

Structure-Aware Multikernel Learning for Hyperspectral Image Classification

Chengle Zhou¹, Graduate Student Member, IEEE, Bing Tu¹, Member, IEEE, Nanying Li, Student Member, IEEE, Wei He, Member, IEEE, and Antonio Plaza², Fellow, IEEE

Abstract—Recently, the inclusion of spatial information has drawn increasing attention in hyperspectral image (HSI) applications due to its effectiveness in terms of improving classification accuracy. However, most of the techniques that include such spatial knowledge in the analysis are based on spatial-spectral weak assumptions, i.e., all pixels in a spatial region are assumed to belong to the same class, and close pixels in spectral space are assigned the same label. This article proposes a novel structure-aware multikernel learning (SaMKL) method for HSI classification, which takes into account structural issues in order to effectively overcome the aforementioned weak assumptions and introduce a true multikernel learning process (based on multiple features derived from the original HSI), thus improving the spectral separability of such features. The proposed SaMKL method is composed of the following main steps. First, multiple (i.e., spectral, spatial, and textural) features are extracted from the original HSI based on various filtering operators. Then, a k -peak density approach is designed to define superpixel regions that can properly capture the structural information of HSIs and overcome the aforementioned weak assumptions. Next, three sets of composite kernels are separately constructed to make full use of the spectral, spatial, and textural information. Meanwhile, these three sets of composite kernels are independently incorporated into a support vector machine classifier to obtain their corresponding classification results. Finally, majority voting is used as a simple and effective method to obtain the final classification labels. Experimental results on real

HSI datasets indicate that the SaMKL outperforms other well-known and state-of-the-art classification approaches, in particular, when very limited labeled samples are available *a priori*.

Index Terms—Hyperspectral image (HSI) classification, multikernel learning, structure-aware learning.

I. INTRODUCTION

LAND use and land cover classification using different kinds of earth observation data, e.g., hyperspectral images (HSIs) [1], synthetic aperture radar [2], light detection and ranging [3], and others [4], is a challenging task in geoscience and remote sensing. Since HSIs can provide a wealth of spectral information about the physical properties of the observed materials, it is now widely brought into focus in terms of data processing, such as anomaly detection [5], spectral unmixing [6], dimensionality reduction [7], and image classification [8], and used in plentiful application fields, including national defense [9], precision agriculture [10]–[12], and environment monitoring [13], [14].

HSI classification is a hot topic in the remote sensing community, playing an essential role in many practical application scenarios. Actually, HSI classification methods can be broadly divided into two categories: supervised and unsupervised. Compared with unsupervised methods, supervised ones can obtain higher classification accuracies. In recent years, extensive works have been proposed to develop accurate pixelwise classifiers for the analysis of HSIs, such as random forests [15], [16], neural networks [17]–[19], multinomial logistic regression (MLR) [20], [21], support vector machines (SVMs) [22], and sparse representation (SR)-based classifiers [23]–[25]. Among these methods, the SVM classifier has shown significant performance in terms of classification accuracy due to the following main advantages: 1) it requires relatively few labeled samples to achieve good classification accuracies; and 2) it is robust to the spectral dimensionality of HSIs [22]. Although the SVM is widely recognized as a powerful classification tool, it is unable to overcome the problems caused by the lack of spectral separability of features.

To improve the spectral separability of HSI features, many kernel tricks have been introduced to increase the classification performance in supervised tasks. For instance, Li *et al.* proposed a generalized composite kernel (GCK) method for HSI classification [26] that exploits nonlinear transformations (spectral and contextual) combined with the SVM or the MLR. The designed classifiers are defined as SVM-CK and MLR-GCK,

Manuscript received May 28, 2021; revised August 14, 2021; accepted September 4, 2021. Date of publication September 10, 2021; date of current version October 11, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61977022, in part by the Science Foundation for Distinguished Young Scholars of Hunan Province under Grant 2020JJ2017, in part by the Key Research and Development Program of Hunan Province under Grant 2019SK2012, in part by the Foundation of Department of Water Resources of Hunan Province under Grants XSKJ2021000-12 and XSKJ2021000-13, in part by the Natural Science Foundation of Hunan Province under Grant 2019JJ50211, 2019JJ50212, 2020JJ4340, and 2021JJ40226, in part by the Hunan Provincial Innovation Foundation for Postgraduate under Grant QL20210254, and in part by the Foundation of Education Bureau of Hunan Province under Grants 19B245, 19B237, 20B257, and 20B266. This work was also supported by Junta de Extremadura under Grant GR18060, and by Spanish Ministerio de Ciencia e Innovación through project PID2019-110315RB-I00 (APRISA). (*Corresponding author: Bing Tu.*)

Chengle Zhou, Bing Tu, and Wei He are with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414000, China (e-mail: chengle_zhou@foxmail.com; tubing@hnist.edu.cn; hewei@hnist.edu.cn).

Nanying Li is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: nanying_li@foxmail.com).

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

Digital Object Identifier 10.1109/JSTARS.2021.3111740

which model spatial information according to extended multiattribute profiles (EMPs). Experiments on real HSIs showed that the classification accuracies obtained by SVM-CK and MLR-GCK can outperform those obtained by the SVM classifier alone. In [27], superpixel-based classification via multiple kernels (SC-MK) was introduced to improve classification accuracy by taking advantage of spectral–spatial information. The SC-MK aims to construct three separate kernels to utilize spatial information within and between superpixels (mean operation within each superpixel and weight fusion between different superpixels). Furthermore, a kernel collaborative representation (CR) with Tikhonov regularization was developed in [28] aimed at increasing the separability of spectral information and incorporating spatial information of neighboring locations into kernel space. Ma *et al.* [29] designed a discriminative kernel CR with Tikhonov regularization method to increase the discriminability of different classes in the HSI classification task. In [30], a patch-based low-rank component induced spatial–spectral kernel method (termed LRCISSK) was proposed for HSI classification; the LRCISSK aims to alleviate misclassification phenomenon of the pixels that lie at the boundaries of class.

Recently, spectral–spatial classification methods have been adopted by many researchers in the HSI domain due to their remarkable classification accuracy and efficiency. In fact, a key step in spectral–spatial classification is the introduction of spatial neighborhood information, such as fixed windows and self-adaptive regions.

- 1) Based on the contextual information provided by fixed windows, Chen *et al.* [31] introduced a joint SR method to incorporate the spatial information around each pixel under test in the sparse recovery optimization problem. In [32], a maximum likelihood estimation (MLE)-based joint sparse representation (MLEJSR) was developed to replace the traditional quadratic loss function with an MLE-like estimator able to measure the joint approximation error. Specifically, the theoretical and experimental results demonstrated the effectiveness of the MLEJSR method, particularly in the presence of high noise. Additionally, the deep learning (DL)-based spectral–spatial classification method also immits the spectral information of the connected pixels around the pixel when performing the classification task. For instance, Hong *et al.* [33] developed a minibatch graph convolutional network, which is capable of inferring out-of-sample data without retraining networks and improving classification performance. Xu *et al.* [34] proposed a random patches network DL method for HSI classification, which directly regards the random patches taken from the HSI as the convolution kernels without any training. He *et al.* [35] developed a multiscale covariance map model for HSI classification, which is aimed at exploiting both spatial and spectral information in local patches to improve the classification of HSI using convolutional neural networks.
- 2) Based on the contextual information provided by self-adaptive regions (which can extract HSI spatial information more effectively because of the advantages provided

by shape adaptation), Zhou *et al.* [36] designed a spatial peak-aware collaborative representation (SPaCR) method for HSI classification; the SPaCR introduces spectral–spatial information among superpixel clusters into regularization terms to construct a new CR-based closed-form solution. Furthermore, Jia *et al.* [37] designed a cascade superpixel-based regularized Gabor feature fusion method for HSI classification, where two types of Gabor features (magnitude and phase) were jointly integrated and exploited in terms of the discriminative information they convey. In [38], a classification framework that combined superpixel-level SR and multitask learning was developed, which utilized prior class-level sparsity for multiple-feature fusion and increasing correlation and distinctiveness of HSI pixels in a local spatial region.

Furthermore, spectral–spatial joint feature extraction has also been widely researched in the remote sensing community. For example, Tu *et al.* [39] proposed a novel feature extraction method via sharing of 3-D block characteristics (3-D-BCS) for HSI classification. Experiments performed on several real HSI datasets with limited training samples show that the 3-D-BCS method outperforms other types of classification method. In [40], a novel HSI feature extraction method based on joint adaptive structure density was developed to make full use of texture features and effectively utilize pixel-level structural feature. Moreover, the spectral–spatial joint feature extraction method based on morphological attributes and spatial filtering has been successfully applied to HSI data processing in recent years. Morphological attribute-based feature extraction methods include morphological profiles [41], its extended versions [42], [43], attribute profiles (APs) [44], and extinction profiles (EPs) [45], [46]. Among the aforementioned methods, EPs have drawn significant attention due to their following key advantages: 1) they can simultaneously preserve the geometrical characteristics of the raw HSI data while removing unimportant details; and 2) they can deliver better recognition performance than that of traditional spatial–spectral AP-based [44] feature extraction methods. Spatial-filtering-based feature extraction methods include recursive filtering (RF) [47], adaptive total variation (ATV) [48], guidance filtering (GuF) [49], and 2-D Gabor filtering [50]. Among the aforementioned methods, Gabor filtering has been successfully applied to HSI classification as a powerful spatial–spectral feature extraction tool [51]–[53]. This is mainly due to the fact that the characteristics of Gabor wavelets (particularly for frequency and orientation representations) are similar to those of the human visual system [54].

Although these spectral–spatial classification and feature extraction techniques have achieved significant interpretation effects in HSI classification, the spectral–spatial weak assumptions of the local spatial information in the HSIs is still a topic worthy of in-depth study. Additionally, while alleviating the spectral–spatial weak assumptions, it is also a challenging task to use multikernel technology to learn various HSI features to enhance the spectral separability between pixels. In this article, we introduce a new structure-aware multikernel learning (SaMKL) method for HSI classification, which exploits structure-aware design technology to effectively overcome weak assumptions in

spectral–spatial classification and introduces a new kind of multikernel learning approach to exploit multiple features extracted from the original HSI data to improve the spectral separability. The proposed SaMKL method is composed of the following main steps. First, multiple features (i.e., spectral, spatial, and textural) are extracted from the original HSI based on different operators (i.e., average fusion, EPs, and 2-D Gabor). Then, a k -nearest-neighbor-based local density peak approach (named k -peak density) is applied to superpixel regions to define the structural information of the HSI and overcome the aforementioned weak assumptions. Next, three sets of composite kernels are separately constructed to fully exploit the spectral, spatial, and textural information contained in the HSI. Meanwhile, the three sets of composite kernels are, respectively, fed to an SVM classifier to obtain the corresponding classification results. Finally, majority voting (as a simple and effective method) is employed to obtain the final classification results. The main contributions of the proposed approach can be summarized as follows.

- 1) A new structure-aware technology (driven by superpixels) is proposed that can effectively overcome spectral–spatial weak assumptions within each superpixel block according to the density distribution of pixels in a local view. It was found that the nonhomogeneous pixels within the superpixel present a lower density, which means that the superpixel is not an absolute homogeneous area from the experimental point of view.
- 2) While alleviating the spectral–spatial weak assumptions, we utilize spectral fusion, morphological attributes, and spatial filtering techniques to extract the spectral, spatial, and textural features of the original HSI data. We found that all these features are indispensable, and that the lack of any of them leads to suboptimal classification accuracies. As a result, the higher classification accuracies obtained by our framework result from the effective integration of these three kinds of features.
- 3) A simple and effective multikernel learning method is designed to achieve our newly proposed classification framework. Above all, any dual-kernel-based SVM cannot achieve the classification accuracies obtained by our framework. It can be, thus, inferred that the kernels constructed by the spectral, spatial, and texture features have the distinct advantage of complementary information.

The rest of this article is organized as follows. Section II briefly reviews some related works. Section III describes the proposed SaMKL method in detail. Section IV presents the experimental results conducted on real HSI datasets. Section V concludes this article with some remarks and hints at plausible future research lines.

II. RELATED WORK

This section briefly describes some related algorithms, i.e., the extinction filters (EFs), 2-D Gabor filters (GFs), density peak clustering (DPC), and SVM using composite kernels. These algorithms play a relevant role in the design of the proposed method.

A. Extinction Filters

The EFs are extrema-oriented connected idempotent filters¹ that retain the relevant leaf node of the connected component corresponding to a binary or grayscale image. Assume that $\text{Max}(\mathbf{I}_g) = \{M_1, M_2, \dots, M_N\}$ is denoted as the set of the minimum connected components of the image \mathbf{I}_g , where N refers to the number of minimum connected components. $M_{i_a} (i_a = 1, 2, \dots, N)$ represents the regional maximum that is used to estimate the extinction value ε_{i_a} with regard to the increasing attribute being analyzed. For the input image \mathbf{I}_g , the EFs preserve the N maxima with the highest extinction values [i.e., $\text{EF}^N(\mathbf{I}_g)$]. This can be defined as follows:

$$\text{EF}^N(\mathbf{I}_g) = R_g^\delta(\mathbf{I}_g) \quad (1)$$

where $R_g^\delta(\mathbf{I}_g)$ refers to the reconstruction by dilation [55] of the mask image $\mathbf{G} = \max_{i_e=1}^n(M'_{i_e})$ (here, max is the pixelwise maximum operation) stemming from the marker image \mathbf{I}_g . M'_1 is the maximum with the highest extinction value and M'_2 has the second highest value.

