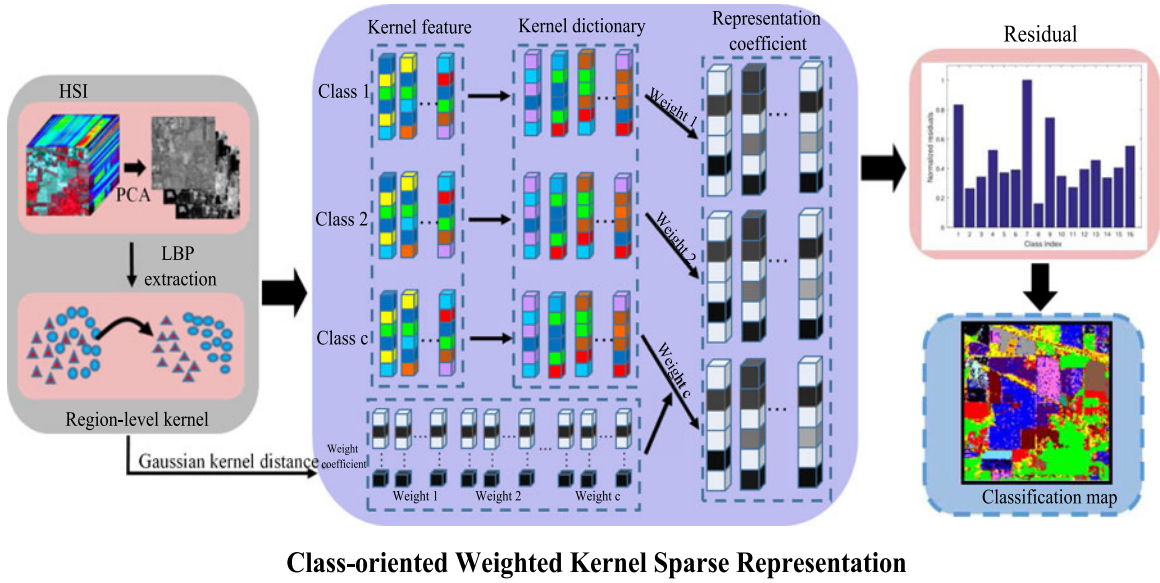

Fig. 1. Schematic illustration of the proposed CoWKSRC classification algorithm with region-level kernel based on LBP histogram feature.

where $\mathbf{\Gamma}_{\boldsymbol{y}}$ is a weighted ℓ_1-regularization term, which acts as the locality adapter that penalizes the distance between $\boldsymbol{y}$ and each atom $\boldsymbol{d}_i$, and $\text{dist}(\boldsymbol{y}, \boldsymbol{d}_i) = \|\boldsymbol{y} - \boldsymbol{d}_i\|_2^2$ denotes the Euclidean distance. Once the weighted coefficient vector $\boldsymbol{\alpha}$ is obtained, the class label of $\boldsymbol{y}$ is determined by the lowest representation error $\arg \min_{i=1,\ldots,c} \|\boldsymbol{y} - \mathcal{D}\delta_i(\boldsymbol{\alpha})\|_2$.

C. Local Binary Patterns

Given a center pixel (scalar value) in a local patch, the calculation of the LBP codes can be computed by comparing its value with its neighboring pixels in a single scan, which can be given by [41]

$$LBP_{u,r} = \sum_{u=0}^{u-1} s(g_p - g_c)2^p$$
$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{6}$$

where g_c is the gray value of the center pixel, g_p is the value of its neighboring pixels in a local circle, u is the number of neighboring pixels, and r is the radius of the neighborhood. The output of the LBP operator in (6) is an u-bit binary string (including 2^u distinct values) or a corresponding decimal label extracted from the local neighboring pixel comparisons, reflect texture orientation and smoothness in a local patch. Generally, the occurrences of LBP operator in a local patch of size $s \times s$ are collected into a histogram feature. The classification is then performed by computing histogram similarities.

III. Class-Oriented Weighted Kernel Sparse Representation Classifier

In this paper, we combine region-level spatial kernel and local structure information into a unified SR-based classification framework, and propose a weighted kernel sparse representation classifier (WKSRC), which is solved by class-specific weighted ℓ_1-minimization problem. Fig. 1 shows the framework of the proposed CoWKSRC. Given a HSI, region-level kernel feature-based uniform LBP histogram representation are applied to each pixel over a local patch. Based on local kernel function, we then construct the kernel subdictionary with regard to the training label class by class. Simultaneously, the class-specific weighting matrix is calculated via Gaussian kernel distance. In the representation stage, class specific weight coefficients are learned by class-oriented optimization strategy based on each kernel subdictionary. Finally, class label is determined according to the minimum total residual of individual class.

A. Region-Level Kernel

In most cases, LBP is used in the form of histogram features counted in a local window, and the similarity is preferred to calculate the distance between two LBP histogram features for better utilizing the advantages of LBP features. Selecting a suitable nonlinear kernel to calculate the distance between two histogram features, we discover the relationship between the test sample and the training samples of the same class. Inspired by [32], we adopt the HI [42] kernel to depict the similarity of two LBP-based histogram feature for HSI classification, which is given by

$$\mathcal{K}_{HI}(\boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{L} \min(a_i, b_i) \tag{7}$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ ($\sum_i^L a_i = \sum_i^L b_i = 1$) denote any two normalized histogram vector with L bins, and it can be verified that $0 \leq k_{HI} \leq 1$. To evaluate the validity of HI kernel, two commonly used kernels (i.e., $linear$ and RBF kernels) in HSI are also applied to the proposed SR-based framework. The $linear$ kernel can be given as

$$\mathcal{K}_{\text{linear}}(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a}^\top \boldsymbol{b} \tag{8}$$

where $\top$ denotes the transpose operation, and the RBF kernel can be given as

$$\mathcal{K}_{RBF}(\boldsymbol{a},\boldsymbol{b}) = \exp(-\gamma\|\boldsymbol{a}-\boldsymbol{b}\|_2^2) \tag{9}$$

where $\gamma > 0$ denotes the width of RBF kernel and is adaptive to the dictionary sets. What is more, all feature vectors should be normalized to unit length with the ℓ_2-norm. For a HSI, the LBP operator is applied to each selected subspace after dimensionality reduction (DR) to form several LBP images, and LBP histogram features are often computed for the pixel of interest in its corresponding local LBP image region.

B. Weighted Kernel Sparse Representation Classifier

Suppose $\mathcal{D} = \{\mathcal{D}_i\}_{i=1}^c \in \mathcal{R}^{m\times n}$ is a training dictionary based on LBP histogram feature for HSI, where $\mathcal{D}_i \in \mathcal{R}^{m\times n_i}$ is the subdictionary for the ith class, m is the feature dimensionality, n_i is the number of atom in $\mathcal{D}_i$, and $n = \sum_i^c n_i$ is the total number of training dictionary atoms for all of the c classes. To capture the nonlinear similarity information of samples in the LBP-based feature space, we map the LBP histogram feature into the region-level kernel induced feature space, and the typical SRC in spatial region-level kernel space is extended as a spatial KSRC. The locality structure of samples contains more discriminative information which is equally essential for SRC. In this paper, we propose an adaptive weighted SRC in the spatial region-level kernel induced space for HSI, namely, weighted kernel sparse representation-based classifier (WKSRC), which can combine the locality structure information and LBP-based spatial KSRC together. The new problem can be formulated as the following weighted ℓ_1-minimization problem, i.e.,

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \arg\min_{\boldsymbol{\alpha}}\|\phi(\boldsymbol{y}) - \Phi(\mathcal{D})\boldsymbol{\alpha}\|_2 + \lambda\|\boldsymbol{\Gamma}_{\phi(\boldsymbol{y})}\boldsymbol{\alpha}\|_1 \\ &= \arg\min_{\boldsymbol{\alpha}}\|\boldsymbol{k}(\mathcal{D},\boldsymbol{y}) - \boldsymbol{K}\boldsymbol{\alpha}\|_2 + \lambda\|\boldsymbol{\Gamma}_{\phi(\boldsymbol{y})}\boldsymbol{\alpha}\|_1 \end{aligned} \tag{10}$$

$$\boldsymbol{\Gamma}_{\boldsymbol{\phi}(\boldsymbol{y})} = \begin{bmatrix} \mathrm{dist}(\phi(\boldsymbol{y}),\phi(\boldsymbol{d}_1)) & & 0 \\ & \ddots & \\ 0 & & \mathrm{dist}(\phi(\boldsymbol{y}),\phi(\boldsymbol{d}_n)) \end{bmatrix} \tag{11}$$

