

FU, H., SUN, G., REN, J., ZHANG, A. and JIA, X. 2020. Fusion of PCA and segmented-PCA domain multiscale 2-D-SSA for effective spectral-spatial feature extraction and data classification in hyperspectral imagery. *IEEE transactions on geoscience and remote sensing* [online], 60, article 5500214. Available from: <https://doi.org/10.1109/TGRS.2020.3034656>

Fusion of PCA and segmented-PCA domain multiscale 2-D-SSA for effective spectral-spatial feature extraction and data classification in hyperspectral imagery.

FU, H., SUN, G., REN, J., ZHANG, A. and JIA, X.

2020

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Fusion of PCA and Segmented-PCA Domain Multiscale 2-D-SSA for Effective Spectral-Spatial Feature Extraction and Data Classification in Hyperspectral Imagery

Hang Fu, Genyun Sun¹, *Member, IEEE*, Jinchang Ren², *Senior Member, IEEE*,
Aizhu Zhang¹, *Member, IEEE*, and Xiuping Jia³, *Senior Member, IEEE*

Abstract—As hyperspectral imagery (HSI) contains rich spectral and spatial information, a novel principal component analysis (PCA) and segmented-PCA (SPCA)-based multiscale 2-D-singular spectrum analysis (2-D-SSA) fusion method is proposed for joint spectral-spatial HSI feature extraction and classification. Considering the overall spectra and adjacent band correlations of objects, the PCA and SPCA methods are utilized first for spectral dimension reduction, respectively. Then, multiscale 2-D-SSA is applied onto the SPCA dimension-reduced images to extract abundant spatial features at different scales, where PCA is applied again for dimensionality reduction. The obtained multiscale spatial features are then fused with the global spectral features derived from PCA to form multiscale spectral-spatial features (MSF-PCs). The performance of the extracted MSF-PCs is evaluated using the support vector machine (SVM) classifier. Experiments on four benchmark HSI data sets have shown that the proposed method outperforms other state-of-the-art feature extraction methods, including several deep learning approaches, when only a small number of training samples are available.

Index Terms—Classification, dimension reduction, feature fusion, hyperspectral imagery (HSI), multiscale 2-D-singular spectrum analysis (2-D-SSA).

I. INTRODUCTION

THE development of hyperspectral sensors provides hyperspectral imagery (HSI) containing detailed spectral and

spatial information simultaneously. HSI has hundreds of contiguous spectral bands acquired from visible light to infrared, and the spectral profile of each pixel can act as a fingerprint for identification and discrimination of various materials [1]. This characteristic makes HSI great use in many areas, such as mineralogy [2], agriculture [3], target detection [4], and land-cover classification [5]. However, how to efficiently classify HSI is still a major challenge.

Over the past decades, a lot of HSI classification techniques have been developed, including support vector machine (SVM) [6], multinomial logistic regression (MLR) [7], the k-nearest neighbor (KNN) [8], and sparse representation classifier (SRC) [9]. Among them, the SVM classifier has been widely used since its low generalization error rate and simple operation. However, the high spectral dimensionality of HSI usually causes the model parameters difficult to estimate when labeled samples are limited for training, which results in the “Hughes phenomenon” [10]. Meanwhile, the data redundancy and noise are also unavoidable, which brings obstacles to the classification [11]. Thus, effective spectral and spatial feature extraction is essential for HSI classification.

Specific to the abovementioned problems, some spectral feature extraction methods have been developed to reduce the spectral dimensionality. For instance, several widely used linear transformation methods include the principal component analysis (PCA) [12], the independent component analysis (ICA) [13], and linear discriminant analysis (LDA) [14], as well as some nonlinear dimension reduction methods, i.e., manifold learning [15]. PCA can transform high-dimensional data into linearly uncorrelated variables, which is beneficial for data compression and visualization, as well as mining of compact effective information. There are also some improved PCA methods, including structured covariance PCA (SC-PCA) [16], segmented PCA (SPCA) [17], and folded-PCA (FPCA) [18], which not only reduces computational load and memory requirements but also incorporates local spectral characteristics.