B. 2-D Gabor Filters

2-D GFs² offer a powerful tool for extracting image texture features in the remote sensing domain [56]–[58]. A Gaussian function and a sinusoidal plane wave comprise 2-D GFs, which can effectively capture texture and orientation information from a set of Gabor-filtered images. The generating function of the 2-D GFs can be expressed as follows:

$$\Phi_{u,v}(x, y) = \frac{c^2}{\pi\zeta\eta} \exp(-(\alpha_o^2 x'^2 + \beta_o^2 y'^2)) \exp(j2\pi zx') \quad (2)$$

$$\begin{cases} x' = (x - \frac{m_o+1}{2}) \cos \theta + (y - \frac{n_o+1}{2}) \sin \theta \\ y' = -(x - \frac{m_o+1}{2}) \sin \theta + (y - \frac{n_o+1}{2}) \cos \theta \end{cases} \quad (3)$$

where c refers to the distribution coefficient of the frequency domain in the generating function. $\zeta = z/\alpha_o$ and $\eta = z/\beta_o$ are defined as the constraint conditions keeping the ratio between the frequency and sharpness, in which α_o is the sharpness of the Gaussian function along the major axis parallel to the wave, and β_o is the sharpness of the Gaussian function along the short axis perpendicular to the wave. θ represents the rotation angle of the Gaussian function. It is should be noted that α_o and β_o are assumed to be equal, and $\zeta = \eta = \sqrt{2}$ was adopted in most previous studies [59]. Besides, m and n represent the size of the GF, which is usually set to be equal to an odd number c_d , which refers to the size of the GF, i.e., $c_d = m_o = n_o$.

C. Density Peak Clustering

The DPC algorithm³ is based on two salient features of clusters: 1) the cluster center is surrounded by most high-density data samples in a certain range, and its density is the highest with respect to these data samples; and 2) the distance between

¹[Online]. Available: <http://pedram-ghamisi.com>

²[Online]. Available: <https://github.com/mhaghagh/gabor>

³[Online]. Available: https://people.sissa.it/~laio/Research/Res_clustering.php

cluster centers of interclass data samples is larger than the distance among intraclass data samples [60]. Suppose that the local density of data samples is presented as ρ , and the distance between data samples is denoted as δ . Then, the DPC can be summarized as follows:

$$d_{co} = \|s_c - s_o\|_2^2 \quad (4)$$

where s_c and s_o are the data samples that belong to a set $\mathbb{D} = \{s_c\}_{c=1}^n$, n refers to the number of data samples, and d_{ij} is the Euclidean distance between data samples s_c and s_o . According to the above distance information, the kernel-based local density ρ_c of data sample s_c can be obtained using one of the following two strategies:

$$\rho_c = \begin{cases} \sum_o \chi(d_{co} - d_c), & \text{cutoff kernel} \\ \sum_o e^{-\left(\frac{d_{ij}}{d_c}\right)^2}, & \text{Gaussian kernel} \end{cases} \quad (5)$$

where d_c is the cutoff distance measuring the radius of the search region. Unlike the cutoff kernel, the Gaussian kernel can decrease the negative impact of the statistical errors introduced by the limited availability of samples. Actually, the Gaussian-kernel-based local density has been broadly used for HSI data interpretation [61]–[63]. It is worth noting that ρ_c reflects the number of data samples that are closer than d_c to the data sample r_c .

After computing ρ_c , another important indicator (δ_c) can be calculated as follows:

$$\delta_c = \begin{cases} \max_o(d_{co}), & \text{if } \rho_c = \max(\rho) \\ \min_{o:\rho_o>\rho_c}(d_{co}), & \text{otherwise.} \end{cases} \quad (6)$$

Specifically, if the data sample s_c has the highest local density, then δ_c is set to the maximum distance from the data sample s_c to any other point. Otherwise, δ_c is defined as the minimum distance from the data sample s_c to any other data sample with higher local density. Finally, a score (γ_c) used to determine cluster centers is given as follows:

$$\gamma_c = \rho_c \times \delta_c \quad (7)$$

where γ_c considers ρ_c and δ_c jointly. It should be noted that the larger the score of γ_c , the more likely that the associated point (s_c) is a cluster center. Therefore, a hint for choosing the number of centers is given by the solution of γ_c , ordered in decreasing order [60].

D. SVM With Composite Kernel

The SVM, a supervised learning model, has been widely used in HSI classification. The algorithm aims to determine the label of a test sample by establishing a suitable hyperplane through the data. To address linearly nonseparable problems involving HSI pixels, a kernel trick can be employed to map pixels into a high-dimensional feature space [22]. Assume that $\{\mathbf{r}_i, c_i\}_{i=1}^n \in (\mathbb{R}^{m \times n} \times \mathbf{c})$ is a training set, where \mathbf{r}_i refers to a training sample and c_i is the corresponding class label. m represents the number of spectral dimensions, and n denotes the number of training

samples. The SVM aims to solve the following problem:

$$\begin{aligned} \min_{\omega, \mu_i, d} & \left\{ \frac{1}{2} \|\omega\|^2 + Q \sum_i \mu_i \right\} \text{ subject to} \\ & c_i (\langle \phi(\mathbf{r}_i), \omega \rangle + d) \geq 1 - \mu_i \quad \forall i = 1, 2, \dots, n \\ & \mu_i \geq 0 \quad \forall i = 1, 2, \dots, n \end{aligned} \quad (8)$$

where ω and d construct the decision rule (classifier) in the feature space, μ_i refers to slack variables addressing the nonseparability of data, and Q is a regularization item controlling the generalization ability of the classifier. $\phi(\cdot)$ is a mapping function that transforms the low-dimensional input data, $\mathbf{r}_i \in \mathbb{R}^m$, into a high-dimensional space $\phi(\mathbf{r}_i) \in \mathbb{R}^{m^*}$ ($m^* > m$). To solve computational problems in high-dimensional spaces, the inner product is employed in SVM training. A kernel trick K can be defined as follows:

$$K(\mathbf{r}_i, \mathbf{r}_j) = \langle \phi(\mathbf{r}_i), \phi(\mathbf{r}_j) \rangle. \quad (9)$$

Then, a nonlinear SVM classifier can be designed by adopting a kernel function, considering the mapping $\phi(\cdot)$ explicitly. The radial basis function (RBF) kernel is a widely used one that can be introduced into the SVM to construct a nonlinear classifier. The kernel is computed by

$$K(\mathbf{r}_i, \mathbf{r}_j) = \exp(-\|\mathbf{r}_i - \mathbf{r}_j\|^2/2\sigma^2). \quad (10)$$

Next, the decision rule for any test pixel can be obtained by solving a dual Lagrangian problem and by incorporating (9) into (8) as follows:

$$f(\mathbf{r}) = \sum_{i=1}^n c_i \vartheta_i K(\mathbf{r}_i, \mathbf{r}) + d \quad (11)$$

where ϑ_i are the Lagrange multipliers in (8), which can be estimated by quadratic programming methods [64]. Furthermore, to introduce the spatial information contained in HSIs, Camps-Valls *et al.* [65] designed an SVM method with a composite kernel strategy,⁴ which can also be called a one-against-one multiple classifier SVM method for HSI classification. Assume that the spectral information and the spatial features from the original training samples are, respectively, denoted as $\{\mathbf{r}_1^s, \mathbf{r}_2^s, \dots, \mathbf{r}_n^s\}$ and $\{\mathbf{r}_1^w, \mathbf{r}_2^w, \dots, \mathbf{r}_n^w\}$. Then, the corresponding two types of spectral and spatial kernels can be denoted as $K_s(\mathbf{r}_i^s, \mathbf{r}_j^s)$ and $K_w(\mathbf{r}_i^w, \mathbf{r}_j^w)$, respectively. A composite kernel can be constructed based on a weighted average operation as follows:

$$K_{cw}(\mathbf{r}_i^s, \mathbf{r}_j^s) = \mu_s K_s(\mathbf{r}_i^s, \mathbf{r}_j^s) + \mu_w K_w(\mathbf{r}_i^w, \mathbf{r}_j^w) \quad (12)$$

where μ_s and μ_w , respectively, refer to the weights for the spectral kernel $K_s(\mathbf{r}_i^s, \mathbf{r}_j^s)$ and the spatial kernel $K_w(\mathbf{r}_i^w, \mathbf{r}_j^w)$, and $\mu_s + \mu_w = 1$. Finally, the composite kernel is embedded into (9) to generate a new decision rule for classification purposes. Due to its joint consideration of the spatial-spectral information contained in the HSI, composite-kernel-based methods can achieve very promising classification performances when compared to single-kernel-based methods.

⁴[Online]. Available: <http://www.disi.unitn.it/rslab>

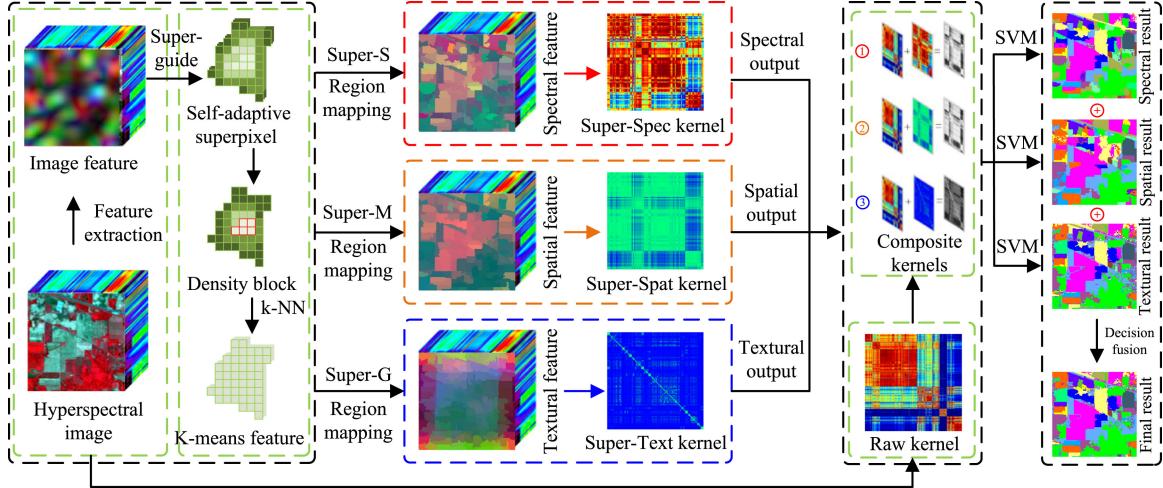


Fig. 1. Overview of the proposed SaMKL method for HSI classification. Super-S, Super-M, and Super-G refer to HSI raw spectral, morphological attribute, and 2-D Gabor textural features based on superpixel guidance, respectively. Super-Spec, Super-Spat, and Super-Text represent superpixel spectral, spatial, and textural kernels, respectively.

III. PROPOSED METHOD

As shown in Fig. 2, the proposed SaMKL method⁵ is composed of three main steps: 1) multiple feature extraction; 2) structure-aware modeling with k -peak density; and 3) multiple kernel learning. A pseudocode of our newly developed SaMKL is given in Algorithm 1. In the following, the implementation details of these three main steps are fully described.

A. Multiple Feature Extraction

In this section, we elaborate on the extraction of various types of features from the original HSI data by the proposed SaMKL method.

1) *Spectral Feature Extraction*: Given an HSI $\mathbf{I} \in \mathbb{R}^m$, in which m stands for the number of dimension of the \mathbf{I} , its dimensions are reduced with an averaging method to obtain a dimension-reduced HSI \mathbf{I}_{BGF} . First, the original HSI is partitioned into M subsets of adjacent bands. M is the number of band subsets. The t th ($t = 1, 2, \dots, M$) subset, \mathbf{H}^t , is given by

$$\mathbf{H}^t = \begin{cases} \left(\mathbf{I}^{(t-1)\lfloor \frac{m}{M} \rfloor + 1}, \dots, \mathbf{I}^{t\lfloor \frac{m}{M} \rfloor} \right), & t\lfloor \frac{m}{M} \rfloor \leq m \\ \left(\mathbf{I}^{(t-1)\lfloor \frac{m}{M} \rfloor + 1}, \dots, \mathbf{I}^m \right), & t\lfloor \frac{m}{M} \rfloor > m \end{cases} \quad (13)$$

where $\mathbf{I} = \{\mathbf{I}^1, \mathbf{I}^2, \dots, \mathbf{I}^m\}$ refers to an HSI with m spectral bands and $\lfloor \frac{m}{M} \rfloor$ is the value of the largest integer not greater than m/M . M is set to 25 in our work; if the SaMKL method is migrated to the new HSI dataset, we recommend setting the value of M to between 25 and 30. Generally, adjacent spectral bands in HSI data are strongly correlated (and often redundant). Once band subsets \mathbf{H}^t are obtained, a dimensionally reduced HSI \mathbf{I}_{BGF} can be obtained by performing the mean operation on

the band subsets \mathbf{H}^t as follows:

$$\mathbf{I}_{\text{BGF}}^t = \frac{1}{n_t} \sum_{i_b=1}^{n_t} \mathbf{H}_{i_b}^t \quad (14)$$

where $\mathbf{I}_{\text{BGF}} = \{\mathbf{I}_{\text{BGF}}^1, \mathbf{I}_{\text{BGF}}^2, \dots, \mathbf{I}_{\text{BGF}}^M\}$, $\mathbf{H}_{i_b}^t$ refers to the i_b th band of the t th subset of the original HSI bands, and n_t represents the total number of bands in the t th subset. Note that the band grouping fusion method, which can effectively reduce the band dimensionality without degenerating the physical properties of the original data, is only adopted as a simple spectral feature extraction tool. In case other more effective spectral feature extraction approaches were incorporated by the proposed framework, the performance of the SaMKL may be even better.