where $\boldsymbol{k}(\mathcal{D},\boldsymbol{y}) = [\mathcal{K}(\boldsymbol{d}_1,\boldsymbol{y}),...,\mathcal{K}(\boldsymbol{d}_n,\boldsymbol{y})]^T \in \mathcal{R}^{n\times 1}$ denotes the kernel vector in spatial region-level kernel space and $\boldsymbol{K} \in \mathcal{R}^{n\times n}$ denotes the corresponding spatial kernel Gram matrix with $\boldsymbol{K}_{i,j} = \mathcal{K}(\boldsymbol{d}_i,\boldsymbol{d}_j)$. $\|\boldsymbol{\Gamma}_{\phi(\boldsymbol{y})}\boldsymbol{\alpha}\|_1$ denotes the weighted ℓ_1-regularization term, $\mathrm{dist}(\phi(\boldsymbol{y}),\phi(\boldsymbol{d}_i))$ denotes the distance or similarity between each atom and the test sample in the region-level kernel space, and λ controls the weight between representation error term $\|\phi(\boldsymbol{y}) - \Phi(\mathcal{D})\boldsymbol{\alpha}\|_2$ and the weighted regularizer term. In this paper, we adopt the Gaussian kernel distance to capture the nonlinear distance information between test sample $\boldsymbol{y}$ and each atom $\boldsymbol{d}_j$ $(j = 1,...,n)$ in the spatial kernel space, which is as follows:

$$\begin{aligned}&\mathrm{dist}_g(\phi(\boldsymbol{y}),\phi(\boldsymbol{d}_j)) \\ &\quad= \exp(-\|\phi(\boldsymbol{y}) - \phi(\boldsymbol{d}_j)\|_2^2)/2\sigma^2 \\ &\quad= \exp(-[\mathcal{K}(\boldsymbol{y},\boldsymbol{y}) + \mathcal{K}(\boldsymbol{d}_i,\boldsymbol{d}_i) - 2\mathcal{K}(\boldsymbol{y},\boldsymbol{d}_i)])/2\sigma^2 \end{aligned} \tag{12}$$

where σ denotes the Gaussian kernel width. Once the locality adaptor coefficient $\boldsymbol{\alpha}$ is obtained, analogous to (3), the decision rule of WKSRC is determined by the lowest representation error $\arg\min_{i=1,...,c}\|\boldsymbol{k}(\mathcal{D},\boldsymbol{y}) - \boldsymbol{K}\delta_i(\boldsymbol{\alpha})\|_2$.

C. Class-Oriented Optimization

In the HSI, samples from the same class and different categories are both highly correlated. For a test sample, it may be faithfully reconstructed based on several different subdictionaries from various classes. The representation error for different classes are similar. To deal with the problem, class-dependent strategy [43] is applied to SR-based classification framework. Different from traditional SR-based models, which mainly consider minimizing the reconstruction error, we focus on learning a more discriminative representation coefficient based on class specific dictionary. For the aforementioned problem (10), originally solved by the weighted ℓ_1-minimization problem, and is substituted by a class-oriented weighted ℓ_1-minimization problem (namely, CoWKSRC), which can be formulated as

$$\hat{\boldsymbol{\alpha}}_i = \arg\min_{\boldsymbol{\alpha}_i}\|\phi(\boldsymbol{y}) - \Phi(\mathcal{D}_i)\boldsymbol{\alpha}_i\|_2 + \lambda\|\boldsymbol{\Gamma}_{i,\phi(\boldsymbol{y})}\boldsymbol{\alpha}_{\boldsymbol{i}}\|_1 \tag{13}$$

$$\boldsymbol{\Gamma}_{\boldsymbol{i},\boldsymbol{\phi}(\boldsymbol{y})} = \begin{bmatrix} \mathrm{dist}_g(\phi(\boldsymbol{y}),\phi(\boldsymbol{d}_1)) & & 0 \\ & \ddots & \\ 0 & & \mathrm{dist}_g(\phi(\boldsymbol{y}),\phi(\boldsymbol{d}_{n_i})) \end{bmatrix} \tag{14}$$

where $\boldsymbol{\alpha}_i$ denotes a $n_i \times 1$ vector of weight coefficients, $\Phi(\mathcal{D}_i)$ denotes the class specific dictionary in the region-level kernel space, $\boldsymbol{\Gamma}_{\boldsymbol{i},\boldsymbol{\phi}(\boldsymbol{y})}$ denotes the weighted matrix specific to class i. By introducing the kernel Gram matrix $\boldsymbol{K}_i = \Phi_i^T\Phi_i$ and kernel vector $\boldsymbol{k}(\cdot,\boldsymbol{y}) = \Phi_i^T\phi(\boldsymbol{y})$, (13) becomes

$$\hat{\boldsymbol{\alpha}}_i = \arg\min_{\boldsymbol{\alpha}_i}\boldsymbol{\alpha}_i^T\boldsymbol{K}_i\boldsymbol{\alpha}_i + 2\boldsymbol{k}^T\boldsymbol{\alpha}_i + \lambda\|\boldsymbol{\Gamma}_{i,\phi(\boldsymbol{y})}\boldsymbol{\alpha}_{\boldsymbol{i}}\|_1 \tag{15}$$

which can be solved by weighted ℓ_1-minimization solvers class by class. Once obtain the class specific weight coefficient $\boldsymbol{\alpha}_i$, the reconstruction of $\boldsymbol{y}$ in the kernel space is $\phi(\hat{\boldsymbol{y}}_i) = \Phi(\mathcal{D}_i^{(n_i)})\boldsymbol{\alpha}_i$, the label of $\boldsymbol{y}$ is determined by the minimal total residual, that is

$$\mathrm{class}(\boldsymbol{y}) = \arg\min_{i=1,...,c} r_i(\boldsymbol{y}) \tag{16}$$

where $r_i(\boldsymbol{y}) = \|\phi(\boldsymbol{y}) - \phi(\hat{\boldsymbol{y}}_i)\|_2$ is the residual between the sparse approximation of class i and the corresponding test samples. The proposed CoWKSRC algorithm is described in Algorithm 1.

TABLE I
PARAMETERS SETTING FOR FEATURE EXTRACTION METHODS

Feature	Indian pines	University of Pavia	KSC
LBP [41]	No. of Base image: 3 Neighborhood and radius: $(8, 2)$ Local patch size: 21×21	No. of Base image: 7 Neighborhood and radius: $(8, 2)$ Local patch size: 17×17	No. of Base image: 7 Neighborhood and radius: $(8, 1)$ Local patch size: 21×21
Gabor [44]	No. of Base image: 10 Scale: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 Direction: 30°, 60°, 90°, 120°, 150°, 180°	No. of Base image: 10 Scale: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 Direction: 30°, 60°, 90°, 120°, 150°, 180°	No. of Base image: 10 Scale: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 Direction: 30°, 60°, 90°, 120°, 150°, 180°

Algorithm 1: Class-oriented WKSRC for HSI.

Input: Training dictionary $\mathcal{D} = \{\mathcal{D}_i\}_{i=1}^{c} \in \mathcal{R}^{m \times n}$, where $\mathcal{D}_i = \{\boldsymbol{d}_{ij}\}_{j}^{n_i} \in \mathcal{R}^{m \times n_i}$ is subdictionary associated with class i, test samples $\boldsymbol{y} \in \mathcal{R}^m$, λ, σ
Select a region-level kernel function $\mathcal{K}(\cdot, \cdot)$ and its parameters,
Compute its kernel Gram matrix $\boldsymbol{K}$ where $\boldsymbol{K}_{ij} = \mathcal{K}(\boldsymbol{d}_i, \boldsymbol{d}_j)$, and vector $\boldsymbol{k}(\cdot, \boldsymbol{y}) = [\mathcal{K}(\boldsymbol{d}_1, \boldsymbol{y}), ..., [\mathcal{K}(\boldsymbol{d}_n, \boldsymbol{y})]^T$
Normalize each column of $\boldsymbol{K}$ and $\boldsymbol{k}(\cdot, \boldsymbol{y})$ to have unit ℓ_2-norm
for $i = 1$ to c **do**
    1) Calculate the locality adaptor matrix $\boldsymbol{\Gamma}_{\boldsymbol{i}, \boldsymbol{\phi}(\boldsymbol{y})}$ specific to class i using Gaussian kernel distance according to (14);
    2) Solve the class-specific weighted ℓ_1-minimization problem (15) to obtain the weight coefficient $\boldsymbol{\alpha}_i$;
    3) Calculate the residual $r_i(\boldsymbol{y}) = \|\phi(\boldsymbol{y}) - \phi(\hat{\boldsymbol{y}}_i)\|_2$
end for
Decide the final label of $\boldsymbol{y}$ based on (16).
Output: The estimated label of $\boldsymbol{y}$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the effectiveness of the proposed method, we adopt three publicly available HSIs[1] the Airborne Visible/Infrared Imaging Spectrometer (Indian Pines) image, the Reflective Optics System Imaging Spectrometer (ROSIS) University of Pavia image, and the AVIRIS Kennedy Space Center (KSC) image. In our experiments, three different types of features (spectral feature, Gabor feature [44], and LBP feature [41]) are adopted to describe each HSI pixel, and PCA (as a dimensional reduction step) is used to obtain the first p-PC base image. The detailed parameters values used in the paper for different types of features are listed in Table I. Two spectral feature-based methods (i.e., NRS [40] and class-dependent sparse representation classifier (cdSRC) [43]) and three LBP feature-based methods (namely as, LBP-KSRC [38], LBP-WSRC and LBP-SVM [45]) are taken as baseline methods for comparison purposes. In order to test the effectiveness of region-level kernel, Gabor feature-based WKSRC (Gabor-WKSRC) was also presented in the following tests. It should be noted that SVM is implemented with the help of the LIBSVM [46] package. As for the region-level kernel, the parameter γ for RBF kernel is set by the median value of $1/(\|\boldsymbol{d}_i - \hat{\boldsymbol{d}}\|_2^2)$, $i = 1, ..., n$, where $\hat{\boldsymbol{d}} = (1/n)\sum_{i=1}^{n} \boldsymbol{d}_i$ is the mean of all available dictionary atoms. To validate the performance of the proposed method, three metrics (i.e., overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ)) are used for following experiments.