The utilization of spatial context information has a potential improvement in classification accuracy. One popular approach for spatial feature extraction is based on the morphological profile (MP) [19], which can extract spatial geometrical

Manuscript received July 30, 2020; revised September 28, 2020; accepted October 2, 2020. This work was supported by the National Key Research and Development Program (2019YFE0126700), the National Natural Science Foundation of China under Grants (41971292, 41801275, 41871270), Dazhi Scholarship of Guangdong Polytechnic Normal University. (Corresponding author: Genyun Sun.)

Hang Fu, Genyun Sun, and Aizhu Zhang are with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China, and also with the Laboratory for Marine Mineral Resources, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China (e-mail: genyun.sun@163.com).

Jinchang Ren is with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou 510665, China, and also with the Department of Electronics and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, U.K. (e-mail: jinchang.ren@ieee.org).

Xiuping Jia is with the School of Engineering and Information Technology, University of New South Wales at Canberra, Canberra, ACT 2600, Australia.

details by opening and closing morphological transformation approaches with a series of structuring elements (SEs). Several extension methods of MP, including morphological attribute profile (MAP) [20], extended MP (EMP) [21], and the extended MAP (EMAP) [22], are developed to extract different kinds of spatial structural information. In addition, another interesting approach, i.e., the 2-D-singular spectrum analysis (2-D-SSA), has been successfully applied for spatial feature extraction of images [23], [24]. Similar to its 1-D case (SSA), based on the singular value decomposition (SVD), it decomposes the given signal into a few components and reconstitutes an approximate description of the original signal using several components [25]. 2-D-SSA can extract the different spatial structural content and fully explore the global spatial correlation of each HSI image, while eliminates the image noise. However, it has several drawbacks, such as its low utilization of spectral features and the undersmoothed or oversmoothed results of land covers due to the fixed embedding window, which limits the popularization of 2-D-SSA. The integration solutions of 2-D-SSA may effectively solve these issues.

Recently, a number of joint spectral–spatial feature extraction methods are preferred for HSI classification, which can further improve the classification performance [26]. These methods can be categorized into two main groups: shallow structure models and deep learning models. The shallow structure models mainly use one or more simple feature extraction techniques to jointly extract spectral feature and single-scale or multiscale spatial features, such as spectral–spatial Gabor surface feature (GSF) fusion [27], intrinsic image decomposition (IID) [28], and SPCA-based Gaussian pyramid features (SPCA-GPs) [29]. Duan *et al.* [30] proposed a novel multiscale total variation (MSTV) method to make full use of multiscale spatial information and spectral features while reducing the influence of textures and noise. Hong *et al.* [31] apply a sequence of well-designed attribute filters to extract invariant features locally from HSI in both spatial and frequency domains, and these features are robust to local pixel or material changes. In [32], a global spatial and local spectral similarity-based manifold learning group sparse representation (GSLs) is proposed to make full use of the intrinsic features of HSI, achieving good classification results.

Deep learning models, which have been proven to be very useful for HSI feature extraction, can hierarchically extract the abstract and discriminative features using multiple non-linear networks [5]. Various deep learning networks have been designed, such as the stacked autoencoder [33], deep belief network (DBN) [34], and convolutional neural network (CNN) [35]. Due to the unique characteristics (e.g., local connections and shared weights) [36], CNN-based methods have been widely used for HSI spectral–spatial classification, such as 3-D CNN [37], ResNet [38], and its extensions spectral–spatial residual network (SSRN) [39] and multipath residual network (MPRN) [40]. Song *et al.* [41] proposed a novel deep feature fusion network (DFFN); combining with residual learning, it extracts features of different scales that can provide complementary yet correlated information for classification. Mu *et al.* [42] proposed a multiscale and multilevel spectral–spatial feature fusion network (MSSN) to

fully integrate spatial and spectral features in different scales, resulting in superior classification performance.