2) *Spatial Feature Extraction*: First of all, to extract HSI spatial information with morphological characteristics, we use principal component analysis (PCA) to achieve the most informative principal components (PCs), and such PCs are used as base images to execute EPs. In fact, the EPs are composed of a sequence of thinning and thickening transformations constructed by a set of EFs with a progressively stricter criterion. The EPs of each PC C_q can be extracted as follows:

$$\text{EP}(\mathbf{C}_q) = \left\{ \underbrace{\varphi^{P_{\lambda_K}}(\mathbf{C}_q), \varphi^{P_{\lambda_{K-1}}}(\mathbf{C}_q), \dots, \varphi^{P_{\lambda_1}}(\mathbf{C}_q)}_{\text{thickening profiles}}, \mathbf{C}_q, \underbrace{\gamma^{P_{\lambda_K}}(\mathbf{C}_q), \gamma^{P_{\lambda_{K-1}}}(\mathbf{C}_q), \dots, \gamma^{P_{\lambda_1}}(\mathbf{C}_q)}_{\text{thinning profiles}} \right\} \quad (15)$$

where $P_{\lambda_K} : \{P_{\lambda_q}\}$ ($q = 1, 2, \dots, K$) is a set of K ordered predicates (i.e., $P_{\lambda_q} \subseteq P_{\lambda_K}$, $q \leq k$) [66]. The result of applying an EF on each \mathbf{C}_q can be expressed as follows: $\mathbf{EP}_1 = \text{EP}(\mathbf{C}_1)$, $\mathbf{EP}_2 = \text{EP}(\mathbf{C}_2)$, and $\mathbf{EP}_3 = \text{EP}(\mathbf{C}_3)$ (for the first three PCs). Due to the fact that the first few PCs generally contain most of the spatial information of the original HSI, the obtained PCs can be

⁵[Online]. Available: <https://github.com/changle-zhou>

used to characterize the HSI and provide distinctive information in terms of the corresponding EPs. Based on this idea, two fusion rules are considered in this article for constructing the EPs:

$$\mathbf{EP} = \begin{cases} \sum_{q=1}^{n_q} \tau_q \mathbf{EP}_q, & \text{weighting fusion} \\ \sum_{\text{stack}} \mathbf{EP}_q, & \text{stacking fusion} \end{cases} \quad (16)$$

where τ_q is the weight coefficient of the q th EP and n_q refers to the number of PCs. In this work, the values of τ_q and n_q are set to 1 and 3, respectively, because it is considered that the structural information contained in the first three PCs of the HSI image is equivalent. Furthermore, we will discuss the pros and cons of these two fusion methods in detail in the experimental part of this article.

3) *Texture Feature Extraction*: First, a set of 2-D GFs with various frequencies and orientations are modeled as follows:

$$\begin{cases} F = \frac{F_{\max}}{\sqrt{2}}, & u = 0, 1, \dots, U - 1 \\ \theta = \frac{v}{8}\pi, & v = 0, 1, \dots, V - 1 \end{cases} \quad (17)$$

where F_{\max} stands for the highest peak frequency of the 2-D Gabor function. F and θ represent the various frequencies and the orientation angles of the 2-D GFs, respectively. U is the number of scales of the 2-D GFs, which is used to control the frequencies. V refers to the number of orientation of 2-D GFs controlling various orientations.

Then, a region \mathbf{R}_w of a fixed window size in the first three PCs C_q is determined based on neighborhood rules for pixels in the spatial domain. Next, the Gabor features $\mathbf{G}_{u,v}$ of the region \mathbf{R}_w are extracted by, respectively, convolving the first three PCs C_q with a family of GFs $\Phi_{u,v}(c_x, c_y)$ as follows:

$$\mathbf{G}_{u,v}(c_x, c_y) = |\mathbf{R}_w * \Phi_{u,v}(c_x, c_y)| \quad (18)$$

where $\mathbf{G}_{u,v}(c_x, c_y)$ refer to the Gabor features obtained as the result of the convolution of 2-D GFs at scale u and orientation v . Here, $|\cdot|$ denotes the magnitude of the convolution. Finally, the textural feature $\mathbf{G}_{u,v}$ can be captured from the HSI after the fixed sliding window traverses the whole image. Specifically, the eight orientations v of 0–180 for the proposed SaMKL method are considered, $[0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}]$. The scale u and the size U of GF are set as 5 and 55, respectively, since they can achieve good balance between high classification accuracies and an acceptable computing burden.

B. Structure-Aware Modeling With k -Peak Density

To conduct this step, we initially perform an entropy rate superpixel (ERS)-based [67] oversegmentation of the HSI using the first few PCs (normally, the first three PCs are enough as they carry out most of the variance information contained in the original HSI). This results in detailed structural information about the scene. However, the ERS is an explicit segmentation algorithm based on image color and brightness. Thus, it is difficult to avoid the presence of spectral–spatial weak assumptions in the resulting (local) superpixel blocks. To address this issue, we exploit the concept of DPC to redefine the superpixel structural information. The details are as follows.

Let us suppose that we have a superpixel region $\mathbf{B}_l = \{\mathbf{r}_l^1, \mathbf{r}_l^2, \dots, \mathbf{r}_l^{n_l}\}$, where n_l is the number of pixels within a superpixel block. The spectral angle can be employed as a measure to obtain information about the distance among different pixels inside each superpixel block as follows:

$$\mathbf{SA}(\mathbf{r}_l^{i_u}, \mathbf{r}_l^{i_v}) = \arccos \left(\frac{\langle \mathbf{r}_l^{i_u}, \mathbf{r}_l^{i_v} \rangle}{\|\mathbf{r}_l^{i_u}\| \|\mathbf{r}_l^{i_v}\|} \right). \quad (19)$$

According to the distance information $\mathbf{SA}(\mathbf{r}_l^{i_u}, \mathbf{r}_l^{i_v})$, the local density ρ^{i_u} of pixel $\mathbf{r}_l^{i_u}$ can be defined as follows:

$$\rho^{i_u} = \sum_{i_u} \exp \left\{ - \left(\frac{\mathbf{SA}(\mathbf{r}_l^{i_u}, \mathbf{r}_l^{i_v})}{d_c} \right)^2 \right\} \quad (20)$$

where d_c stands for a cutoff distance, which is achieved by

$$d_c = \mathcal{M}(\zeta) \quad \text{s.t.} \quad \zeta = \left\langle \eta \cdot \frac{n_l(n_l - 1)}{100} \right\rangle \quad (21)$$

in which \mathcal{M} is a matrix that sorts the nonzero elements in the upper triangular matrix of $\mathbf{SA}(\mathbf{r}_l^{i_u}, \mathbf{r}_l^{i_v})$ from the smallest to the largest, parameter η adjusts the size of the cutoff distance, and $\langle \cdot \rangle$ refers to the round operation.

Next, pixels with high density within each superpixel block are selected. The formula is expressed as follows:

$$\Omega = \arg \max_{i_u \in k_0} \rho^{i_u} \quad (22)$$

where k_0 represents the range of pixels with higher density within a superpixel block. In addition, a mean operation can be applied on the spectral signatures of pixels within each superpixel as follows:

$$\bar{\mathbf{r}}_l = \frac{1}{n_o} \sum_{k=1}^{n_o} \mathbf{r}_l^k \quad (23)$$

where n_o represents the number of pixels with relatively high density for each superpixel block. Finally, the superpixel region can be redefined as $\bar{\mathbf{B}}_l = \{\bar{\mathbf{r}}_l^1, \bar{\mathbf{r}}_l^2, \dots, \bar{\mathbf{r}}_l^{n_l}\}$ and $\mathbf{F} = \{\bar{\mathbf{B}}_1, \bar{\mathbf{B}}_2, \dots, \bar{\mathbf{B}}_{S_n}\}$ is denoted as a feature of the original HSI. Here, S_n is the number of superpixel blocks and $l \in (1, 2, \dots, S_n)$. In order to clearly express the procedure of the k -peak density approach, Fig. 2 gives an outline of the designed method. Moreover, we applied the operations described in Section II-B to the aforementioned spectral (I_{BGF}), spatial (\mathbf{EP}), and textural ($\mathbf{G}_{u,v}$) features to obtain three new kinds of features from the HSI. The new spectral, spatial, and textural features are denoted as \mathbf{F}_{spe} , \mathbf{F}_{ext} , and \mathbf{F}_{gab} , respectively.

C. Multiple Kernel Learning

Let $\{\hat{\mathbf{r}}_1^O, \hat{\mathbf{r}}_2^O, \dots, \hat{\mathbf{r}}_N^O\}$ be a set of pixels randomly selected from the original HSI \mathbf{I} to be used as training samples. Here, N represents the number of training samples. Then, the corresponding \mathbf{F}_{spe} , \mathbf{F}_{ext} , and \mathbf{F}_{gab} for these samples are also selected for the training process and defined as $\{\hat{\mathbf{r}}_1^{\text{spe}}, \hat{\mathbf{r}}_2^{\text{spe}}, \dots, \hat{\mathbf{r}}_N^{\text{spe}}\}$, $\{\hat{\mathbf{r}}_1^{\text{ext}}, \hat{\mathbf{r}}_2^{\text{ext}}, \dots, \hat{\mathbf{r}}_N^{\text{ext}}\}$, and $\{\hat{\mathbf{r}}_1^{\text{gab}}, \hat{\mathbf{r}}_2^{\text{gab}}, \dots, \hat{\mathbf{r}}_N^{\text{gab}}\}$, respectively. After that, various kernels are obtained for the aforementioned

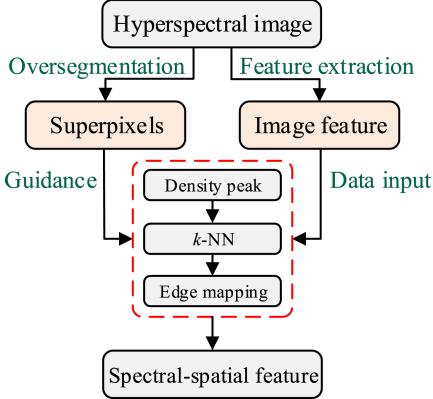


Fig. 2. Illustration of k -peak density on the proposed SaMKL method, which aims to overcome spectral–spatial weak assumptions. The k -NN refers to k nearest neighbor mechanism in machine learning.

four types of training samples using the RBF kernel function as follows:

$$K_O^{\text{train}}(\hat{r}_i^O, \hat{r}_j^O) = \exp\left(-\|\hat{r}_i^O - \hat{r}_j^O\|^2 2\sigma^2\right) \quad (24)$$

$$K_{\text{spe}}^{\text{train}}(\hat{r}_i^{\text{spe}}, \hat{r}_j^{\text{spe}}) = \exp\left(-\|\hat{r}_i^{\text{spe}} - \hat{r}_j^{\text{spe}}\|^2 2\sigma^2\right) \quad (25)$$

$$K_{\text{ext}}^{\text{train}}(\hat{r}_i^{\text{ext}}, \hat{r}_j^{\text{ext}}) = \exp\left(-\|\hat{r}_i^{\text{ext}} - \hat{r}_j^{\text{ext}}\|^2 2\sigma^2\right) \quad (26)$$

$$K_{\text{gab}}^{\text{train}}(\hat{r}_i^{\text{gab}}, \hat{r}_j^{\text{gab}}) = \exp\left(-\|\hat{r}_i^{\text{gab}} - \hat{r}_j^{\text{gab}}\|^2 2\sigma^2\right). \quad (27)$$

Next, three sets of composite kernel are separately generated as follows:

$$K_{CW1} = \alpha K_O^{\text{train}} + (1 - \alpha) K_{\text{spe}}^{\text{train}} \quad (28)$$

$$K_{CW2} = \beta K_O^{\text{train}} + (1 - \beta) K_{\text{ext}}^{\text{train}} \quad (29)$$

$$K_{CW3} = \theta K_O^{\text{train}} + (1 - \theta) K_{\text{gab}}^{\text{train}} \quad (30)$$

where α , β , and θ are the free parameters. Once the three groups of composite kernels are obtained, different classification rules can be constructed by, respectively, feeding the composite kernels to the SVM classifier. Finally, majority voting is used to obtain the final classification results generated by the SVM, trained using different composite kernels.

IV. EXPERIMENTAL RESULTS

A. Experimental Setting

1) *Description of Datasets*: This section describes our four real HSI experimental datasets (i.e., Indian Pines, University of Pavia, Center of Pavia, and Washington DC), which include different kinds of landscapes to prove the practicability of the proposed SaMKL method in different scenarios.

1) *Indian Pines*: The Indian Pines image was captured in 1992 by the Airborne Visible Infrared Imaging Spectrometer over the Indian Pines area in Northwest Indiana. The image has a size of $145 \times 145 \times 220$ pixels, a spatial resolution of 20 m, and a wavelength range of 0.4–2.5 μm . Before our HSI classification experiments, the spectral

Algorithm 1: SaMKL.

Inputs: 1) Original HSI data $I \in \mathbb{R}^m$; 2) Position indexes of the training samples; 3) Parameter settings, i.e., S_n , k_0 , α , β , and θ ; **Outputs:** Classification results obtained by a decision fusion-based SVM classifier using composite kernels.

- 1: Spectral feature extraction using (13) and (14).
- 2: Spatial feature extraction using (15) and (16).
- 3: Textural feature extraction using (17).
- 4: ERS-based oversegmentation.
- 5: **For** each $l = 1 : S_n$ **do**.
- 6: Compute spectral angle $\text{SA}(r_l^u, r_l^v)$.
- 7: Obtain local density ρ^u and cutoff distance d_c .
- 8: Select pixels with high density
 $\hat{u} = \arg \max_{u \in k_0} \rho^u$.
- 9: Perform mean operation on each superpixel according to (21).
- 10: **End For**
- 11: $K_{CW1} = \alpha K_O^{\text{train}} + (1 - \alpha) K_{\text{spe}}^{\text{train}}$
- 12: $K_{CW2} = \beta K_O^{\text{train}} + (1 - \beta) K_{\text{ext}}^{\text{train}}$
- 13: $K_{CW3} = \theta K_O^{\text{train}} + (1 - \theta) K_{\text{gab}}^{\text{train}}$
- 14: Perform majority voting-based decision fusion to obtain the final classification results of the SVM using composite kernels.