[1] Available online: http://www.ehu.es/ccwintco/index.php

TABLE II
NUMBER OF TRAINING AND TEST SAMPLES FOR THE INDIAN PINE IMAGE

Class		Samples	
No	Name	Train	Test
1	Alfalfa	6	48
2	Corn-no till	144	1290
3	Corn-min till	84	750
4	Corn	24	210
5	Grass/Pasture	50	447
6	Grass/Trees	75	672
7	Grass/Pasture-mowed	3	23
8	Hay-windrowed	49	440
9	Oats	2	18
10	Soybeans-no till	97	871
11	Soybeans-min till	247	2221
12	Soybean-clean till	62	552
13	Wheat	22	190
14	Woods	130	1164
15	Bldg-grass-tree-drives	38	342
16	Stone-steel towers	10	85
Total		1043	10366

A. Experiment on AVIRIS Indian Pines Image

1) Data description: The first HSI in our experiment was gathered by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene, covers a mixed agricultural/forest area, and consists of 145×145 pixels and 220 spectral bands range from 0.4 to 2.5 μm with a spatial resolution of 20 m/pixel. In the following experiments, the number of bands is reduced to 200 by removing 20 water absorption bands. A total of 10 366 samples containing 16 different land-cover classes are available, most of which are different types of crops (e.g., corns, soybeans, and wheats), as are detailed in Table II. This scene constitutes a very challenging classification problem due to the significant presence of mixed pixels and unbalanced labeled classes. For each class, we randomly choose around 10% of the labeled samples for training and the remaining samples used for testing. The number of training and test samples for each class is shown in Table II, and the spatial distribution of them is visually shown in Fig. 2(b).

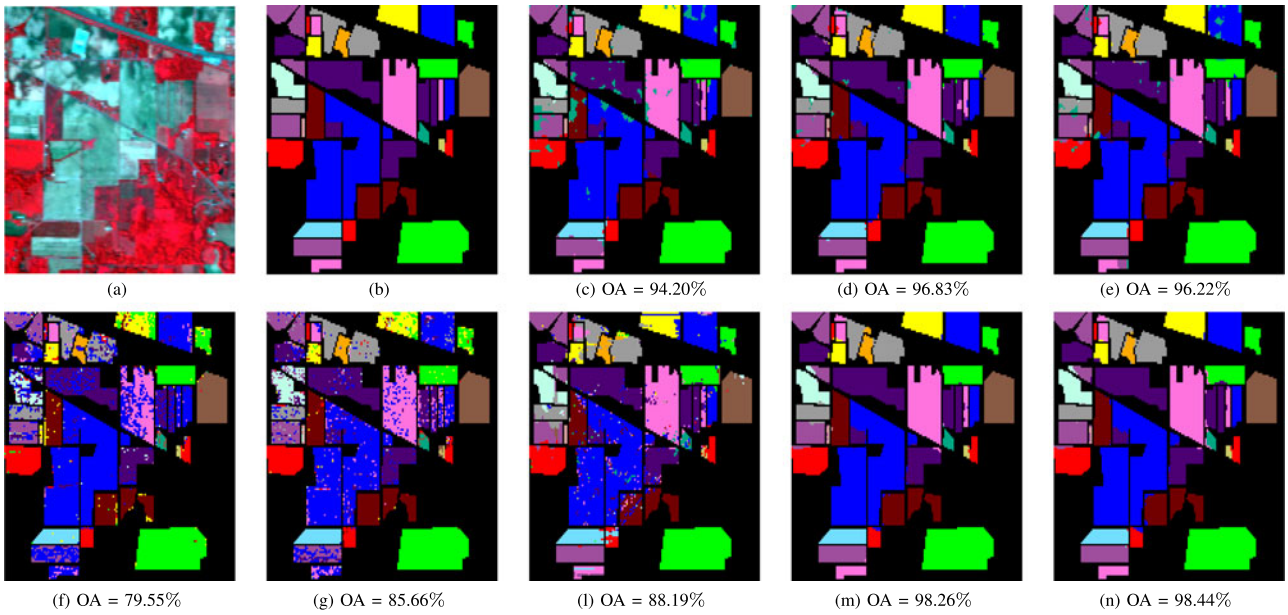

Fig. 2. Indian image: (a) False color image. (b) Ground-truth map (with 16 classes); and the classification map obtained by: (c) LBP-KSRC, (d) LBP-WSRC, (e) LBP-SVM, (f) NRS, (g) cdSRC, (h) Gabor-WKSRC, (i) WKSRC, and (j) CoWKSRC.

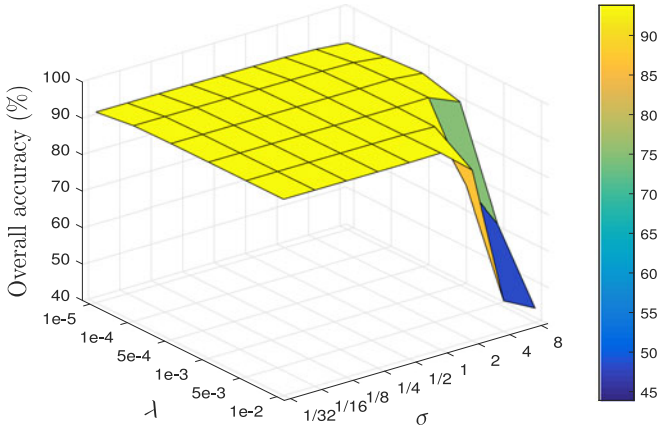

Fig. 3. Parameter tuning of λ and σ for the proposed CoWKSRC for AVIRIS Indian pines image.

2) Parameter analysis: First of all, we evaluate the effect of the parameters λ and σ on the OA performance of the proposed CoWKSRC algorithm on the Indian Pines image. The ℓ_1-regularization parameter λ is varied in the range $\{1e-5, 1e-4, 5e-4, 1e-3, 5e-3, 1e-2\}$, and the Gaussian kernel width σ is varied in the range $\{1/32, 1/16, 1/8, 1/4, 1/2, 1, 2, 4, 8\}$. The OA performance versus different λ and σ is depicted in Fig. 3, so we set $\lambda = 1e-4$ and $\sigma = 2$ for the following experiments.

3) Classification results: The classification results of the proposed HI-based region-level kernel methods (WKSRC and CoWKSRC) with several baseline classifiers (LBP-KSRC, LBP-WSRC, LBP-SVM, NRS, and cdSRC) on the AVIRIS Indian Pines image are summarized in Table III. From Table III, the proposed HI-based region-level kernel methods outperform other two state-of-the-art spectral-based methods (i.e., NRS and cdSRC). As a class specific method, different information (representation error and euclidean distance) are integrated into SR-based classification framework in cdSRC, and it achieves a higher accuracy than the CR-based method with local structure information (i.e., NRS). As ℓ_1-norm-based methods, the coding coefficients reflect the correlation between the test sample and the selected dictionary. Based on the whole or class-specific dictionary, the nonzero position and the relative value of the coding coefficients change small. Hence, WKSRC has similar classification result with its class-oriented optimization variant (i.e., CoWKSRC). WKSRC and CoWKSRC have a better classification performance than LBP-based methods with different state-of-the-art classifiers (namely, LBP-KSRC, LBP-WSRC, and LBP-SVM). LBP-KSRC (based on RBF kernel) achieve a relevant satisfying classification result of 93.63% by mapping into a nonlinear kernel space. As an extension of typical SRC, LBP-WSRC integrates the local structure information between the test pixel and each dictionary atom as the prior information into the SRC model. Compared to LBP-KSRC, LBP-WSRC (Euclidean distance) is more effective with 2.38% higher accuracy. Based on the similar classification model, LBP-based method (WKSRC) has a much higher accuracy than Gabor-filtering feature method (Gabor-WKSRC). The corresponding classification maps are shown in Fig. 2(c)–(n).