Although deep learning-based methods could exploit high-level features and have made remarkable progress in HSI classification, there are several drawbacks in HSI applications. One of the key challenges with deep modeling is the lack of insufficient training samples. The deep network structure used for feature extraction and classification is a time-consuming and error-prone process with a huge number of hyperparameters (e.g., learning rate and kernel sizes of convolutional filters). Furthermore, the extracted deep features can hardly be linked to any semantic meanings compared with handcrafted features. In other words, these deep features form a latent representation of the high-dimensional data, and thus, to some degree, they are human inexplicable. There is still a long way to go to study the interpretability of the deep learning model.

No matter shallow structured models or deep learning models, the extraction of the representative features of HSI is the key to improve the classification performance. Accordingly, several factors have been widely used in existing work for the extraction of representative features, including the joint utilization of spectral and spatial information, multiscale feature extraction, and the combination of the global and local features. In addition, considering the problems in the deep models, shallow models are still desirable because of their simple structures and fewer parameters. To this end, we aim to integrate several simple techniques for efficient and effective feature extraction and data classification of the HSI.

In this article, a fusion framework, PCA and SPCA domain multiscale 2-D-SSA, is proposed for this purpose. PCA and SPCA are used to extract overall and adjacent spectral correlation characteristics of objects, respectively, and dimension reduction. Besides, 2-D-SSA uses several embedding windows of different size to extract abundant multiscale spatial trend features by applying on the SPCA dimension-reduced images. Postprocessing via PCA is applied to reduce the dimension and enhance the differences of spectral pixels. The multiscale 2-D-SSA spatial features, as well as overall spectral features obtained by PCA, are fused as final spectral–spatial features (called MSF-PCs), as the input of SVM classifier. The major contributions of this method can be concluded as follows.

- 1) By combining PCA, SPCA, and 2-D-SSA, a novel approach is proposed to extract more concentrated spectral–spatial features of the HSI.
- 2) With the multiscale 2-D-SSA, spatial characteristics of complex scenes at different scales can be extracted for more accurate classification.
- 3) The global and local features of HSI are taken into account simultaneously to explore the interconnections between pixels or objects, which features multiscale 2-D-SSA for extracting spatial features and PCA/SPCA for spectral feature extraction.
- 4) The proposed method provides superior classification performance to several state-of-the-art methods,

including CNN-based approaches, even with a small number of training samples.

The remainder of this article is organized as follows. The principles of PCA, SPCA, and 2-D-SSA are briefly reviewed in Section II. Section III describes our proposed method. Experimental setup, results, and analysis of the four HSI data sets are given in Section IV and V, respectively. Finally, some concluding remarks are drawn in Section VI.

II. RELATED WORKS

As the three main techniques adopted in our approach, PCA and its expanded version SPCA, as well as 2-D-SSA, are briefly introduced in the following two sections.

A. PCA and SPCA

As a standard linear transformation method, PCA has been widely used for data dimension reduction in HSI. The main processing of PCA is to construct a linear transformation matrix to preserve most spectral features of the HSI in the principal components (PCs) of fewer dimensions.

Given an HSI data $H \in \mathfrak{R}^{B \times N}$ with N pixels and B bands, though mean-adjusted, this matrix is first used for the covariance matrix C calculation, which is defined as

$$C = \overline{HH^T} \in \mathfrak{R}^{B \times B} \quad (1)$$

where \bar{H} is the mean-adjusted matrix H . The eigendecomposition is applied for the covariance matrix C to obtain the first k ($k < B$) eigenvectors corresponding to bigger eigenvalues. These eigenvectors constitute the linear transformation matrix M with the size of $B \times k$. This linear transformation can be written as

$$Y = M^T H \in \mathfrak{R}^{k \times N} \quad (2)$$

where Y is the dimension-reduced matrix after PCA.

Unlike conventional PCA, segmented-PCA (SPCA) is proposed to consider the local useful spectral characteristics of the HSI. In SPCA, the original HSI data are first partitioned into K subgroups, which consists of highly correlated bands. PCA is conducted separately on each band subgroup. By selecting several main components from each subgroup, it can avoid the low correlations between the highly correlated blocks and reduce the dimension while essentially preserve the important information in each subgroup of bands. The components selected can be regrouped and transformed again to compress the data further. Further details can be found in [17].