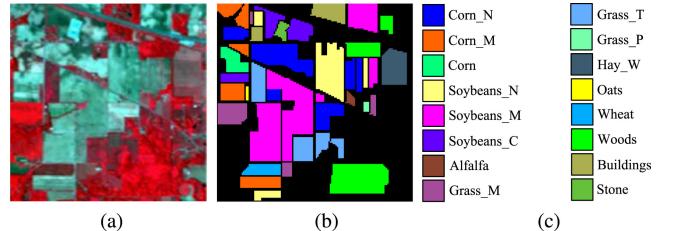


Fig. 3. Indian Pines test image. (a) Three-band false-color composite image. (b) Ground truth. (c) Color coding.

bands covering the water absorption region in the original HSI data (104–108, 150–163, and 220) were removed, so that the number of spectral bands is reduced from 220 to 200. The Indian Pines image includes agricultural fields, forest, and many different kinds of vegetation species. Fig. 3(a)–(c), respectively, show a false-color composite, the corresponding labeled data, and the class labels for the Indian Pines scene.

2) *University of Pavia*: The University of Pavia image was collected by the Reflective Optics System Imaging Spectrometer (ROSIS-3) over an urban area centered at the University of Pavia, Italy. The image has a spatial resolution of 1.3 m and a wavelength range of 0.43–0.86 μm . The image size is 610×340 pixels and contains 115 spectral bands. Before our experiments, the 12 noisiest bands were removed. Fig. 4(a)–(c), respectively, show a false-color composite, the corresponding labeled data, and the class labels for the University of Pavia scene.

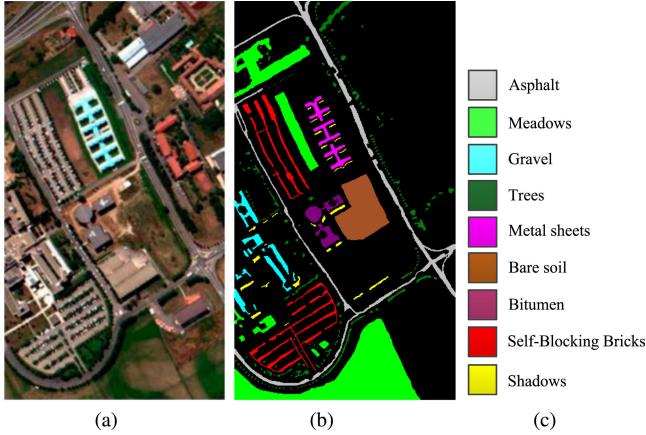


Fig. 4. University of Pavia test image. (a) Three-band false-color composite image. (b) Ground truth. (c) Color coding.

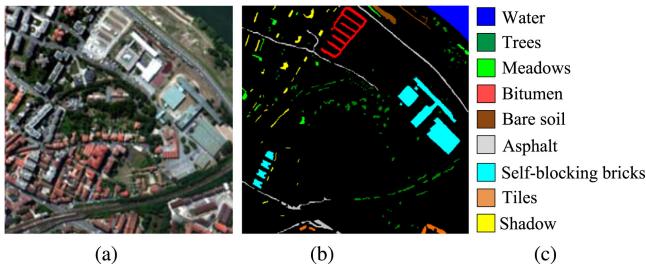


Fig. 5. Center of Pavia test image. (a) Three-band false-color composite image. (b) Ground truth. (c) Color coding.

3) *Center of Pavia*: The Center of Pavia image was also collected by ROSIS-3. After preprocessing, the image comprises 102 bands. The size of the image is 1096×715 pixels, and it has nine different classes. It is worth mentioning that the local perspective of the original scene with a size of 492×492 is used in our experiment, and it also contains nine different classes. Fig. 5(a)–(c), respectively, show a false-color composite, the corresponding labeled data, and the class labels of the Center of Pavia scene.

4) *Washington DC*: The Washington DC image was acquired by the Hyperspectral Digital Image Collection Experiment. It contains a total of 210 bands ranging from 0.2 to $2.4 \mu\text{m}$. The image size is $208 \times 307 \times 210$ pixels. Fig. 6(a)–(c), respectively, show a false-color composite, the corresponding labeled data, and the class labels of the Washington DC scene.

2) *Benchmark Classifier*: The SVM is a classic pixel-by-pixel classifier and is employed in this work as a baseline for evaluating the performance of the proposed SaMKL. In addition, fivefold cross validation is used to determine the parameters of the SVM. As shown in Table I, random training samples are selected from each class in our experiments. In order to make the results statistically significant, the average results after ten repeated Monte Carlo experiments are reported.

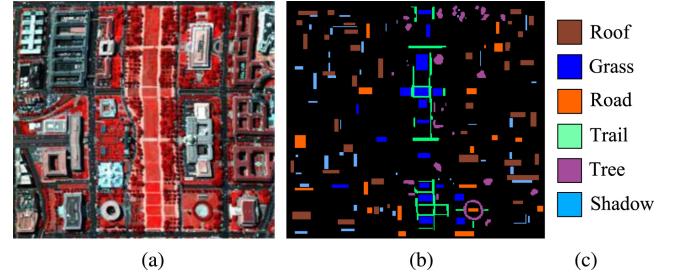


Fig. 6. Washington DC test image. (a) Three-band false-color composite image. (b) Ground truth. (c) Color coding.

3) *Quality Metrics*: In this experiment, three commonly used accuracy evaluation metrics: overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa) are adopted. OA represents the probability that the classification result is consistent with the reference classification result. AA is the mean value of the percentage of each category being correctly classified. The Kappa coefficient is another commonly used index for evaluating classification accuracy. Its advantage is that it considers the impact of uncertainty on the classification results when estimating the recognition accuracy.

B. Parameter Analysis

In this section, we conduct experiments on the aforementioned HSI datasets to analyze the sensitivity of the two most important parameters of the proposed SaMKL method, i.e., S_n (the number of superpixel blocks) and k_0 (the number of pixels with high density within a superpixel block). The numbers of randomly selected training samples from each class of the various HSI datasets are given in Table I. It can be clearly seen from Fig. 7 that the optimal classification accuracies (94.23%, 94.01%, 92.58%, and 92.58%, respectively) are achieved when S_n is set to 150, 300, 120, and 1100 on the four test images. When S_n is larger than the optimal value, the number of superpixels in the image will be too high, and the size of each superpixel will become smaller. Therefore, the spatial information will not be fully exploited in this case. When S_n is lower than the optimal value, the number of superpixels in the image will be too low, and the size of each superpixel will become larger. Therefore, a superpixel may contain different classes with the same spectral information, resulting in a decrease in classification accuracy. Meanwhile, it can be clearly seen that, when S_n is set to an optimal value, the corresponding values of k_0 in the four datasets are 70%, 50%, 30%, and 30%, respectively. This not only means that spatial information is fully exploited, but also that interference from different classes with the same spectral information is eliminated to a great extent.

A second experiment is also performed using these four HSI datasets to test the influence of three kernel coefficients, α , β , and θ , on the proposed SaMKL method. As shown in Fig. 8, it can be seen that the classification accuracy first increases and then decreases as the values of α , β , and θ gradually increase in the four test datasets. This means that the introduction of spectral, spatial, and textural kernels for the proposed method is

TABLE I
NUMBER OF TRAINING AND TEST SAMPLES SELECTED FROM DIFFERENT CLASSES IN THE FOUR DATASETS

Class	Indian Pines			University of Pavia			Center of Pavia			Washington DC		
	Name	Train	Test	Name	Train	Test	Name	Train	Test	Name	Train	Test
1	Grass_T	3	43	Asphalt	19	6612	Water	4	3271	Roof	47	3082
2	Grass_P	13	1415	Meadows	19	18630	Trees	3	2640	Street	27	1763
3	Corn	9	821	Gravel	19	2080	Meadows	1	891	Grass	22	1380
4	Soybeans_N	3	234	Trees	19	3045	Self_BB	3	2137	Tree	19	1245
5	Grass_T	5	478	Sheets	19	1326	Bare soil	1	800	Water	18	1176
6	Grass_P	6	724	Soil	19	5010	Asphalt	3	2962	Square	17	1103
7	Hay_W	3	25	Bitumen	19	1311	Bitumen	7	7280			
8	Oats	5	473	Bricks	19	3663	Tiles	2	603			
9	Soybeans_M	3	17	Shadows	19	928	Shadows	2	1415			
10	Soybeans_C	8	964									
11	Alfalfa	25	2430									
12	Grass_M	6	587									
13	Wheat	3	202									
14	Woods	11	1254									
15	Buildings	4	382									
16	Stone	3	90									
	Total	110	10139	Total	171	42605	Total	26	21999	Total	114	228

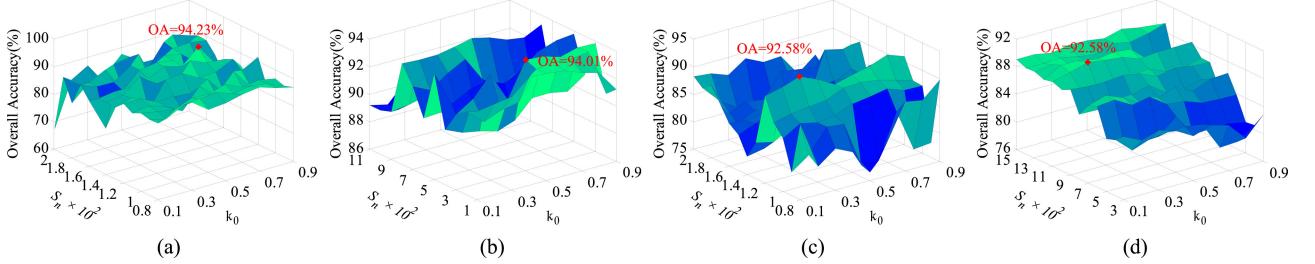


Fig. 7. Influence of parameters S_n and k_0 on the performance of the proposed method. (a) Indian Pines image. (b) University of Pavia image. (c) University of Pavia Center image. (d) Washington DC image.

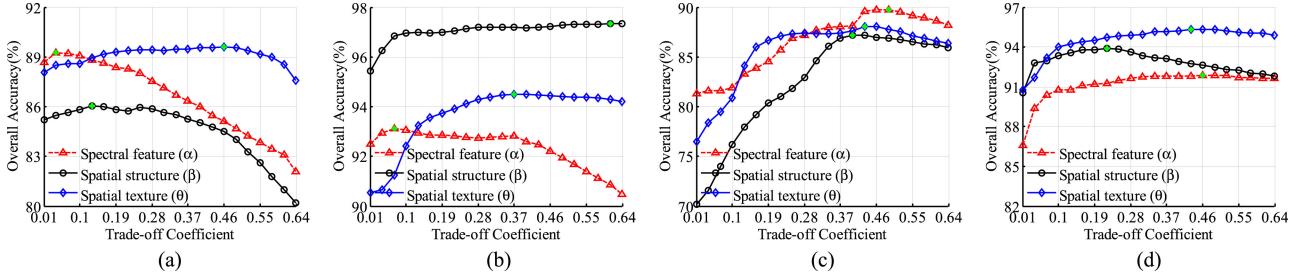


Fig. 8. Influence of parameters α , β , and θ on the performance of the proposed method. (a) Indian Pines image. (b) University of Pavia image. (c) University of Pavia Center image. (d) Washington DC image.

effective in terms of improving HSI classification. Focusing on Fig. 8(c), the proposed SaMKL method can achieve the highest classification accuracy on the Center of Pavia image when the values of α , β , and θ are set to 0.49, 0.4, and 0.43, respectively. Moreover, since the image structures in the University of Pavia, Center of Pavia, and Washington DC images are more detailed than those in the Indian Pines image, the optimal value of their kernel coefficients is larger than that of the Indian Pines image.

C. Ablation Analysis

In this section, we investigate the contribution of each stage of the proposed SaMKL method on the Indian Pines test image

using 1.0% randomly selected training samples from each class. Specifically, we consider three stages: stage I: spectral information with multikernel learning, stage II: spatial information with multikernel learning, and stage III: textural information with multikernel learning. Table II reports the classification accuracy obtained after conducting each stage in the proposed method and the final classification accuracy obtained by the SaMKL method. It can be seen that the classification performance of the proposed SaMKL method is always better than that obtained for any individual stages of the SaMKL. For example, the proposed SaMKL method exhibits the highest classification accuracy, with an OA value reaching 90.94%, while the OAs of other stages (i.e., Stage I, Stage II, and Stage III) are 84.08%, 86.23%, and

TABLE II
CLASSIFICATION ACCURACY (IN PERCENT) FOR EACH STEP IN THE PROPOSED METHOD

Index	Stage I	Stage II	Stage III	SaMKL
OA	84.08(1.90)	86.23(4.19)	87.01(3.08)	90.94(1.27)
AA	85.80(2.53)	87.86(2.20)	89.24(3.06)	92.03(1.22)
Kappa	81.85(2.11)	84.28(4.78)	84.28(4.78)	89.63(1.47)

Stage I: spectral information with multikernel learning. Stage II: spatial information with multikernel learning. Stage III: textural information with multikernel learning.

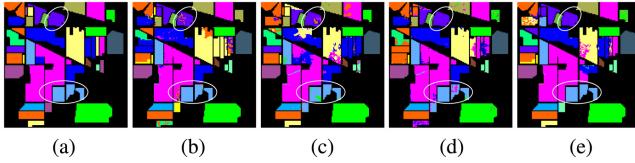


Fig. 9. Comparison between the reference maps and the classification maps achieved from each stage of the proposed method. (a) Reference map. (b) Classification map obtained by stage 1: spectral information with multikernel learning of the proposed method. (c) Classification map obtained by stage 2: spatial information with multikernel learning of the proposed method. (d) Classification map obtained by stage 3: textural information with multikernel learning of the proposed method. (e) Final classification map obtained by the proposed method.