4) Evaluation of multiple region-level kernel variants: To illustrate the performance of the region-level kernel-based method with multiple kernel variants, three kernels, i.e., RBF, $linear$, and HI kernel, are visually and quantitatively compared, which are both optimized by class oriented strategy, and denoted by CoWKSRC-RBF, CoWKSRC-$linear$, and CoWKSRC-HI, respectively. We randomly choose a test pixel located at (23,107) belonging to class 8 in Indian pines image. The representation coefficient of the proposed CoWKSRC with three different kernels are visually depicted in Fig. 4(a)–(c) based on 1043 training samples, the corresponding sample

TABLE III
CLASSIFICATION ACCURACY (AVERAGED ON TEN MONTE CARLO RUNS) OBTAINED BY DIFFERENT CLASSIFIERS FOR THE AVIRIS INDIAN PINES IMAGE WITH 10% TRAINING SAMPLES PER CLASS

Class	LBP-KSRC	LBP-WSRC	LBP-SVM	NRS [40]	cdSRC [43]	Gabor-WKSRC	WKSRC	CoWKSRC
1	95.22 ± 3.18	92.36 ± 3.18	94.44 ± 7.89	72.50 ± 4.97	76.67 ± 8.12	62.92 ± 9.25	86.67 ± 9.50	$\mathbf{96.67 \pm 3.78}$
2	92.87 ± 0.89	95.87 ± 1.16	94.75 ± 0.65	68.82 ± 1.33	79.40 ± 2.03	84.02 ± 1.94	97.07 ± 1.44	$\mathbf{97.67 \pm 1.14}$
3	89.29 ± 2.43	90.07 ± 0.87	96.67 ± 0.67	52.93 ± 3.64	71.15 ± 3.03	81.55 ± 5.72	$\mathbf{97.95 \pm 2.15}$	97.01 ± 2.69
4	90.00 ± 3.12	95.73 ± 3.43	93.89 ± 0.27	39.14 ± 7.88	69.05 ± 8.30	89.90 ± 4.09	$\mathbf{98.10 \pm 1.58}$	$\mathbf{98.10 \pm 1.65}$
5	90.90 ± 1.37	94.41 ± 2.66	94.04 ± 1.75	90.43 ± 2.88	93.42 ± 0.89	81.16 ± 2.81	$\mathbf{96.73 \pm 1.88}$	95.30 ± 3.97
6	95.39 ± 1.86	96.28 ± 2.10	96.26 ± 0.83	95.00 ± 1.53	$\mathbf{98.42 \pm 0.34}$	90.03 ± 2.06	97.35 ± 1.06	98.36 ± 1.09
7	89.86 ± 5.02	91.30 ± 4.13	$\mathbf{93.21 \pm 1.62}$	49.57 ± 22.51	32.17 ± 14.29	71.30 ± 13.61	90.43 ± 16.95	92.17 ± 13.19
8	98.17 ± 0.80	99.70 ± 0.19	98.82 ± 0.13	98.64 ± 0.48	99.82 ± 0.19	97.50 ± 1.66	$\mathbf{100 \pm 0.00}$	$\mathbf{100 \pm 0.00}$
9	67.04 ± 4.56	66.67 ± 9.03	60.81 ± 8.49	34.44 ± 7.24	$\mathbf{70.00 \pm 23.77}$	42.22 ± 12.17	47.78 ± 12.79	68.89 ± 6.33
10	94.07 ± 2.05	96.21 ± 1.37	95.67 ± 1.89	51.76 ± 2.54	80.71 ± 1.10	84.45 ± 2.48	$\mathbf{97.91 \pm 2.38}$	97.89 ± 2.36
11	94.69 ± 0.32	96.15 ± 0.28	96.03 ± 0.51	94.65 ± 0.74	85.28 ± 2.05	86.79 ± 1.34	$\mathbf{99.42 \pm 0.29}$	99.12 ± 0.41
12	91.85 ± 1.26	94.21 ± 1.34	93.68 ± 2.14	63.55 ± 3.76	81.92 ± 2.12	83.44 ± 3.18	$\mathbf{98.70 \pm 1.28}$	97.79 ± 1.34
13	93.16 ± 5.79	98.07 ± 1.61	98.30 ± 1.22	95.79 ± 2.07	99.47 ± 0.02	95.89 ± 3.23	98.00 ± 2.09	$\mathbf{100 \pm 0.00}$
14	96.83 ± 0.78	98.71 ± 0.42	97.89 ± 0.1	96.60 ± 0.84	95.74 ± 1.44	97.66 ± 1.23	99.42 ± 0.52	$\mathbf{99.85 \pm 0.21}$
15	95.86 ± 1.22	97.56 ± 0.74	95.68 ± 0.35	61.99 ± 5.41	74.33 ± 6.37	91.40 ± 5.51	96.55 ± 2.37	$\mathbf{98.54 \pm 0.88}$
16	82.35 ± 8.24	89.02 ± 3.75	89.20 ± 4.44	84.71 ± 5.88	85.88 ± 7.30	$\mathbf{96.71 \pm 3.26}$	69.65 ± 6.83	86.12 ± 6.30
OA	93.63 ± 0.21	96.01 ± 0.37	95.88 ± 0.35	79.27 ± 0.63	85.35 ± 0.53	87.66 ± 0.47	97.92 ± 0.34	$\mathbf{98.22 \pm 0.28}$
AA	92.78 ± 0.34	93.29 ± 2.01	92.90 ± 1.05	71.91 ± 2.13	80.84 ± 2.52	83.56 ± 1.75	91.98 ± 1.49	$\mathbf{95.22 \pm 0.78}$
κ	93.43 ± 0.24	95.56 ± 0.35	95.49 ± 0.40	75.88 ± 0.75	83.26 ± 0.59	85.95 ± 0.53	97.63 ± 0.39	$\mathbf{97.97 \pm 0.32}$

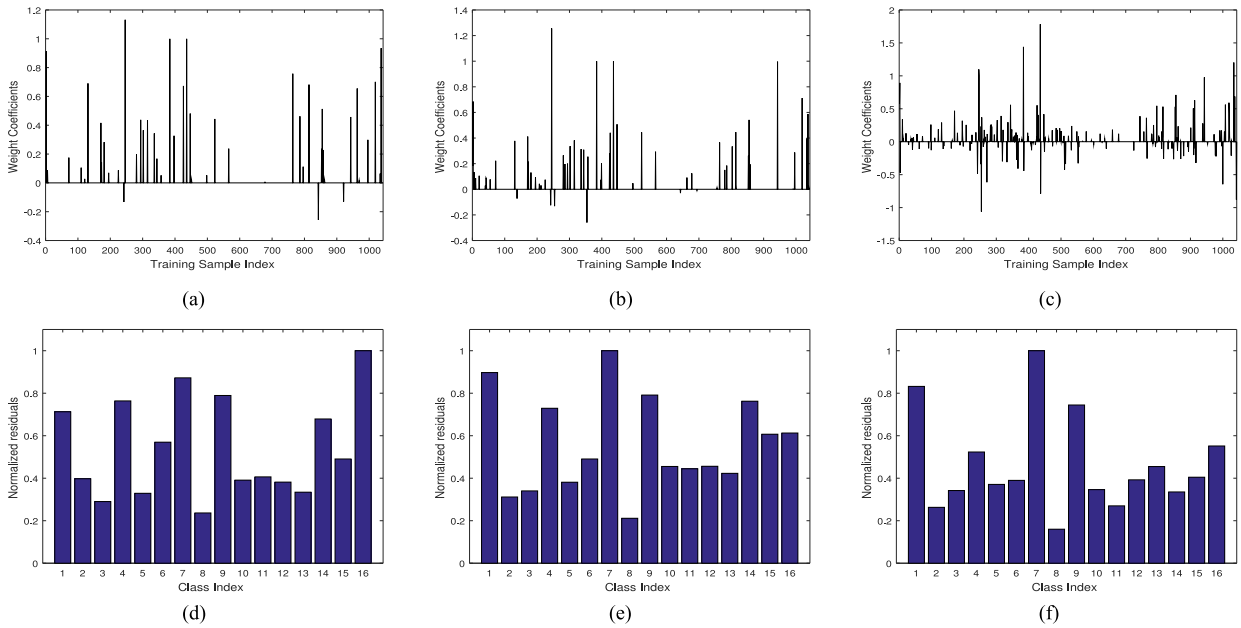

Fig. 4. Representation coefficients obtained by (a) CoWKSRC-RBF, (b) CoWKSRC-linear, (c) CoWKSRC-HI. Normalized residuals obtained by (d) CoWKSRC-RBF. (e) CoWKSRC-$linear$, and (f) CoWKSRC-HI.

index for class 8 (marked by red color) is from 386 to 435. The reconstruction error (i.e., normalized residual) for each class of the proposed method with different kernels are given in Fig. 4(d)–(f), respectively. From the figure, it can be seen that all the spatial region-level kernel methods can classify the test pixel properly, but the obtained ℓ_1-norm coefficients in different region-level kernel feature space are different. The weight coefficients in HI-kernel feature space are uniformly distributed among the dictionary atoms from different classes, and it is evident that the determination in HI-kernel space is more robust for classification task than in $linear$- and RBF-kernel feature space. Moreover, the label of this test pixel is determined by the minimum residual criterion. The OA performance of the proposed CoWKSRC with different kernel methods is reported in Table IV. The HI kernel-based method achieves the best classification performance.