B. 2-D-Singular Spectrum Analysis

As a model-free technique to analyze 1-D time series, singular spectrum analysis (SSA) can extract different components of given signals, including trend, oscillations, and noise, leading to smoothing and denoising results. Its 2-D extension version, 2-D-SSA, is first presented in [43], which is believed to have similar capabilities with SSA. In 2010, Golyandina and Usevich [44] enriched the algorithm and theory of 2-D-SSA and popularized its application, especially in images.

Given a grayscale with a size $h \times w$, define the embedding window of size $u \times v$ (with $1 \leq u \leq h$ and $1 \leq v \leq w$), which

moves from the top left to the bottom right of the image to construct the trajectory matrix T . The pixels in the window are expanded and joined as a vector $V \in \mathfrak{R}^{uv \times 1}$ and as a column in the trajectory matrix, which is shown as follows:

$$T = \begin{pmatrix} V_{1,1}, V_{1,2}, \dots, V_{1,w-v+1}, V_{2,1}, \dots, V_{h-u+1,w-v+1} \\ \vdots \\ V_{u,w \times (h-u+1)(w-v+1)} \end{pmatrix} \in \mathfrak{R}^{u \times (h-u+1)(w-v+1)}. \quad (3)$$

Note that the trajectory matrix T has a structure called HbH, i.e., Hankel by Hankel.

Then, the eigenvalues of TT^T and their corresponding eigenvectors are denoted as $(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L)$ and (U_1, U_2, \dots, U_L) , respectively. The trajectory matrix can be written as follows:

$$T = T_1 + T_2 + \dots + T_L \\ (T_i = \sqrt{\lambda_i} U_i V_i^T, V_i = X^T U_i / \sqrt{\lambda_i}) \quad (4)$$

where U_i and V_i are the empirical orthogonal functions and the PCs of the trajectory matrix, respectively. Afterward, we select T_1 as an approximation to T , mainly because it contains the most important spatial information that benefits for classification [23]. Finally, the matrix T_1 is converted to a new image of size $h \times w$ again, as the reconstructed image, by a two-step diagonal averaging process in the matrix antidiagonals in both each block and between blocks.

2-D-SSA has several advantages. First, it simultaneously considers the local and global spatial relevant information on remote sensing images. Different Hankel blocks involve horizontal or vertical spatial information of images, while lagged vectors contain local spatial characteristics of ground objects. This local information is put into the HbH matrix structure together to preserve the global spatial correlation. In addition, 2-D-SSA can effectively reduce the influence of serious image noise, and its reconstructed images by the key component have shown the great noise robustness both in HSI and non-HSI images [23], [45].

III. PROPOSED METHODOLOGY

The flowchart of our proposed spectral-spatial HSI classification method is shown in Fig. 1, which consists of three following steps. First, PCA and SPCA are applied to extract different spectral correlation characteristics and reduce the dimension of HSI, respectively. Then, the dimension-reduced images obtained by SPCA are used to extract and excavate spatial features via multiscale 2-D-SSA. Finally, the multiscale spatial features, as well as spectral features obtained by PCA, are fused to MSF-PCs features, and the SVM classifier is adopted to evaluate its performance.

A. Spectral Feature Extraction

In HSI, the object identification information contained in the spectral profile is beneficial for classification. However, the narrow hyperspectral bands and wide sensitization range have caused a huge dimension, which brings higher computational complexity. In order to reduce the feature dimension and preserve the expression ability of the feature, PCA is employed to extract global spectral features. The spectral