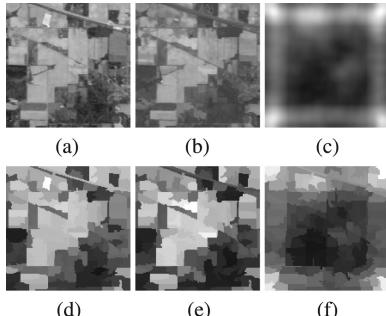


Fig. 10. HSI feature maps obtained for each stage of the proposed method. (a) Raw image. (b) Image obtained after applying the EF. (c) Image obtained after applying 2-D Gabor filtering. (d) Superpixel guidance fusion-based image obtained from (a). (e) Superpixel guidance fusion-based image obtained from (b). (f) Superpixel guidance fusion-based image obtained from (c).

87.01%, which means a decrease of 6.86%, 4.71%, and 3.93%, respectively.

In addition, Fig. 9 shows a comparison between the reference maps and the classification maps obtained from each stage in the proposed method. As can be seen, Fig. 9(b) and (c) shows significant misclassification results for the Grass_T and Soybeans_C in the circled areas and Fig. 9(d) achieves better classification results in this area. Specifically, Fig. 9(e) presents the best classification results obtained after conducting majority voting on the results in Fig. 9(b)–(d) with respect to results in Fig. 9(b)–(d). The reason is that different types of features (i.e., spectral, spatial, and textural) contain complementary information (see Fig. 10). More importantly, this means that the majority voting conducted after obtaining different features can further promote the effectiveness and robustness of the proposed method in terms of classification accuracy.

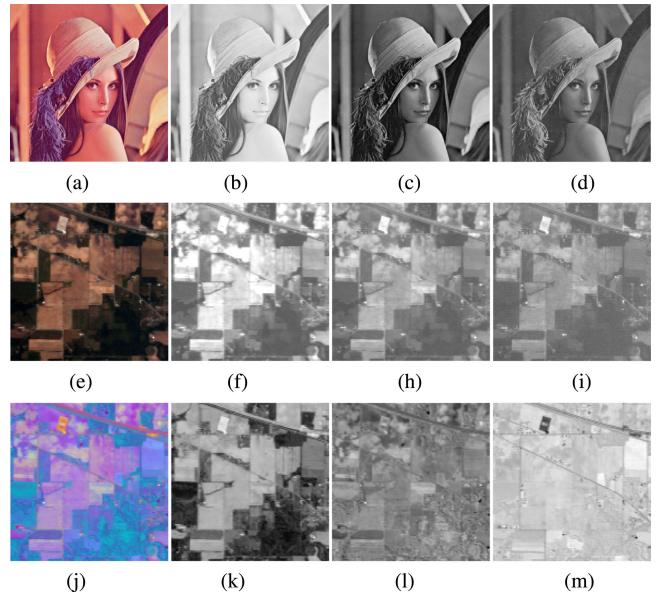


Fig. 11. (a) Natural image “Lena.” (b)–(d) R, G, and B channels of the “Lena” natural image. (e) RGB image synthesized from the Indian Pines HSI image. (f)–(i) R (625–750 nm), G (495–570 nm), and B (435–450 nm) channels of the Indian Pines HSI image. (j) Pseudocolor image synthesized from pre-three PCs of the Indian Pines HSI image. (k)–(m) First, second, and third PCs of the Indian Pines HSI image.

D. Discussion on Different SaMKL Aspects

1) *Virtual RGB Images and PCs in HSI Filtering:* In this section, we investigate the advantage of using virtual RGB (VRGB) images [68] over using the first three PCs in the filtering process of the proposed SaMKL method, using the Indian Pines image. As shown in Fig. 11(a)–(d), we extract grayscale images of the different channels of the natural image “Lena.” It can be seen that the brightness of the gray images of different channels is quite diverse, and their spatial details are complementary. Motivated by this, we generate a VRGB image from the original HSI according to their corresponding spectral wavelengths. The details of this operation can be seen in [68]. Compared with the first three PCs [see Fig. 11(k)–(m)], the VRGB image is more suitable for extraction of texture features and exhibits richer spatial features in terms of visual interpretation [see Fig. 11(e)–(i)].

Then, several experiments are conducted on the Indian Pines dataset to test the effectiveness of VRGB and PCs on the proposed SaMKL method by extracting different types of features (i.e., spatial and texture-based ones). The adopted morphological operations for spatial-based feature extraction are the EMP [69], DMP [70], AP [44], and EP [45]. The operations for textural feature extraction are RF [47], ATV [48], GuF [49], and 2-D GF [50]. The training samples are randomly selected, accounting for 1.0% (Indian Pines), 0.42% (University of Pavia), 0.1% (University of Center), and 1.0% (University of Houston) of the available labeled samples. In Tables III and IV, the classification accuracies of the proposed SaMKL method using VRGB and PCA on the different operations (spatial and texture-based) are presented. It can be seen that the proposed SaMKL method

TABLE III
CLASSIFICATION RESULTS (IN PERCENTAGE) OBTAINED BY THE PROPOSED METHOD USING DIFFERENT MORPHOLOGICAL FEATURE EXTRACTION ALGORITHMS

Methods	EMP		DMP		AP		EP	
	VRGB	PCs	VRGB	PCs	VRGB	PCs	VRGB	PCs
OA	84.03(3.36)	87.06(5.54)	85.01(2.72)	88.02(2.64)	84.69(4.76)	86.67(2.66)	83.91(2.39)	88.31(4.48)
AA	86.27(3.66)	88.42(4.51)	87.37(2.26)	90.58(2.96)	86.41(3.87)	88.59(2.06)	86.28(2.06)	90.21(4.62)
Kappa	81.68(3.97)	85.21(6.46)	82.82(3.15)	86.32(3.02)	82.39(5.58)	84.71(3.12)	81.57(2.74)	86.60(5.22)

The number in the parenthesis is the standard variation of the accuracies obtained after ten repeated experiments.

TABLE IV
CLASSIFICATION RESULTS (IN PERCENTAGE) OBTAINED BY THE PROPOSED METHOD USING DIFFERENT FILTERING FEATURE EXTRACTION ALGORITHMS

Methods	RF		ATV		GuF		2-D-GF	
	VRGB	PCs	VRGB	PCs	VRGB	PCs	VRGB	PCs
OA	79.44(1.97)	80.97(8.61)	77.12(3.26)	82.02(7.26)	77.36(6.25)	80.87(6.72)	82.54(4.84)	87.39(2.19)
AA	83.02(1.42)	81.61(10.8)	80.67(2.37)	83.71(8.43)	80.27(6.15)	83.16(5.34)	84.62(4.36)	89.70(2.14)
Kappa	76.44(2.27)	78.04(10.3)	73.71(3.84)	79.36(8.58)	73.71(7.83)	77.98(8.01)	79.91(5.69)	85.63(2.48)

The number in the parenthesis is the standard variation of the accuracies obtained after ten repeated experiments.

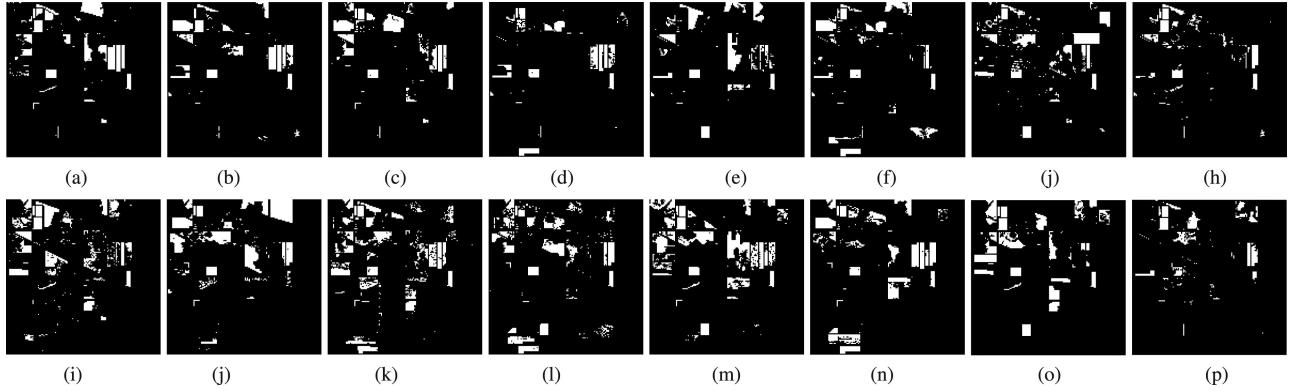


Fig. 12. Comparison of misclassification results using different morphological algorithms and filtering methods on the Indian Pines image. Misclassification results for morphological algorithms: (a) EMP-RGB (OA = 85.01%), (b) EMP-PCA (OA = 87.14%), (c) DMP-RGB (OA = 86.26%), (d) DMP-PCA (OA = 89.85%), (e) AP-RGB (OA = 86.63%), (f) AP-PCA (OA = 87.54%), (g) EP-RGB (OA = 84.96%), and (h) EP-PCA (OA = 91.12%). Misclassification results for filtering methods: (i) RF-RGB (OA = 81.53%), (j) RF-PCA (OA = 79.27%), (k) ATV-RGB (OA = 79.52%), (l) ATV-PCA (OA = 76.52%), (m) GF-RGB (OA = 82.45%), (n) GF-PCA (OA = 76.52%), (o) 2D-Gabor-RGB (OA = 81.25%), and (p) 2D-Gabor-PCA (OA = 90.65%).

(using PCs) can achieve the highest classification accuracy on any spatial and textural information extraction operation. In addition, we give the misclassification results of the proposed SaMKL method using different morphological algorithms and filtering methods on the Indian Pines test image in Fig. 12. It is worth noting that the PCA-based SaMKL method exhibits significant misclassification results (white part). This is because, although VRGB images have good visual appearance, they cannot include the main spectral and spatial details in the HSI images as compared to PCs.

2) *Discussion on Kernel Learning*: Here, we discuss the impact on the classification results obtained by the proposed SaMKL method with kernel learning using various structure-aware techniques with the considered HSI datasets. The training and test samples used are given in Table I. In order to objectively and fairly compare the obtained classification results, all experiments were repeated ten times. Table V shows the classification results obtained by the proposed SaMKL method using different kernel learning rules on the four HSI test images.

Stages I, II, and III refer to the spectral, spatial, and textural feature-based structure-aware techniques, respectively. In Table V, it can be seen that the classification accuracy achieved by any two-stage kernel learning combination is lower than that achieved by the three-stage kernel learning combination, for any considered HSI dataset. For instance, on the Indian Pines dataset, the classification accuracy of F-Ss is only 82.09%, while the classification accuracy of F-Ss-T can reach 88.56%. The reason is that two-stage kernel learning only fuses two types of features, while three-stage kernel learning can fully extract the spectral, spatial, and texture features from HSIs. In addition, it can be seen that the kernel with fusion strategy is better than the kernel with simple feature stacking. Specifically, for the University of Pavia dataset, the OA of the proposed F-Sf-T method is 96.39%, while OA of the F-Ss-T is only 88.85%. This further demonstrates that the kernel with simple feature stacking cannot achieve better classification accuracy than the multikernel learning strategy, which can make full use of the spectral, spatial, and textural information contained in the original HSIs.

TABLE V
CLASSIFICATION RESULTS (IN PERCENT) OBTAINED FROM SUPERPIXEL-BASED MULTIKERNEL REPRESENTATION USING DIFFERENT FEATURE AND DECISION FUSION APPROACHES ON THE FOUR CONSIDERED HSI IMAGES

		Indian Pines		
Fusion Strategy		OA	AA	Kappa
SaMKL	F-Ss	82.09(5.91)	84.84(6.30)	79.43(6.93)
	F-Sf	83.85(3.73)	86.97(3.16)	81.56(4.24)
	Ss-T	84.10(2.44)	86.86(2.16)	81.82(2.87)
	Sf-T	84.72(6.78)	87.16(5.89)	82.51(7.92)
	F-T	85.20(4.64)	88.31(2.62)	83.15(5.22)
	F-Ss-T	85.16(5.47)	87.25(4.92)	82.95(6.37)
	F-Sf-T	88.56(2.42)	90.30(2.00)	86.95(2.79)
		University of Pavia		
Fusion Strategy		OA	AA	Kappa
SaMKL	F-Ss	86.31(2.45)	90.35(2.27)	82.39(2.99)
	F-Sf	84.71(5.18)	88.59(4.20)	80.40(6.42)
	Ss-T	92.26(4.13)	95.52(1.98)	90.03(5.07)
	Sf-T	94.75(1.74)	96.43(1.13)	93.13(2.22)
	F-T	88.85(4.70)	92.01(3.37)	85.54(5.99)
	F-Ss-T	88.85(4.70)	92.01(3.37)	85.54(5.99)
	F-Sf-T	96.39(1.11)	97.90(0.70)	95.26(1.43)
		Center of Pavia		
Fusion Strategy		OA	AA	Kappa
SaMKL	F-Ss	82.75(3.81)	71.59(5.28)	78.69(5.00)
	F-Sf	84.07(3.58)	73.86(4.36)	80.54(4.26)
	Ss-T	86.09(3.78)	79.18(4.49)	82.88(4.73)
	Sf-T	86.50(2.76)	80.03(3.92)	83.26(3.68)
	F-T	84.40(2.60)	73.25(6.47)	80.86(3.36)
	F-Ss-T	87.51(2.64)	77.54(4.58)	84.59(3.41)
	F-Sf-T	88.15(1.98)	78.92(4.12)	85.36(2.48)
		Washington DC		
Fusion Strategy		OA	AA	Kappa
SaMKL	F-Ss	91.82(1.34)	92.28(1.35)	89.82(1.66)
	F-Sf	91.79(1.51)	92.21(1.43)	89.78(1.87)
	Ss-T	93.30(1.94)	93.07(2.56)	91.64(2.45)
	Sf-T	93.02(1.27)	93.17(1.25)	91.30(1.57)
	F-T	91.50(1.78)	92.11(1.80)	89.44(2.20)
	F-Ss-T	94.07(1.03)	94.10(1.01)	92.61(1.27)
	F-Sf-T	94.34(2.32)	94.31(2.13)	92.94(2.90)

The number in the parenthesis is the standard variation of the accuracies obtained after ten repeated experiments. ⁵F: Spectral-feature-based structure-aware stage (termed as Stage I). S: Spatial-feature-based structure-aware stage (termed as Stage II). T: Textural-feature-based structure-aware stage (termed as Stage III). Moreover, for the EPs conducted on each PC, Ss refers to feature stacking and Sf denotes feature fusion in Stage II.