B. Experiment on ROSIS University of Pavia Image

The second HSI was acquired by the Reflective Optics System Imaging Spectrometer-03 (ROSIS-03) sensor, which covers the area over the University of Pavia, Italy, with a pixel response in 115 spectral channels (with spectral range from 0.43 to 0.86 μm) [1]. After removing the 12 noisy bands, the remaining 103 channels are reserved. The image consists of 610×340 pixels with a spatial resolution of 1.3 m per pixel, and

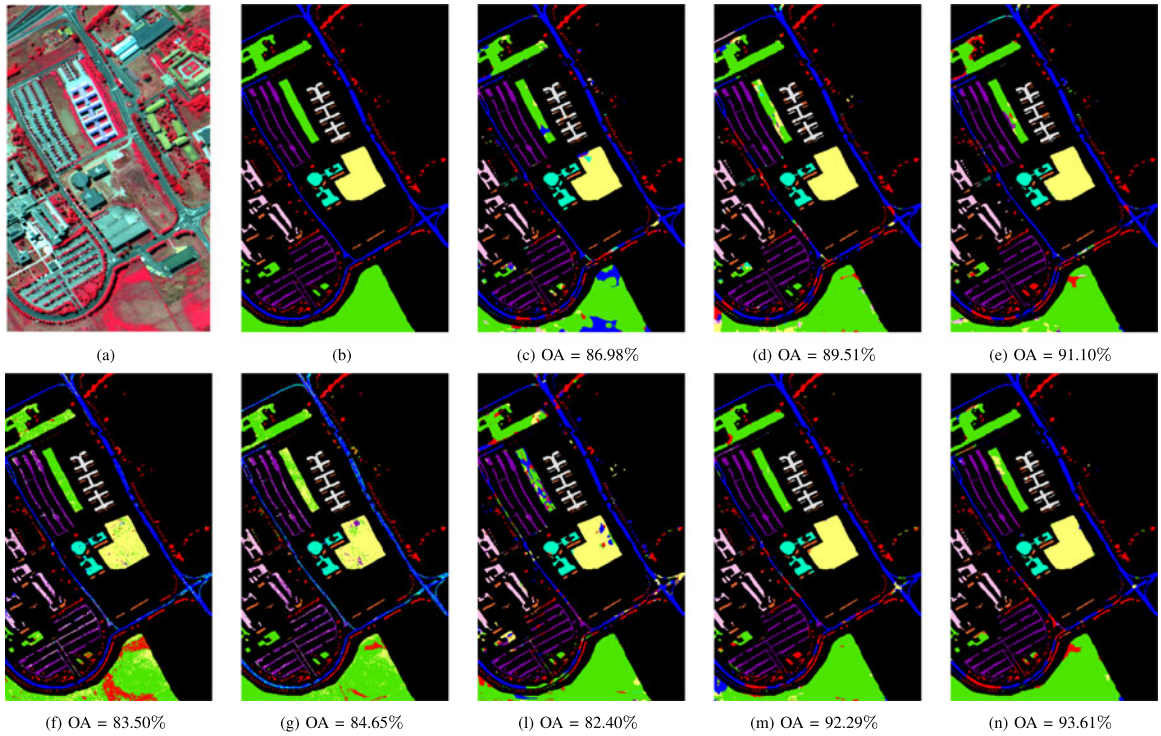

Fig. 5. University of the Pavia image: (a) False color image. (b) Ground-truth map (with nine classes); and the classification map obtained by: (c) LBP-KSRC, (d) LBP-WSRC, (e) LBP-SVM, (f) NRS, (g) cdSRC, (h) Gabor-WKSRC, (i) WKSRC, and (j) CoWKSRC.

TABLE IV
OVERALL CLASSIFICATION ACCURACY OBTAINED BY THE PROPOSED COWKSRC WITH MULTIPLE KERNEL METHODS AS A FUNCTION OF THE NUMBER OF TRAINING SAMPLES PER CLASS FOR AVIRIS INDIAN PINES IMAGE

CoWKSRC	*linear*	*RBF*	*HI*
5%	91.87 ± 0.57	89.85 ± 0.41	$\mathbf{93.49 \pm 0.63}$
10%	95.70 ± 0.74	94.11 ± 0.36	$\mathbf{98.22 \pm 0.28}$
15%	97.51 ± 0.62	96.26 ± 0.56	$\mathbf{98.71 \pm 0.51}$
20%	97.94 ± 0.68	96.90 ± 0.45	$\mathbf{98.84 \pm 0.39}$

TABLE V
NUMBER OF TRAINING AND TESTING SAMPLES FOR THE UNIVERSITY OF PAVIA IMAGE

Class		Samples	
No	Name	Train	Test
1	Alfalfa	60	6571
2	Meadows	60	18589
3	Gravel	60	2039
4	Trees	60	3004
5	Metal sheet	60	1285
6	Bare soil	60	4969
7	Bitumen	60	1270
8	Brick	60	3622
9	Shadows	60	887
Total		540	42236

contains nine ground-truth classes. The false color composite image and the ground-truth map are shown in Fig. 5(a) and (b), respectively. We randomly choose 60 samples for each class as the training samples and use the remaining as the test samples (shown in Table V).

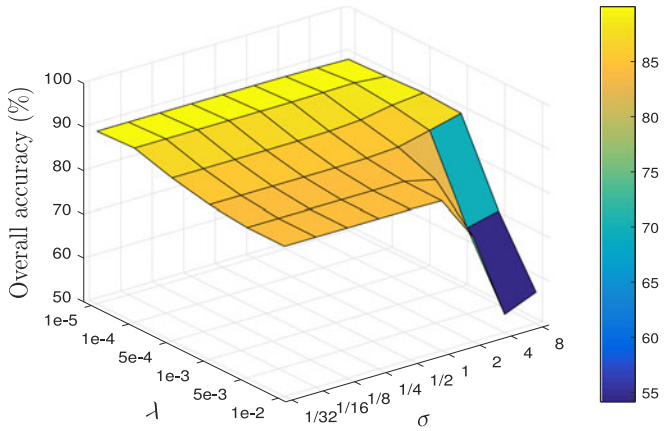

Fig. 6. Parameter tuning of λ and σ for the proposed CoWKSRC for the University of Pavia.

Fig. 6 illustrates the OA performance of the proposed CoWKSRC algorithm versus different λ and σ. We set $\lambda = 1e-4$ and $\sigma = 4$ for the University of Pavia image. Likewise, Table VI gives the OA performance of the proposed CoWKSRC-based region-level kernel with three different kernel methods under different training sample sizes. It is apparent that the HI-based method achieves the best performance among the three kernel-based methods. The numerical classification results using LBP-KSRC, LBP-WSRC, LBP-SVM, NRS, cdSRC, Gabor-WKSRC, WKSRC, and CoWKSRC are summarized in Table VII. Fig. 5(c)–(j) illustrates the corresponding classification maps obtained by various methods. As is shown

TABLE VII
CLASSIFICATION ACCURACY (AVERAGED ON TEN MONTE CARLO RUNS) OBTAINED BY DIFFERENT CLASSIFIERS OF ROSIS UNIVERSITY OF PAVIA IMAGE WITH 60 TRAINING SAMPLES PER CLASS

Class	LBP-KSRC	LBP-WSRC	LBP-SVM	NRS [40]	cdSRC [43]	Gabor-WKSRC	WKSRC	CoWKSRC
1	87.65 ± 2.16	84.16 ± 3.74	74.14 ± 5.70	83.55 ± 0.79	84.15 ± 3.80	69.01 ± 2.09	84.82 ± 2.74	$\mathbf{85.51 \pm 2.14}$
2	82.58 ± 1.26	87.08 ± 1.70	94.02 ± 1.05	78.49 ± 4.68	82.34 ± 4.23	83.73 ± 1.80	$\mathbf{93.11 \pm 2.33}$	92.57 ± 2.72
3	96.41 ± 1.43	96.07 ± 1.48	96.25 ± 1.07	75.64 ± 3.74	76.32 ± 5.65	92.32 ± 2.41	96.75 ± 1.53	$\mathbf{97.05 \pm 1.30}$
4	66.48 ± 3.28	74.76 ± 2.47	90.08 ± 3.37	95.50 ± 2.62	95.97 ± 1.04	59.21 ± 2.34	79.67 ± 4.45	83.75 ± 2.91
5	93.56 ± 1.72	94.16 ± 2.64	97.17 ± 0.93	99.60 ± 0.15	99.58 ± 0.12	97.17 ± 2.28	94.21 ± 1.87	94.69 ± 2.46
6	98.61 ± 1.14	98.37 ± 1.39	95.69 ± 0.40	88.16 ± 3.41	85.64 ± 4.51	90.20 ± 3.20	$\mathbf{97.14 \pm 2.37}$	95.87 ± 3.10
7	98.30 ± 1.40	98.50 ± 0.85	95.04 ± 2.52	91.42 ± 2.67	92.05 ± 1.83	93.40 ± 4.13	98.08 ± 1.34	$\mathbf{99.24 \pm 0.53}$
8	94.46 ± 2.81	94.82 ± 2.02	94.29 ± 1.31	84.11 ± 5.80	83.35 ± 4.48	88.55 ± 5.63	$\mathbf{97.17 \pm 1.25}$	95.97 ± 1.06
9	83.29 ± 3.41	85.52 ± 2.76	74.63 ± 6.56	99.41 ± 0.20	$\mathbf{99.59 \pm 0.23}$	88.32 ± 1.33	81.24 ± 3.08	83.63 ± 3.16
OA	86.62 ± 0.50	88.70 ± 0.95	90.69 ± 0.79	83.44 ± 1.98	84.95 ± 1.57	82.08 ± 1.05	91.80 ± 1.16	$\mathbf{91.82 \pm 1.04}$
AA	89.04 ± 0.44	90.38 ± 0.58	90.02 ± 0.94	88.43 ± 0.22	88.78 ± 0.28	84.66 ± 1.11	91.35 ± 0.66	$\mathbf{92.03 \pm 0.51}$
κ	82.71 ± 0.62	85.34 ± 1.19	88.24 ± 1.04	78.79 ± 2.33	80.57 ± 1.85	76.76 ± 1.31	89.23 ± 1.48	$\mathbf{89.27 \pm 1.32}$