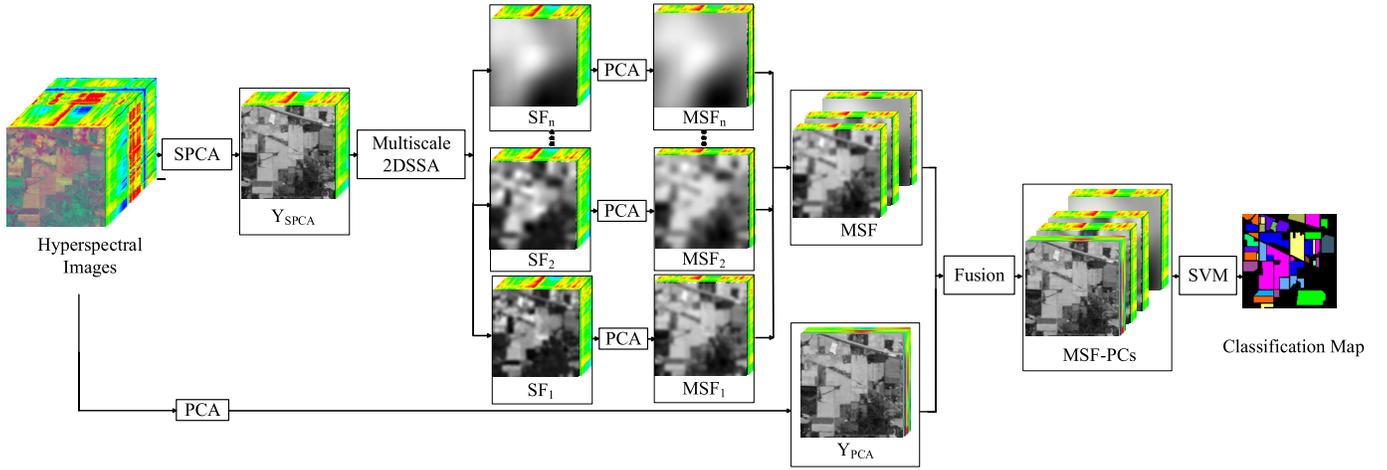


Fig. 1. Flowchart of our proposed spectral-spatial classification method. Generally, the proposed method consists of three main steps, i.e., spectral feature extraction, 2-D-SSA-based multiscale spatial feature extraction, and feature fusion and classification.

feature constituted by the former PCs after PCA can be fused with the spatial features for classification.

However, the global spectral feature obtained by PCA ignores the diversity of spectral profiles in a different wavelength range. In other words, PCA can often fail to extract the local useful characteristics of the HSI. Using a unified projection in dimension reduction for an entire spectral vector is inappropriate. In addition, there are also highly relevant characteristics in adjacent spectral bands, which has brought information redundancy. Therefore, SPCA is employed to solve the abovementioned problems and extract local spectral features. In SPCA, highly correlated bands are selected as the subgroups based on pairwise separability measures, such as the Bhattacharyya distance, while this type of band selection usually requires a time-consuming clustering process. Based on the fact that the correlations between neighboring spectral bands are generally higher than for bands further apart [17], we use a simple averaging method to select subsets.

Denote $X \in \mathfrak{R}^{W \times H \times B}$ as an HSI data cube, where W and H represent the rows and columns of data, the B bands are first partitioned equally into K subsets, and each subset X_{sub} is given as follows:

$$X_{\text{sub}}^k = \begin{cases} \{X^{(k-1)[B/K]+1}, \dots, X^{k[B/K]}\} & k[B/K] \leq B \\ \{X^{(k-1)[B/K]+1}, \dots, X^B\} & k[B/K] > B \end{cases} \quad (k = 1, 2, \dots, K) \quad (5)$$

where $[B/K]$ represents the value of the smallest integer greater than B/K . X^i is the i th band of the HSI data. Then, PCA is employed to each adjacent subset, and the first PCs of each band subsets, which contains the main information of local adjacent bands, are stacked together as local spectral features. SPCA concentrates several adjacent spectral contents and preserves the diversity of spectral features, which is beneficial to the subsequent spatial feature mining. Besides, it also reduces the data dimension and eliminates spectral redundancy, which can significantly improve the efficiency of the following-on analysis.