E. Comparison of Different Methods

In this section, we compare the proposed SaMKL method with several commonly used spectral-based and spatial-spectral classification algorithms, including the SVM [22], generalized composite kernel (SVM-CK and MLR-GCK) [26], edge-preserving filtering (EPF) [71], EPs-fusion method (EPs-F) [72], superpixel-based classification via multiple kernels (SC-MK) [27], hierarchical GuF-based ensemble learning (HiFi-We) [73], and structural-kernel collaborative representation (SKCR) [74] on the four real HSI datasets. For all the compared methods, we used the default parameter settings given by the authors in the corresponding published papers.

For the Indian Pines dataset, the 16 categories of ground reference data are randomly divided into two sample sets with different proportions. The training sample and test samples

account for 1.0% and 99.0% of the labeled samples, respectively. Table I shows the specific number of training samples and test samples for each class. In addition, Table VI shows the OA, AA, and Kappa coefficient of different competitive methods and the proposed SaMKL method for each class of Indian Pines scene. As shown in Table VI, when the training sample accounts for 1.0% of the ground reference data, the SVM can effectively distinguish the ground objects with large spectral discrimination. For example, the recognition accuracies of Oats and Stone reach more than 90%. However, since the spatial information of HSIs is not considered, the accuracy of pixel-based spectral classifiers is not high for some spectrally similar features. For instance, the OAs of Grass_T, Soybeans_N, and Hay_W only reached 26.71%, 30.46%, and 43.12%, respectively. By using the spatial information in the image, SVM-CK, MLR-GCK, EPF, and EPs-F can always achieve higher classification accuracies than pixel-level features during classification. However, the OA values achieved by the four considered spatial-spectral classification methods are all lower than 80%, and the classification accuracy of each class is not optimal. Compared with other spectral feature extraction methods, the OA, AA, and Kappa coefficients of the proposed SaMKL method are all the highest, and the result of the SaMKL method is 28.91% higher than that of SVM in terms of OA. Fig. 13 shows the classification results obtained by different comparison methods. As shown in Fig. 13, compared with other competitive methods (including classic and state-of-the-art approaches), the proposed SaMKL method performs better in terms of classification performance. Although SC-MK, HiFi-We, and SKCR methods can achieve higher classification accuracies, some salt-and-pepper noise remains in the classification results. In contrast, the proposed SaMKL method can effectively remove this noise-like misclassification and obtain higher classification accuracy. The reason is that the SaMKL method effectively integrates spectral, spatial, and textural information by introducing multikernel learning and thus effectively avoiding misclassification.

For the University of Pavia dataset, the classification accuracy (including the accuracies of each class, OA, AA, and Kappa coefficients) of different algorithms are all shown in Table VII. In this experiment, only 0.42% of the ground reference data are selected for the training. The number of specific training and test samples for each class are also shown in Table I. Obviously, when few training samples are used, it is more difficult to achieve accurate HSI classification. However, the classification accuracy of the proposed SaMKL method is still superior to that achieved by the other compared methods, reaching 96.39%. Moreover, the proposed SaMKL method exhibits the best classification accuracy for Asphalt, Gravel, Soil, and Shadows classes (97.96%, 97.57%, 99.70%, and 100.0%), respectively. Fig. 14 shows the classification results obtained by various methods in the University of Pavia scene. As shown in the figure, although some compared methods (e.g., the EPF and EPs-F methods) utilized spectral-spatial information, there are still many misclassifications in the final classification results. In contrast, the proposed SaMKL method can effectively avoid salt-and-pepper noise and achieve the highest classification accuracy.

TABLE VI

CLASSIFICATION ACCURACY (IN PERCENT) OBTAINED FOR THE INDIAN PINES IMAGE BY THE SVM [22], SVM-CKK [26], MLR-GCK [26], EPF [71], EPs-F [72], SC-MK [27], HiFi-WE [73], SKCR [74], AND SAMKL METHODS

The number of training samples is 1.0% of the reference data.									
Class	SVM	SVM-CK	MLR-GCK	EPF	EPs-F	SC-MK	HiFi-We	SKCR	SaMKL
1	26.71(7.24)	69.77(15.4)	79.07(12.4)	70.32(37.0)	96.74(1.63)	100.0(0.00)	97.91(0.74)	100.0(0.00)	100.0(0.00)
2	52.53(7.21)	52.66(5.18)	53.24(4.38)	64.47(18.1)	70.23(10.0)	80.14(8.22)	77.56(7.50)	82.36(12.2)	80.92(9.96)
3	48.67(11.2)	41.38(10.3)	39.15(8.10)	77.77(17.5)	70.93(8.14)	76.36(6.19)	84.24(10.2)	67.78(18.0)	87.33(5.84)
4	30.46(12.0)	33.03(6.81)	45.43(20.0)	48.31(40.0)	45.38(10.3)	74.91(19.4)	68.38(18.0)	72.31(28.9)	77.82(20.7)
5	80.26(11.9)	48.37(13.3)	36.95(6.88)	95.60(5.05)	74.04(16.6)	75.88(8.13)	75.42(8.77)	79.16(7.04)	77.51(7.04)
6	73.70(8.88)	82.04(4.15)	62.27(13.3)	79.01(4.25)	90.44(3.55)	90.68(13.8)	93.55(4.40)	94.94(12.0)	95.86(5.21)
7	43.12(28.0)	92.00(4.62)	85.20(10.8)	72.76(35.6)	92.40(1.26)	99.60(1.26)	96.80(2.53)	99.60(1.26)	99.20(1.69)
8	91.01(5.68)	90.04(7.04)	73.40(13.1)	97.13(3.67)	98.63(3.47)	100.0(0.00)	90.93(7.19)	96.91(9.76)	94.65(17.0)
9	16.22(9.12)	75.88(16.7)	71.18(21.7)	52.97(39.3)	80.59(31.4)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)
10	54.64(9.04)	49.34(7.92)	58.74(9.93)	73.88(11.9)	59.18(18.6)	72.43(12.7)	72.31(12.4)	70.74(18.6)	76.48(8.01)
11	56.71(3.78)	82.08(5.36)	73.05(5.72)	54.90(5.07)	90.60(4.84)	84.73(5.41)	86.35(4.27)	94.60(2.52)	94.08(4.93)
12	33.77(8.46)	25.47(6.28)	33.58(12.4)	60.90(23.5)	49.63(11.2)	56.58(19.1)	67.84(15.0)	80.17(11.1)	70.70(8.08)
13	79.38(6.90)	90.00(4.76)	61.44(18.5)	97.91(2.49)	98.81(0.67)	90.69(14.8)	98.61(1.51)	100.0(0.00)	99.95(0.16)
14	85.49(4.59)	92.40(6.11)	85.19(6.56)	86.50(5.72)	94.39(9.06)	90.90(6.30)	96.17(6.43)	90.55(9.82)	99.45(0.84)
15	33.20(14.2)	39.32(9.28)	36.47(11.8)	74.43(19.3)	78.12(8.33)	67.38(20.3)	80.58(12.1)	59.97(28.4)	81.07(13.1)
16	97.95(3.26)	71.11(9.99)	59.33(14.4)	96.77(4.02)	82.22(3.92)	98.56(1.05)	98.44(1.50)	97.89(0.35)	97.89(0.35)
OA	59.13(1.43)	65.62(2.07)	60.62(1.89)	67.67(4.33)	79.44(4.07)	81.58(1.95)	83.79(2.33)	84.88(5.19)	88.04(2.33)
AA	56.49(3.57)	64.68(2.13)	59.60(3.83)	75.23(5.34)	79.52(5.17)	84.93(2.15)	86.57(2.52)	86.69(5.19)	89.56(2.69)
Kappa	52.83(1.45)	60.06(2.37)	54.80(2.23)	61.88(5.27)	76.34(4.91)	79.01(2.20)	81.51(2.65)	82.68(6.09)	86.32(2.64)

The number in the parenthesis is the standard variation of the accuracies obtained after ten repeated experiments.

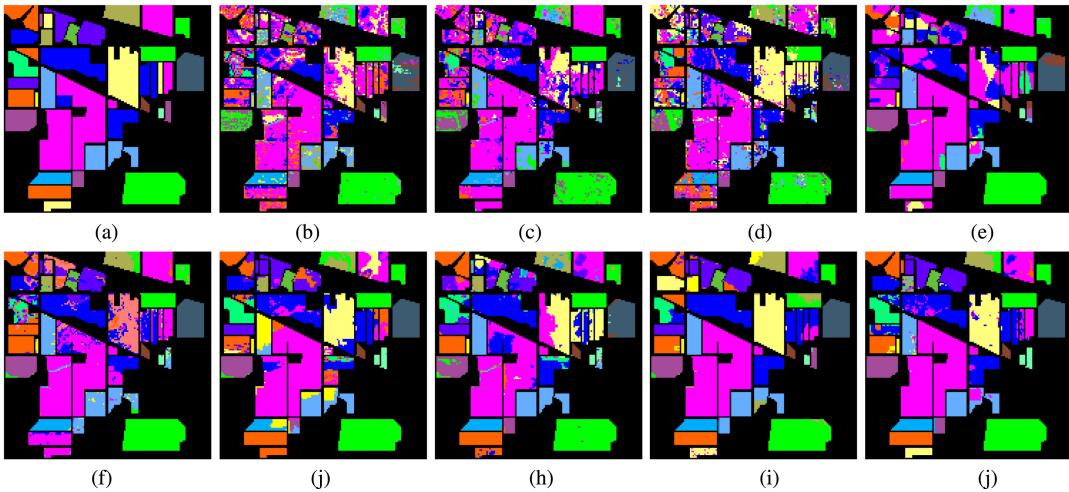


Fig. 13. Classification maps obtained by different methods trained with 1.0% of the available labeled data on the Indian Pines test image. (a) Reference map. (b) SVM; OA = 56.88%. (c) SVM-CK; OA = 63.42%. (d) MLR-GCK; OA = 63.78%. (e) EPF; OA = 64.17%. (f) EPs-F; OA = 79.17%. (g) SC-MK; OA = 79.47%. (h) HiFi-We; OA = 81.42%. (i) SKCR; OA = 84.78%. (j) SaMKL; OA = 89.66%

TABLE VII

CLASSIFICATION ACCURACY (IN PERCENT) OBTAINED FOR THE UNIVERSITY OF PAVIA IMAGE BY THE SVM [22], SVM-CKK [26], MLR-GCK [26], EPF [71], EPs-F [72], SC-MK [27], HiFi-WE [73], SKCR [74], AND SAMKL METHODS

The number of training samples is 0.42% of the reference data.									
Class	SVM	SVM-CK	MLR-GCK	EPF	EPs-F	SC-MK	HiFi-We	SKCR	SaMKL
1	95.23(1.45)	70.64(4.07)	93.40(3.33)	97.04(1.26)	95.81(2.62)	94.33(2.87)	82.53(5.57)	85.93(7.12)	97.96(1.54)
2	94.95(1.09)	81.29(5.60)	88.98(6.17)	95.26(2.42)	92.03(6.18)	87.68(4.73)	85.05(6.00)	82.58(9.19)	94.26(2.43)
3	66.53(6.06)	72.76(6.51)	96.75(1.34)	84.93(11.1)	99.93(0.06)	96.74(2.38)	90.47(5.17)	87.88(5.62)	97.57(4.01)
4	78.27(6.82)	94.07(2.53)	96.26(1.29)	73.87(14.4)	90.85(4.26)	93.65(4.73)	88.09(3.59)	90.96(3.96)	94.80(3.14)
5	95.50(2.63)	99.47(0.15)	99.25(0.58)	92.32(6.32)	99.99(0.02)	99.68(0.82)	96.58(2.30)	96.82(5.29)	99.90(0.12)
6	68.08(4.71)	82.39(4.47)	88.90(4.42)	66.60(17.5)	99.09(2.28)	96.82(3.84)	91.29(3.86)	85.82(9.37)	99.70(0.21)
7	59.88(5.13)	91.62(2.81)	99.63(0.17)	77.63(8.56)	99.97(0.04)	99.99(0.02)	97.95(2.69)	99.28(0.68)	99.93(0.02)
8	82.52(3.51)	70.79(7.21)	92.13(5.17)	86.51(6.50)	98.70(0.94)	86.74(6.60)	84.40(6.59)	84.60(5.59)	97.01(3.46)
9	99.96(0.08)	99.95(0.06)	98.35(1.43)	97.50(1.30)	100.0(0.00)	98.64(4.27)	99.91(0.16)	99.96(0.06)	100.0(0.00)
OA	85.13(1.58)	80.65(2.69)	91.68(2.65)	85.66(4.42)	94.99(2.47)	91.56(2.19)	86.90(2.38)	85.84(3.56)	96.39(1.11)
AA	82.32(1.54)	84.78(0.98)	94.85(0.70)	85.74(3.54)	97.38(0.70)	94.92(1.13)	90.69(1.33)	90.42(1.58)	97.90(0.70)
Kappa	80.70(1.98)	75.15(3.20)	89.17(3.29)	81.67(5.29)	93.47(3.13)	89.07(2.73)	83.13(2.89)	81.80(4.17)	95.26(1.43)

The number in the parenthesis is the standard variation of the accuracies obtained after ten repeated experiments.