TABLE VI
OVERALL CLASSIFICATION ACCURACY OBTAINED BY THE PROPOSED COWKSRC WITH MULTIPLE KERNEL METHODS AS A FUNCTION OF THE NUMBER OF TRAINING SAMPLES PER CLASS FOR UNIVERSITY OF PAVIA IMAGE

CoWKSRC	$linear$	RBF	HI
40	81.90 ± 1.09	81.29 ± 1.32	$\mathbf{88.35 \pm 1.11}$
50	84.77 ± 1.21	84.80 ± 0.99	$\mathbf{91.33 \pm 1.43}$
60	85.09 ± 1.37	85.30 ± 1.07	$\mathbf{91.82 \pm 1.04}$
70	87.52 ± 1.14	88.19 ± 1.20	$\mathbf{94.48 \pm 0.98}$

TABLE VIII
NUMBER OF TRAINING AND TESTING SAMPLES FOR THE KENNEDY SPACE CENTER IMAGE

Class		Samples	
No	Name	Train	Test
1	Scrub	10	751
2	Willow swamp	10	233
3	Cabbage palm	10	246
4	Cabbage palm/oak	10	242
5	Slash pine	10	151
6	Oak/broadleaf	10	219
7	Hardwood swamp	10	95
8	Graminoid marsh	10	421
9	Spartina marsh	10	510
10	Cattail marsh	10	394
11	Salt marsh	10	409
12	Mud flats	10	493
13	Water	10	917
	Total	90	5121

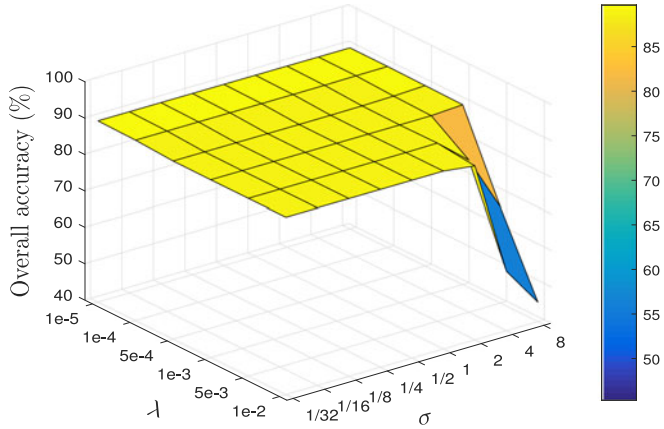

Fig. 7. Parameter tuning of λ and σ for the proposed CoWKSRC for Kennedy Space Center.

TABLE IX
OVERALL CLASSIFICATION ACCURACY OBTAINED BY THE PROPOSED COWKSRC WITH MULTIPLE KERNEL METHODS AS A FUNCTION OF THE NUMBER OF TRAINING SAMPLES PER CLASS FOR KENNEDY SPACE CENTER IMAGE

CoWKSRC	Linear	RBF	HI
5	77.23 ± 1.28	66.89 ± 1.76	$\mathbf{81.75 \pm 1.39}$
10	91.52 ± 1.67	70.32 ± 2.32	$\mathbf{95.39 \pm 1.43}$
15	96.85 ± 2.01	71.58 ± 1.98	$\mathbf{98.90 \pm 1.82}$
20	98.26 ± 1.54	73.48 ± 2.15	$\mathbf{99.62 \pm 1.61}$

in Table VII, the proposed methods (WKSRC and CoWKSRC) based on spatial region-level kernel outperform other spectral-based (i.e., NRS and cdSRC) methods in terms of classification performance. Based on the same spatial feature, the proposed methods have better classification results than other classifiers based LBP methods. Based on different spatial kernel feature, under the same classifier, region-level method (i.e., WKSRC) has a much higher accuracy than Gabor-filtering kernel method (i.e., Gabor-WKSRC).

C. Experiment on AVIRIS Kennedy Space Center Image

The third HSI is gathered by AVIRIS over the Kennedy Space Center (KSC), Florida, USA, on March, 1996. This image contains 224 bands whose wavelength covers the spectral range from 0.4 to 2.5 μm, and the image size is 512×614 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 classification. This scene contains 13 ground-truth classes. As is shown in Table VIII, 10 samples per class are chosen as training samples and the remaining used for testing. The false color image and ground-truth map are shown in Fig. 8(a) and (b).

Fig. 7 illustrates the influence of λ and σ on the OA performance of the proposed region-level kernel method (CoWKSRC). As shown in Table IX, the HI-based method has a

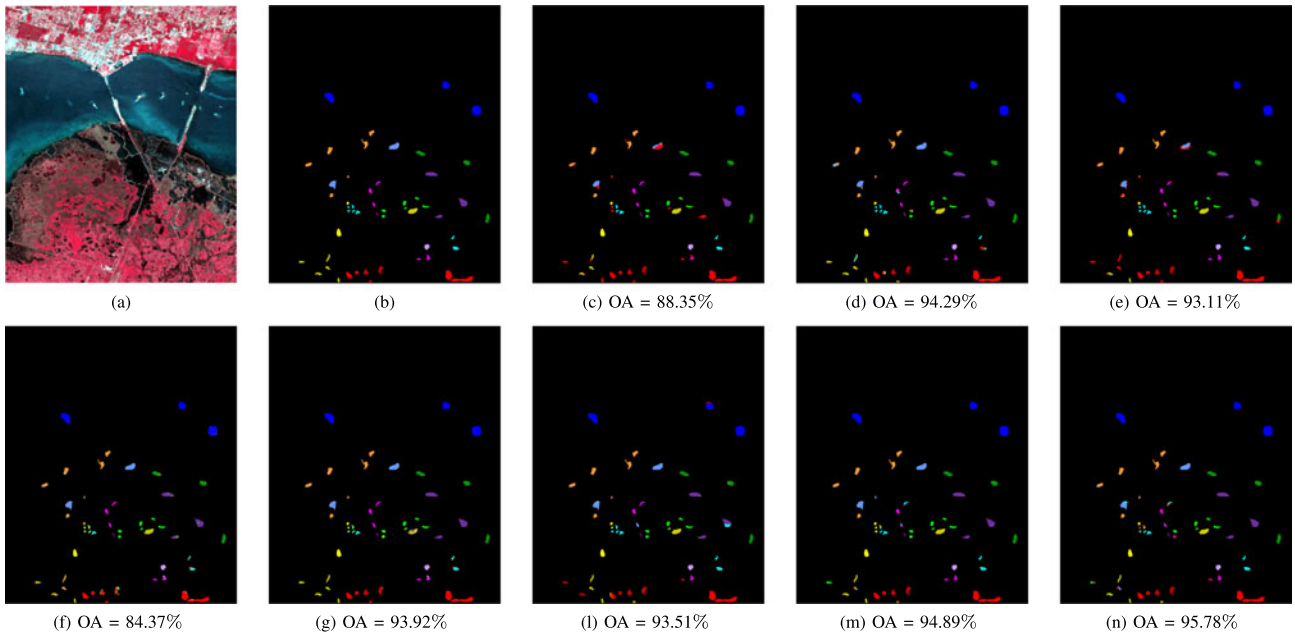

Fig. 8. KSC image: (a) False color image. (b) Ground-truth map (with 13 classes); and the classification map obtained by: (c) LBP-KSRC, (d) LBP-WSRC, (e) LBP-SVM, (f) NRS, (g) cdSRC, (l) Gabor-WKSRC, (m) WKSRC, and (n) CoWKSRC.