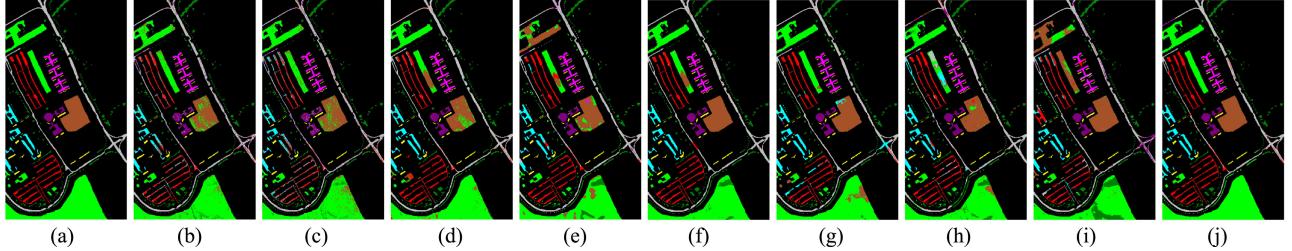


Fig. 14. Classification maps obtained by different methods trained with 0.42% of labeled data on the University of Pavia test image. (a) Reference map. (b) SVM; OA = 84.63. (c) SVM-CK; OA = 82.01. (d) MLR-GCK; OA = 87.14. (e) EPF; OA = 86.82. (f) EPs-F; OA = 84.61. (g) SC-MK; OA = 88.12. (h) HiFi-We; OA = 87.54. (i) SKCR; OA = 85.25. (j) SaMKL; OA = 96.39.

TABLE VIII

CLASSIFICATION ACCURACY (IN PERCENT) OBTAINED FOR THE CENTER OF PAVIA IMAGE BY THE SVM [22], SVM-CKK [26], MLR-GCK [26], EPF [71], EPs-F [72], SC-MK [27], HiFi-We [73], SKCR [74], AND SAMKL METHODS

The number of training samples is 0.1% of the reference data.									
Class	SVM	SVM-CK	MLR-GCK	EPF	EPs-F	SC-MK	HiFi-We	SKCR	SaMKL
1	99.69(0.29)	100.0(0.00)	100.0(0.00)	99.80(0.29)	100.0(0.00)	89.13(3.82)	99.99(0.02)	100.0(0.00)	100.0(0.00)
2	83.55(6.21)	96.27(4.88)	88.78(5.01)	84.98(8.51)	75.45(9.97)	80.08(13.2)	92.54(9.96)	83.55(14.6)	92.62(4.61)
3	72.13(30.8)	14.76(11.7)	23.56(19.1)	74.49(33.4)	21.12(17.5)	21.35(9.27)	44.94(22.1)	19.06(9.99)	24.52(10.7)
4	68.18(10.5)	81.58(5.03)	83.69(7.57)	80.26(16.0)	94.31(16.5)	97.29(3.19)	86.88(7.73)	82.84(17.3)	85.77(10.4)
5	61.70(32.9)	14.34(10.3)	43.65(12.8)	57.17(47.1)	72.61(22.6)	75.55(22.9)	57.19(18.0)	48.33(21.6)	66.00(10.1)
6	59.14(31.7)	77.61(14.7)	91.13(9.25)	85.99(9.72)	53.52(21.6)	79.37(14.1)	89.88(6.31)	79.63(16.2)	85.16(10.9)
7	76.19(15.2)	90.51(2.83)	90.71(6.75)	81.00(14.0)	91.29(8.09)	88.16(5.91)	84.75(5.30)	88.64(6.93)	94.64(3.85)
8	93.42(12.2)	32.48(10.6)	67.95(27.0)	79.47(33.3)	65.13(22.9)	50.58(29.7)	53.84(18.0)	28.82(23.4)	59.42(24.6)
9	99.73(0.32)	99.41(0.64)	89.66(17.1)	99.94(0.12)	80.78(13.8)	69.33(18.1)	95.84(9.54)	66.18(13.7)	99.94(0.16)
OA	79.10(5.79)	83.15(1.55)	86.11(2.91)	83.49(6.21)	80.98(3.68)	81.63(5.15)	86.10(2.73)	80.57(3.34)	88.55(1.85)
AA	79.30(5.76)	67.44(2.22)	75.46(4.36)	82.57(7.62)	72.69(7.46)	72.31(6.21)	78.43(4.97)	66.34(5.31)	78.67(3.99)
Kappa	73.66(8.13)	79.01(2.02)	82.93(3.58)	79.14(8.79)	76.42(4.84)	77.64(6.19)	83.16(3.26)	76.13(4.08)	85.84(2.37)

The number in the parenthesis is the standard variation of the accuracies obtained after ten repeated experiments.

TABLE IX

CLASSIFICATION ACCURACY (IN PERCENT) OBTAINED FOR THE WASHINGTON DC IMAGE BY THE SVM [22], SVM-CKK [26], MLR-GCK [26], EPF [71], EPs-F [72], SC-MK [27], HiFi-We [73], SKCR [74], AND SAMKL METHODS

The number of training samples is 1.0% of the reference data.									
Class	SVM	SVM-CK	MLR-GCK	EPF	EPs-F	SC-MK	HiFi-We	SKCR	SaMKL
1	86.27(2.35)	92.32(3.67)	92.66(3.15)	85.18(1.72)	94.37(1.59)	94.64(1.96)	94.85(2.16)	93.60(4.06)	94.57(3.21)
2	94.91(2.22)	98.29(1.27)	98.77(0.87)	73.77(3.63)	97.63(1.76)	97.21(1.35)	97.90(1.82)	97.94(0.77)	96.45(3.03)
3	77.88(7.33)	87.15(4.92)	86.38(5.17)	86.27(8.19)	82.95(4.21)	86.76(5.98)	83.35(5.84)	89.50(4.14)	88.31(4.15)
4	91.80(6.10)	95.54(3.00)	95.85(3.26)	96.49(3.85)	94.54(2.87)	94.27(4.13)	94.31(4.44)	94.92(2.02)	93.51(5.71)
5	97.06(1.69)	95.89(1.77)	96.53(0.87)	96.60(0.90)	95.77(1.16)	94.85(2.51)	94.17(6.05)	93.88(1.91)	95.71(1.41)
6	99.15(0.34)	95.81(2.10)	94.30(1.81)	99.12(0.56)	96.70(1.47)	95.86(1.24)	96.79(1.20)	94.64(1.54)	97.33(1.05)
OA	89.80(1.84)	93.91(2.00)	93.94(0.91)	85.44(1.51)	93.80(0.78)	94.10(0.97)	93.84(1.85)	94.13(1.55)	94.34(2.32)
AA	91.18(1.62)	94.17(1.80)	94.08(0.83)	89.57(1.66)	93.66(0.76)	93.93(1.13)	93.56(2.25)	94.08(1.14)	94.31(2.13)
Kappa	87.30(2.27)	92.41(2.49)	92.45(1.13)	81.62(1.91)	92.26(0.98)	92.65(1.22)	92.31(2.34)	92.69(1.91)	92.94(2.90)

The number in the parenthesis is the standard variation of the accuracies obtained after ten repeated experiments.

Besides the Indian Pines and University of Pavia datasets, we conducted similar experiments on the Center of Pavia and Washington DC datasets. Specifically, we selected 0.1% and 1.0% of the ground reference data as training samples on the two datasets, respectively. The specific numbers of training samples per class are shown in Table I. Tables VIII and IX show the classification accuracies for each class, along with the OA, AA, and Kappa coefficients of the different methods in the two experiments. It can be clearly seen from these tables that, for small training sets, the OA, AA, and Kappa coefficients of the proposed method are always better than those of other methods. Although the classification accuracy of each class is not optimal on the Washington DC dataset, the distance from the optimal classification accuracy is not obvious. This proves

that the proposed method is indeed an effective spectral–spatial classification method.

F. Effect of Using Different Numbers of Training Samples

In this section, we analyze the effect of using different numbers of training and test samples on the classification accuracies of the proposed SaMKL method. Fig. 15 shows the classification accuracy of the proposed SaMKL method (and that obtained by competitive methods) on the public HSI datasets under various sample conditions. It can be seen from Fig. 15 that the number of training samples affects the accuracy of the classification results. As the number of training samples increases, the classification accuracy of various methods gradually increases, but the

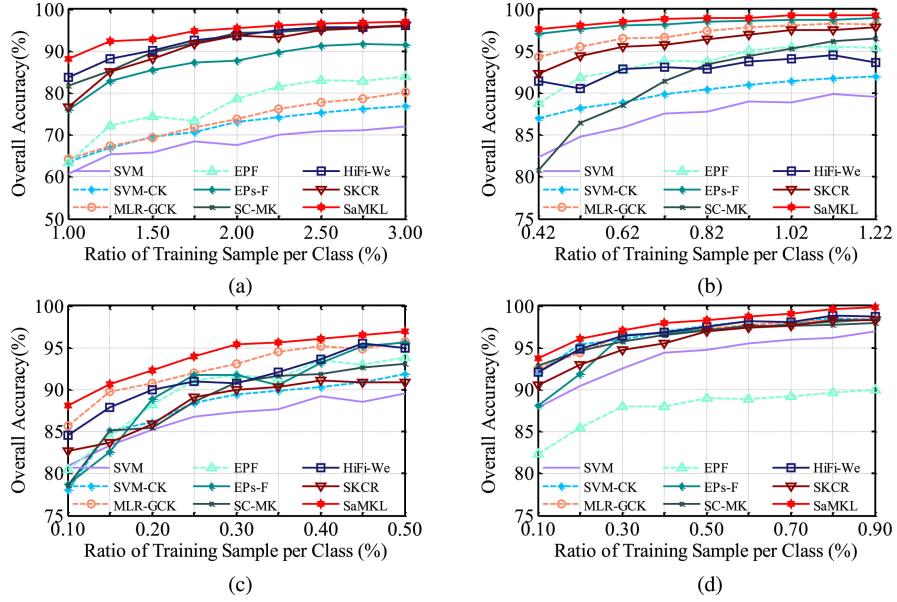


Fig. 15. Classification performance of the compared methods with different numbers of training samples on the four HSI datasets. (a) Indian Pines dataset. (b) University of Pavia dataset. (c) Center of Pavia dataset. (d) Washington DC dataset.

TABLE X
COMPUTING TIME (IN SECONDS) NEEDED FOR THE CLASSIFICATION OF FOUR TEST IMAGES OBTAINED BY SVM, SVM-GCK, MLR-GCK, EPF, EPS-F, SC-MK, HiFi-We, SKCR, AND THE PROPOSED SAMKL METHODS

Image	SVM	SVM-CK	MLR-GCK	EPF	EPS-F	SC-MK	HiFi-We	SKCR	SaMKL
Indian Pines	2.39	1.39	0.43	2.78	1.76	1.43	277.75	0.89	1.44
University of Pavia	5.17	5.24	2.87	6.13	5.77	3.47	156.13	2.38	5.20
Center of Pavia	1.56	3.92	1.28	19.48	3.08	2.00	240.25	1.36	2.80
Washington DC	4.27	2.40	1.31	4.12	2.56	2.73	182.40	1.88	2.26

proposed SaMKL method can always obtain higher classification accuracies (especially when very small training sets are used). For example, the experimental results on the Center of Pavia dataset show that, when only 0.1% of the samples are used for training, the SaMKL method can still obtain a classification accuracy close to 88%, while the SVM-CK can only achieve 78% accuracy. On the University of Pavia data, although only 0.42% of the samples are used for training, the SaMKL method can still obtain an overall classification accuracy close to 99%. Therefore, when the number of training samples is very limited, the proposed SaMKL method (with structure-aware-based multikernel learning) can also provide good classification results.

G. Analysis of Computational Time for Various Methods

In this section, we present the computational complexity of the above classification methods. All the algorithms have been implemented in MATLAB R2014a, and the experimental platform is a computer equipped with Intel(R) i7-7800X CPU 3.50 GHz and 32 GB of RAM. Table X shows the computational times for the proposed SaMKL method and other comparison methods. As shown in the table, the time for classification using the proposed method is relatively fast. Specifically, when processing the Indian Pines image, the entire process only takes 1.44 s. However, compared with other classification algorithms on the University of Pavia dataset, the proposed SaMKL method does not exhibit an advantage in terms of calculation speed. The reason is that

TABLE XI
COMPUTING TIME (IN SECONDS) FOR EACH STEP OF THE PROPOSED METHOD

Image	Stage I	Stage II	Stage III
Indian Pines	2.39	1.39	0.43
University of Pavia	5.17	5.24	2.87
Center of Pavia	1.56	3.92	1.28
Washington DC	4.27	2.40	1.31

Stage I: spectral information fusion with superpixel multikernel representation. Stage II: spatial information fusion with superpixel multikernel representation. Stage III: spatial texture fusion with superpixel multikernel representation.

the proposed SaMKL method not only extracts spectral, spatial, and textural features but also performs multikernel learning on the above features. Moreover, Table XI shows the computational complexity of each stage of the proposed SaMKL method. It can be seen that the computing time of the first stage is higher than that of the other two stages. To solve this problem, we will study how to use graphics processing units (GPUs) to accelerate our algorithm in future developments.

V. CONCLUSION AND FUTURE LINES

This article introduces a new SaMKL method for HSI classification. Our method is structure aware in the sense that it can consider different types of spatial and spectral features to effectively overcome weak assumptions in the integration of

these two sources of information. Besides, our method exploits multikernel learning from multiple features derived from the original HSI to efficiently improve the spectral separability of classes. Our experimental results, conducted on four different HSI datasets, indicate that the proposed SaMKL method outperforms well-known and state-of-the-art classification methods (particularly, when very limited training samples are available).

In the proposed framework, we exploit the concept of nearest-neighbor-based local density peak to define the structural information of the original HSI. However, remaining sources of spatial knowledge within superpixels have not been categorized and/or exploited. This deserves further analysis in future studies in order to exploit more accurately the spatial information for classification purposes. Moreover, our future work will also focus on developing new spatial-based methods able to obtain high-confidence structural information and further improve classification accuracy. Last but not least, our future work will also consider GPU technology to accelerate the computational performance of our method.

ACKNOWLEDGMENT

The authors would like to thank Dr. Nanjun He from Tencent Jarvis Lab and Mr. Siyuan Huang from Sichuan Suitang Technology Co., Ltd. for their enthusiastic help. In addition, the authors appreciate Prof. Jun Li, Prof. Pedram Ghamisi, and Prof. Leyuan Fang for providing the software of the SVM-CK, MLR-GCK, EPF, Eps, EPs-F, and SC-MK. The authors would like also to thank the editors and anonymous reviewers for their insightful comments and suggestions, which have significantly improved this article.

REFERENCES

- [1] B. Rasti *et al.*, “Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox,” *IEEE Trans. Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Apr. 2020.
- [2] J. Kang, D. Hong, J. Liu, G. Baier, N. Yokoya, and B. Demir, “Learning convolutional sparse coding on complex domain for interferometric phase restoration,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 826–840, Feb. 2021.
- [3] R. Huang, D. Hong, Y. Xu, W. Yao, and U. Stilla, “Multi-scale local context embedding for LiDAR point cloud classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 721–725, Jul. 2020.
- [4] D. Hong *et al.*, “More diverse means better: Multimodal deep learning meets remote-sensing imagery classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.
- [5] B. Tu, X. Yang, C. Zhou, D. He, and A. Plaza, “Hyperspectral anomaly detection using dual window density,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8503–8517, Dec. 2020.
- [6] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, “An augmented linear mixing model to address spectral variability for hyperspectral unmixing,” *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.
- [7] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, “Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction,” *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 35–49, 2019.
- [8] J. Jiang, J. Ma, and X. Liu, “Multilayer spectral-spatial graphs for label noisy robust hyperspectral image classification,” *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2020.3029523](https://doi.org/10.1109/TNNLS.2020.3029523).
- [9] J. M. Bioucas-Dias *et al.*, “Hyperspectral remote sensing data analysis and future challenges,” *IEEE Trans. Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [10] M. A. Lee, Y. Huang, H. Yao, S. J. Thomson, and L. M. Bruce, “Determining the effects of storage on cotton and soybean leaf samples for hyperspectral analysis,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2562–2570, Jun. 2014.
- [11] M. Kanning, B. Siegmann, and T. Jarmer, “Regionalization of uncovered agricultural soils based on organic carbon and soil texture estimations,” *Remote Sens.*, vol. 8, no. 11, pp. 927–943, Nov. 2016.
- [12] J. Chi and M. M. Crawford, “Spectral unmixing-based crop residue estimation using hyperspectral remote sensing data: A case study at Purdue University,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2531–2539, Jun. 2014.
- [13] J. Ryan, C. Davis, N. Tufillaro, R. Kudela, and B. C. Gao, “Application of the hyperspectral imager for the coastal ocean to phytoplankton ecology studies in Monterey Bay, CA, USA,” *Remote Sens.*, vol. 6, no. 2, pp. 1007–1025, Dec. 2014.
- [14] A. Brook and E. B. Dor, “Quantitative detection of settled dust over green canopy using sparse unmixing of airborne hyperspectral data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 884–897, Feb. 2016.
- [15] J. Ham, Yangchi M. M. ChenCrawford, and J. Ghosh, “Investigation of the random forest framework for classification of hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Feb. 2005.
- [16] M. Dalponte, H. O. Ørka, T. Gobakken, D. Ganelle, and E. Næsset, “Tree species classification in boreal forests with hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2632–2645, Oct. 2013.
- [17] Y. Jun, “Adaptive control of nonlinear PID-based analog neural networks for a nonholonomic mobile robot,” *Neurocomputing*, vol. 71, no. 7, pp. 1561–1565, Mar. 2008.
- [18] J. Zhu, L. Fang, and P. Ghamisi, “Deformable convolutional neural networks for hyperspectral image classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1254–1258, Aug. 2018.
- [19] J. Xie, N. He, L. Fang, and P. Ghamisi, “Multiscale densely-connected fusion networks for hyperspectral images classification,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 246–259, 2021.
- [20] M. Khodadadzadeh, J. Li, A. Plaza, and J. M. Bioucas-dias, “A subspace-based multinomial logistic regression for hyperspectral image classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2105–2109, Dec. 2014.
- [21] J. Li, J. M. Bioucas-Dias, and A. Plaza, “Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 318–322, Mar. 2013.
- [22] 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.
- [23] Z. Wu, Q. Wang, A. Plaza, J. Li, L. Sun, and Z. Wei, “Parallel spatial-spectral hyperspectral image classification with sparse representation and Markov random fields on GPUs,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2926–2938, Jun. 2015.
- [24] B. Tu, S. Huang, L. Fang, G. ZhangandJ. Wang, and B. Zheng, “Hyperspectral image classification via weighted joint nearest neighbor and sparse representation,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4063–4075, Nov. 2018.
- [25] J. Fan, T. Chen, and S. Lu, “Superpixel guided deep-sparse-representation learning for hyperspectral image classification,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3163–3173, Nov. 2018.
- [26] 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.
- [27] 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.
- [28] 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, Jun. 2015.
- [29] Y. Ma, C. Li, H. Li, X. Mei, and J. Ma, “Hyperspectral image classification with discriminative kernel collaborative representation and Tikhonov regularization,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 587–591, Apr. 2018.
- [30] L. Sun *et al.*, “Low rank component induced spatial-spectral kernel method for hyperspectral image classification,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3829–3842, Oct. 2020.

- [31] 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.
- [32] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.
- [33] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
- [34] Y. Xu, B. Du, F. Zhang, and L. Zhang, "Hyperspectral image classification via a random patches network," *ISPRS J. Photogramm. Remote Sens.*, vol. 142, pp. 344–357, 2018.
- [35] N. He *et al.*, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 755–769, Feb. 2019.
- [36] C. Zhou, B. Tu, Q. Ren, and S. Chen, "Spatial peak-aware collaborative representation for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: [10.1109/LGRS.2021.3083416](https://doi.org/10.1109/LGRS.2021.3083416).
- [37] S. Jia, Z. Lin, B. Deng, J. Zhu, and Q. Li, "Cascade superpixel regularized Gabor feature fusion for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1638–1652, May 2020.
- [38] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multi-task joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.
- [39] B. Tu, C. Zhou, X. Liao, Q. Li, and Y. Peng, "Feature extraction via 3-D block characteristics sharing for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3042274](https://doi.org/10.1109/TGRS.2020.3042274).
- [40] B. Tu, C. Zhou, J. Peng, G. Zhang, and Y. Peng, "Feature extraction via joint adaptive structure density for hyperspectral imagery classification," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 5006916.
- [41] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
- [42] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Feb. 2005.
- [43] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Feb. 2005.
- [44] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Jun. 2010.
- [45] 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, Jul. 2016.
- [46] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1641–1645, Aug. 2016.
- [47] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014.
- [48] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.
- [49] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [50] S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on Gabor filters," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1160–1167, Oct. 2002.
- [51] 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, Jan. 2016.
- [52] S. Jia *et al.*, "3-D Gaussian-Gabor feature extraction and selection for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8813–8826, Jul. 2019.
- [53] S. Jia, K. Wu, J. Zhu, and X. Jia, "Spectral-spatial Gabor surface feature fusion approach for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1142–1154, Sep. 2019.
- [54] L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification," *Image Vis. Comput.*, vol. 25, no. 5, pp. 553–563, May 2007.
- [55] P. Soille, "Morphological image analysis: Principles and applications," *Sensor Rev.*, vol. 28, no. 5, pp. 800–801, 1999.
- [56] C. Li, S. Li, X. Kang, and T. Lu, "Gabor filtering based deep network for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Nov. 2017, pp. 1808–1811.
- [57] L. He, C. Liu, J. Li, Y. Li, S. Li, and Z. Yu, "Hyperspectral image spectral-spatial-range Gabor filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4818–4836, Nov. 2020.
- [58] C. Liu, J. Li, L. He, A. Plaza, S. Li, and B. Li, "Naive Gabor networks for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 376–390, Jan. 2021.
- [59] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.
- [60] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
- [61] K. Sun, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 998–1002, May 2015.
- [62] R. X. X. Luo and J. Yin, "Information-assisted density peak index for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 41, pp. 1870–1874, Oct. 2017.
- [63] B. Tu, C. Zhou, D. He, S. Huang, and A. Plaza, "Hyperspectral classification with noisy label detection via superpixel-to-pixel weighting distance," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4116–4131, Jun. 2020.
- [64] O. Bousquet, "New approaches to statistical learning theory," *Ann. Inst. Statist. Math.*, vol. 55, no. 2, pp. 371–389, 2003.
- [65] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [66] C. Vachier and F. Meyer, "Extinction value: A new measurement of persistence," in *Proc. IEEE Workshop Nonlinear Signal Image Process.*, Jun. 1995, pp. 254–257.
- [67] M. Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy-rate clustering analysis via maximizing a submodular function subject to matroid constraint," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 99–112, Jan. 2014.
- [68] L. Liu, Z. Shi, B. Pan, N. Zhang, H. Luo, and X. Lan, "Multiscale deep spatial feature extraction using virtual RGB image for hyperspectral imagery classification," *Remote Sens.*, vol. 12, no. 2, Jan. 2020, Art. no. 280.
- [69] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
- [70] H. G. AkÇay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2097–2111, Jul. 2008.
- [71] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
- [72] L. Fang, N. He, S. Li, P. Ghamisi, and J. A. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1803–1815, Mar. 2018.
- [73] B. Pan, Z. Shi, and X. Xu, "Hierarchical guidance filtering-based ensemble classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4177–4189, Jul. 2017.
- [74] B. Tu, C. Zhou, X. Liao, G. Zhang, and Y. Peng, "Spectral-spatial hyperspectral classification via structural-kernel collaborative representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 861–865, May 2021.



Chengle Zhou (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering and automation in 2019 from the Hunan Institute of Science and Technology, Yueyang, China, where he is currently working the M.S. degree in information and communication engineering.

His research interests include image processing, pattern recognition, hyperspectral image classification, anomaly detection, and noisy label detection.

Dr. Zhou is the recipient of the Excellent Student Award at the 2021 IEEE Geoscience and Remote Sensing Society Summer School on Modeling in Microwave and Optical Remote Sensing.



Wei He (Member, IEEE) received the B.S. degree in computer science and technology in 2006 from Hunan Institute of Science and Technology, Yueyang, China; the M.S. degree in computer software and theory in 2009 from Changsha University of Science and Technology, Changsha, China, and the Ph.D. degree in information and communication Engineering in 2020 from Hoseo University, Asan, South Korea.

He was a Visiting Researcher with the Computer Vision Laboratory, University of Nevada, Reno, NV, USA, from August 2015 to August 2016. He is currently an Associate Professor with the School of Information Science and Engineering, Hunan Institute of Science and Technology. His current research interests include computer vision, pattern recognition, and video surveillance applications.



Bing Tu (Member, IEEE) received the M.S. degree in control science and engineering from the Guilin University of Technology, Guilin, China, in 2009, and the Ph.D. degree in mechatronic engineering from the Beijing University of Technology, Beijing, China, in 2013.

From 2015 to 2016, he was a Visiting Researcher with the Department of Computer Science and Engineering, University of Nevada, Reno, NV, USA, which is supported by the China Scholarship Council. Since 2018, he has been an Associate Professor with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang, China. His research interests include sparse representation, pattern recognition, and analysis in remote sensing.



Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is currently the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored more than 600 publications, including more than 200 JCR journal papers (more than 160 in IEEE journals), 23 book chapters, and around 300 peer-reviewed conference proceeding papers. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was a member of the Editorial Board of IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER from 2011 to 2012 and *IEEE Geoscience and Remote Sensing Magazine* in 2013. He was also a member of the Steering Committee of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He was a recipient of the recognition of Best Reviewer of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2009 and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2010. He was an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2007 to 2012. He was also a recipient of the Best Column Award of *IEEE Signal Processing Magazine* in 2015, the 2013 Best Paper Award of IEEE JSTARS, and the Most Highly Cited Paper (2005–2010) in *Journal of Parallel and Distributed Computing*. He received the Best Paper Awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He was the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012 and the President of the Spanish Chapter of the IEEE GRSS from 2012 to 2016. He has reviewed more than 500 manuscripts for more than 50 different journals. He was the Editor-in-Chief for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2013 to 2017. He has Guest Edited ten special issues on hyperspectral remote sensing for different journals. He is an Associate Editor for IEEE ACCESS (receiving the recognition as an Outstanding Associate Editor for the journal in 2017).



Nanying Li (Student Member, IEEE) received the B.S. degree in automation and the M.Sc. degree in information and communication engineering from the Hunan Institute of Science and Technology, Yueyang, China, in 2017 and 2020, respectively. She is currently working toward the Ph.D. degree in computer science and technology with the Shenzhen University, Shenzhen, China.

Her research interests include hyperspectral image classification and anomaly detection.