TABLE X
CLASSIFICATION ACCURACY (AVERAGED ON TEN MONTE CARLO RUNS) OBTAINED BY DIFFERENT CLASSIFIERS OF KENNEDY SPACE CENTER IMAGE WITH TEN TRAINING SAMPLES PER CLASS

Class	LBP-KSRC	LBP-WSRC	LBP-SVM	NRS [40]	cdSRC [43]	Gabor-WKSRC	WKSRC	CoWKSRC
1	96.32 ± 0.26	91.21 ± 3.34	96.80 ± 0.36	82.77 ± 10.00	**94.11 ± 3.57**	84.05 ± 5.41	91.24 ± 6.85	88.28 ± 9.28
2	87.70 ± 8.94	84.12 ± 9.66	90.70 ± 5.07	85.41 ± 5.68	**87.38 ± 5.38**	83.61 ± 10.61	84.72 ± 8.63	86.35 ± 12.22
3	79.00 ± 4.17	96.75 ± 5.63	92.28 ± 3.14	89.11 ± 4.01	88.86 ± 5.60	89.92 ± 14.19	94.88 ± 7.39	**98.21 ± 2.52**
4	95.73 ± 5.00	95.73 ± 3.84	90.91 ± 3.79	45.54 ± 14.98	67.93 ± 5.27	**96.94 ± 3.49**	86.61 ± 10.26	86.03 ± 9.45
5	93.60 ± 11.09	100 ± 0.00	93.60 ± 5.09	67.68 ± 9.77	85.70 ± 6.92	99.07 ± 2.07	96.16 ± 8.59	**100 ± 0.00**
6	99.54 ± 0.79	98.45 ± 0.36	100 ± 0.00	57.53 ± 11.45	75.80 ± 20.09	90.87 ± 5.05	99.91 ± 0.20	**100 ± 0.00**
7	95.31 ± 1.53	100 ± 0.00	100.00 ± 0.00	91.16 ± 5.70	95.58 ± 2.72	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**
8	66.11 ± 8.61	82.82 ± 7.17	77.51 ± 9.51	74.44 ± 6.02	87.46 ± 1.28	**93.25 ± 9.91**	91.69 ± 8.75	88.22 ± 10.66
9	91.44 ± 4.93	94.77 ± 8.22	87.84 ± 5.06	89.29 ± 13.86	**99.41 ± 0.69**	89.96 ± 7.61	97.69 ± 2.42	98.08 ± 4.19
10	89.59 ± 4.59	98.98 ± 1.76	95.43 ± 5.67	87.46 ± 2.85	98.58 ± 0.29	98.68 ± 2.81	91.37 ± 15.64	**100 ± 0.00**
11	65.53 ± 3.60	97.15 ± 3.39	99.35 ± 1.13	95.01 ± 3.17	97.16 ± 0.84	96.82 ± 3.80	95.89 ± 4.49	**99.32 ± 1.16**
12	79.58 ± 0.82	82.76 ± 5.92	80.32 ± 3.03	79.55 ± 3.39	**97.69 ± 1.31**	93.43 ± 7.56	95.50 ± 7.55	97.20 ± 3.59
13	97.26 ± 1.12	**100 ± 0.00**	**100 ± 0.00**	97.93 ± 0.00	**100 ± 0.00**	97.60 ± 2.26	**100 ± 0.00**	99.98 ± 0.01
OA	87.68 ± 0.63	93.48 ± 1.73	92.70 ± 0.87	83.78 ± 2.19	93.27 ± 0.93	92.84 ± 1.41	94.64 ± 1.70	**95.39 ± 1.43**
AA	87.78 ± 0.47	93.98 ± 1.83	92.67 ± 1.00	80.22 ± 2.09	90.44 ± 2.06	93.40 ± 2.09	94.28 ± 1.50	**95.51 ± 1.06**
κ	86.52 ± 0.71	92.67 ± 1.92	91.86 ± 0.98	81.96 ± 2.42	92.50 ± 1.05	92.04 ± 1.58	94.04 ± 1.89	**94.87 ± 1.59**

TABLE XI
VARIANTS OF THE PROPOSED COWKSRC

Notation	Characteristic	Description
CoWKSRC-S	Nonspatial	Weighted Kernel SRC based on spectral feature via Class-oriented Optimization
CoWSRC	Nonkernel	Weighted SRC based on LBP histogram feature via Class-oriented Optimization
CoKSRC	Nonweighted	Kernel SRC with LBP-based region-level kernel via Class-oriented Optimization
WKSRC	Global Optimization	Weighted Kernel SRC with LBP-based region-level kernel via Global Optimization

better classification performance than the other two kernel-based methods. The numerical classification results of the proposed methods (WKSRC and CoWKSRC) with several baseline classifiers (i.e., LBP-KSRC, LBP-WSRC, LBP-SVM, NRS, and cdSRC) on the Kennedy Space Center Image are summarized in Table X. Fig. 8(c)–(j) shows the corresponding classification maps for various methods. Among all the SR-based methods, the proposed spatial region-level kernel methods (WKSRC and CoWKSRC) yield the highest result. Moreover, the proposed methods have a better classification performance than other LBP-feature-based methods. Depend on different spatial features and with the similar classifier, region-level kernel methods (i.e., WKSRC and CoWKSRC) have a better performance than Gabor-filtering feature method.

TABLE XII
OA RESULTS (%) OBTAINED BY MULTIPLE VARIANTS OF COWKSRC WITH DIFFERENT PERTCENT (OR NUMBER) OF DICTIONARY ATOMS

Methods		CoWKSRC-S	CoWSRC	CoKSRC	WKSRC	CoWKSRC
Indian Pines	%5	73.90 ± 2.12	82.87 ± 0.22	94.78 ± 0.78	94.90 ± 0.61	$\mathbf{95.36 \pm 0.34}$
	%10	77.78 ± 1.73	83.08 ± 0.16	97.32 ± 0.16	97.55 ± 0.16	$\mathbf{98.06 \pm 0.06}$
	%15	79.10 ± 1.61	84.21 ± 0.49	97.40 ± 0.32	97.46 ± 0.30	$\mathbf{98.98 \pm 0.20}$
	%20	79.72 ± 1.56	84.99 ± 1.95	98.06 ± 0.11	98.28 ± 0.12	$\mathbf{99.06 \pm 0.15}$
University of Pavia	40	63.33 ± 1.77	87.34 ± 2.20	87.98 ± 2.64	88.65 ± 1.32	$\mathbf{89.38 \pm 1.05}$
	50	64.36 ± 2.09	89.32 ± 1.97	90.51 ± 0.99	90.83 ± 0.96	$\mathbf{91.06 \pm 1.11}$
	60	65.25 ± 1.91	90.31 ± 1.49	91.56 ± 1.16	91.88 ± 1.10	$\mathbf{92.21 \pm 0.84}$
	70	65.47 ± 2.14	91.63 ± 1.02	92.39 ± 0.88	92.64 ± 0.99	$\mathbf{94.68 \pm 0.49}$
KSC	5	73.30 ± 1.77	83.97 ± 0.56	84.51 ± 0.62	85.15 ± 0.31	$\mathbf{86.37 \pm 0.15}$
	10	77.25 ± 1.25	93.23 ± 0.49	93.78 ± 0.39	94.13 ± 0.18	$\mathbf{95.32 \pm 0.24}$
	15	80.38 ± 0.86	96.40 ± 0.37	96.82 ± 0.46	97.03 ± 0.34	$\mathbf{98.15 \pm 0.23}$
	20	80.18 ± 1.19	97.12 ± 0.69	98.21 ± 0.58	98.63 ± 0.22	$\mathbf{99.04 \pm 0.17}$

D. Effect of Combining Multiple Strategies

In this section, we investigate the performance of each component (i.e., spatial feature, region-level kernel, weighted-strategy, and class-oriented optimization) in CoWKSRC. We refer these components to CoWKSRC-S, CoWSRC, CoKSRC, and WKSRC, respectively. The descriptions are listed in Table XI. For the Indian Pines image, 5%, 10%, 15%, and 20% of the labeled samples are randomly chosen as the training dictionary, and the remaining samples are used for the test. For the University of Pavia image, 40, 50, 60, and 70 atoms per class are selected as the training dictionary. For the KSC image, the number of dictionary atoms per class is varied from 5 to 20. From Table XII, without spatial feature, the OA results decrease significantly for the three HSI, thus, demonstrating the importance of LBP. The region-level kernel also plays a significant role in the proposed method. Without considering the region-level kernel, the OA of CoWKSRC is decreased from 95.36% to 89.21% for the Indian Pines image (5% of the labeled samples). WKSRC shows slightly better results than CoKSRC, which indicates that class-oriented optimization is slight more important than weighted strategy. In conclusion, by integrating the components as mentioned above, our proposed method achieves the best performance.

E. Computational Time

Finally, we compare the computational time of the proposed WKSRC and its class-oriented variants. All the experiments are carried out using MATLAB R2016a on Intel(R) Core(TM) i7-4790 CPU PC machine with 16 GB of RAM. To investigate the superiority of class-oriented based methods, the GPU times (in seconds) of two pairwise methods, i.e., WKSRC, WSRC, and its class-oriented variants (CoWKSRC and CoWSRC), are listed in Table XIII, XIV, and XV with different sizes (or percent) of training samples on Indian Pines, University of the Pavia, and KSC images, respectively. The dominant computational time of the four methods come from the learning of sparse coefficients via weighted ℓ_1-norm regularization problem in the region-level kernel or spectral space, which is implemented by a weighted version of LARS via the SPArse Modeling Software (SPAMS) package [47], [48]. Although the proposed spatial region-level kernel method achieves similar classification performance with its class-oriented variant, the class-oriented variant is faster than WKSRC for the three HSIs (seen in Tables XIII, XIV, and XV). That is because CoWKSRC adopt the class-oriented strategy that leads to the monotonical decrease of the objective function and finds the global minimum without any risk of falling into the local minimum.

TABLE XIII
COMPUTATIONAL TIME (AVERAGED ON TEN MONTE CARLO RUNS) OBTAINED BY DIFFERENT CLASSIFIERS AS A FUNCTION OF THE PERCENT OF TRAINING SAMPLES PER CLASS FOR THE INDIAN PINES IMAGE

	CoWKSRC	WKSRC	CoWSRC	WSRC
5%	38.56 ± 0.67	104.18 ± 1.41	33.46 ± 0.27	42.99 ± 0.20
10%	114.11 ± 0.63	400.70 ± 2.20	64.58 ± 1.62	132.35 ± 0.41
15%	272.24 ± 1.86	906.98 ± 1.75	100.07 ± 0.69	250.26 ± 0.79
20%	490.18 ± 2.67	1652 ± 5.18	131.20 ± 1.71	392.13 ± 3.54

TABLE XIV
COMPUTATIONAL TIME (AVERAGED ON TEN MONTE CARLO RUNS) OBTAINED BY DIFFERENT CLASSIFIERS AS A FUNCTION OF THE NUMBER OF TRAINING SAMPLES PER CLASS FOR THE UNIVERSITY OF PAVIA IMAGE

	CoWKSRC	WKSRC	CoWSRC	WSRC
40	125.31 ± 4.03	356.89 ± 19.38	118.43 ± 4.46	153.18 ± 2.53
50	183.71 ± 3.41	549.20 ± 22.02	148.77 ± 8.92	216.38 ± 5.05
60	285.88 ± 11.32	740.56 ± 31.04	234.08 ± 6.27	186.74 ± 5.15
70	386.93 ± 7.34	952.95 ± 41.64	219.65 ± 25.32	289.65 ± 6.03

TABLE XV
COMPUTATIONAL TIME (AVERAGED ON TEN MONTE CARLO RUNS) OBTAINED BY DIFFERENT CLASSIFIERS AS A FUNCTION OF THE NUMBER OF TRAINING SAMPLES PER CLASS FOR THE KENNEDY SPACE CENTER IMAGE

	CoWKSRC	WKSRC	CoWSRC	WSRC
5	2.71 ± 0.06	3.14 ± 0.20	2.44 ± 0.14	2.72 ± 0.13
10	4.82 ± 0.07	5.80 ± 0.27	4.36 ± 0.09	5.23 ± 0.10
15	7.63 ± 0.11	9.54 ± 0.32	7.49 ± 0.02	6.21 ± 0.08
20	10.35 ± 0.19	13.53 ± 0.57	7.91 ± 0.28	10.21 ± 0.35

V. Conclusion

In this paper, we proposed a novel class-oriented WKSRC method based on region-level kernel feature for HSI classification. In the proposed method, region-level kernels are applied to map original spectral space into a high-dimensional kernel space to capture the nonlinear structure of the extracted LBP histogram feature. To further strengthen the discriminative ability, we combine region-level kernels and locality structure information based on Gaussian kernel distance into a unified SR-based framework, and propose a novel class-oriented WKSRC method solved by class specific weighted ℓ_1-minimization problem to obtain more discriminating weighting coefficients for HSI classification. Experimental results on three HSI demonstrate that the proposed algorithms yield better classification performance compared with other state-of-the-art classifiers. The future work will focus on how to construct a more effective multiple kernels for the proposed framework.

Acknowledgment

The authors would like to thank the associated 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, Supplement 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, Mar. 2007.

[3] M. Borengasser, W. S. Hungate, R. Watkins, and O. Beeri, *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, "Semi-supervised 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. 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.

[11] 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.

[12] M. Elad, *Sparse and Redundant Representations—From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[13] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan. 2008.

[14] 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.

[15] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. 2nd Ed. Hoboken, NJ, USA: Wiley, 2012.

[17] 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.

[18] 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.

[19] 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.

[20] J. Li, H. Zhang, and L. Zhang, "Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification," *ISPRS J. Photogramm.*, vol. 94, pp. 25–36, 2014.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. A. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, Oct. 2016.

[26] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, pp. 51–59, 1996.

[27] S. Xie, S. Shan, X. Chen, and W. Gao, "V-LGBP: Volume based local Gabor binary patterns for face representation and recognition," in *Proc. 19th Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[28] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, "Multi-scale block local binary patterns for face recognition," in *Proc. Int. Conf. Biometrics*, 2007, pp. 809–818.

[29] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognit.*, vol. 43, pp. 706–719, 2010.

[30] 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.

[31] R. Min and J.-L. Dugelay, "Improved combination of LBP and sparse representation based classification (SRC) for face recognition," in *Proc. 2011 IEEE Int. Conf. Multimedia Expo.*, 2011, pp. 1–6.

[32] C. Kang, S. Liao, S. Xiang, and C. Pan, "Kernel sparse representation with pixel-level and region-level local feature kernels for face recognition," *Neurocomputing*, vol. 133, pp. 141–152, 2014.

[33] C. Lu, H. Min, J. Gui, L. Zhu, and Y.-K. Lei, "Face recognition via weighted sparse representation," *J. Visual Commun. Image Representation*, vol. 24, pp. 111–116, 2013.

[34] W. Li and Q. Du, "Adaptive sparse representation for hyperspectral image classification," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2015, pp. 4955–4958.

[35] W. Li and Q. Du, "A survey on representation-based classification and detection in hyperspectral remote sensing imagery," *Pattern Recognit. Lett.*, vol. 83, pp. 115–123, 2016.

[36] T. Zhang and F. J. Oles, "Text categorization based on regularized linear classification methods," *Inf. Retrieval*, vol. 4, pp. 5–31, 2001.

[37] P. Du, Z. Xue, J. Li, and A. J. Plaza, "Learning discriminative sparse representations for hyperspectral image classification," *J. Sel. Topics Signal Process.*, vol. 9, pp. 1089–1104, 2015.

[38] S. Gao, I. W. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition," in *Proc. 11th Eur. Conf. Comput. Vis.*, Sep. 2010, vol. 6314, pp. 1–14.
[39] L. Zhang *et al.*, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.
[40] 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.
[41] 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.
[42] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. 8th Eur. Conf. Comput. Vis.*, 2004, pp. 469–481.
[43] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2683–2695, May 2015.
[44] D. A. Clausi and M. E. Jernigan, "Designing Gabor filters for optimal texture separability," *Pattern Recognit.*, vol. 33, pp. 1835–1849, 2000.
[45] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.
[46] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, 2011, Art. no. 27.
[47] J. Mairal, F. R. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.
[48] J. Mairal, F. R. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2009, pp. 689–696.

Le Gan received the M.S. degree in cartography and geographic information system from Yangtze University, Wuhan, China, in 2015. He is currently working toward the Ph.D. degree in cartography and geographic information system 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, time series analysis and machine learning.

Junshi Xia (S'11–M'16) received the B.S. degree in geographic information systems, 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 with the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology, Grenoble, France, in 2014.

From December 2014 to April 2015, he was a Visiting Scientist with the Department of Geographic Information Sciences, Nanjing University, Nanjing, China. From May 2015 to April 2016, he was a Postdoctoral Research Fellow with the University of Bordeaux, Talence, France. From May 2016, he is the JSPS Postdoctoral 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.

In 2017, he received the first place in the IEEE Geoscience and Remote Sensing Society Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee.

Peijun Du (M'07–SM'12)) received the Ph.D. degree in geodesy and survey engineering 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.

Dr. Du is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He also served as the Co-Chair of the Technical Committee of URBAN 2009, IAPR-PRRS 2012, EORSA 2014, IEOAs 2015, and CCGC 2015, the Co-Chair of the Local Organizing Committee of JURSE 2009, WHISPERS 2012, and EORSA 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 2011, and SPIE European Conference on Image and Signal Processing for Remote Sensing during 2012–2016.

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 Universit de Savoie, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory, Delegation Generale de l'Armement (DGA - French National Defense Department). Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005, an Associate Professor from 2005 to 2007, and is currently a Professor of signal and image processing. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory. His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing. He has been a Visiting Scholar at Stanford University (USA), KTH (Sweden) and NUS (Singapore). Since 2013, he has been an Adjunct Professor with the University of Iceland. Since 2015, he has been a Visiting Professor at the University of California, Los Angeles.

Dr. Chanussot is the Founding President of IEEE Geoscience and Remote Sensing French chapter (2007–2010) which received the 2010 IEEE GRS-S Chapter Excellence Award. He was the co-recipient of the NORSIG 2006 Best Student Paper Award, the IEEE 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 a member of the IEEE Geoscience and Remote Sensing Society AdCom (2009–2010), in charge of membership development. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS). He was the Chair (2009–2011) and the Cochair of the GRS Data Fusion Technical Committee (2005–2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008) and 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 (2005–2007) and for Pattern Recognition (2006–2008). 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 (2011–2015). In 2013, he was a Guest Editor for the Proceedings of the IEEE and in 2014 a Guest Editor for the IEEE SIGNAL PROCESSING MAGAZINE. He is a Fellow of the IEEE and a member of the Institut Universitaire de France (2012–2017).