For convenience, the spectral features by PCA and SPCA are denoted as

$$Y_{\text{PCA}} = \text{PCA}(X, P) \in \mathfrak{R}^{W \times H \times P} \quad (6)$$

$$Y_{\text{SPCA}} = \{\text{PCA}(X_{\text{sub}}^1, 1), \dots, \text{PCA}(X_{\text{sub}}^K, 1)\} \in \mathfrak{R}^{W \times H \times K} \quad (7)$$

where $P (P \ll B)$ and $K (K \ll B)$ represent the number of dimension after PCA and SPCA, respectively.

B. 2-D-SSA-Based Multiscale Spatial Feature Extraction

After SPCA, the retained K images have more concentrated spectral diversity characteristics and lower feature dimensions compared with raw HSI. 2-D-SSA with different sizes of embedding windows [23] is applied to Y_{SPCA} , which can extract more spatial features than raw single-band image, as well as spend less computing time.

The 2-D-SSA with embedding window size $u_l \times v_l$ performs on Y_{SPCA} to obtain the l th scale spatial features (SF_l), which is calculated by

$$SF_l = 2\text{D-SSA}(Y_{\text{SPCA}}, u_l \times v_l) \in \mathfrak{R}^{W \times H \times K} \quad (l = 1, 2, \dots, n) \quad (8)$$

where n is the number of total scales for different embedding windows. The size of SF_l is exactly the same as Y_{SPCA} . The multiscale 2-D-SSA can capture the multiscale spatial information of the hyperspectral image effectively. Fig. 2 shows the false-color image constituted by the first three PCs in each 2-D-SSA scale. It can be seen the diversity of spatial features and the discrimination information in different extraction scales. The spatial difference information is stacked to multiscale spatial features, which can supplement the feature information of different land covers.

However, the dimension of stacked features from different scales will have an inevitable increase in the computational cost, as well as redundant information. In addition, the resulted images obtained by different 2-D-SSA scales are continuous smooth, and some texture information is removed, which has a negative effect on the classification performance because the

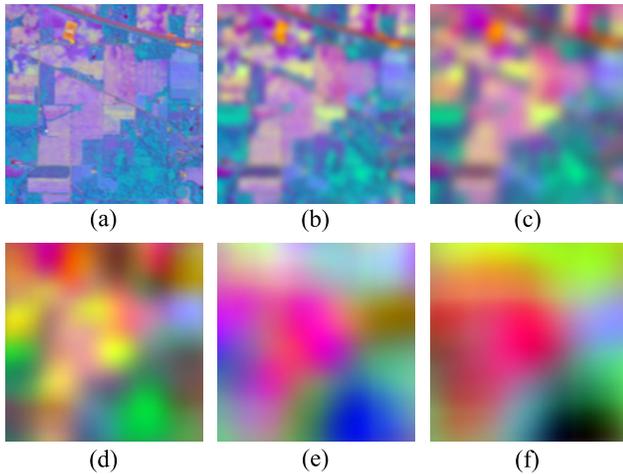


Fig. 2. Feature images obtained by different embedding windows of 2-D-SSA. (a) False-color image of original HSI data. (b)–(f) Feature images with embedding window scales 5×5 , 10×10 , 20×20 , 30×30 , and 40×40 , respectively.

differences of pixels belonging to different classes are reduced distinctly. Therefore, PCA is applied to featured images at each scale again to solve the problem presented earlier. Let MSF_l represents the first L components after applying PCA on the l th scale feature image SF_l as follows:

$$MSF_l = \text{PCA}(SF_l, L) \in \mathfrak{R}^{W \times H \times L} \quad (l = 1, 2, \dots, n) \quad (9)$$

where L ($L \leq K$) represents the number of dimensions after PCA. The final multiscale spatial features can be stacked as

$$MSF = \{MSF_1, \dots, MSF_n\} \in \mathfrak{R}^{W \times H \times (n \times L)}. \quad (10)$$

Compared with SF_l , the MSF_l feature further highlights the local scene differences in different scales because of the ability of PCA in emphasizing the spectral differences among pixels, which is beneficial for classification. The final stacked features can further enrich the spatial features of ground objects.

C. Feature Fusion and Classification

In this section, the obtained multiscale spatial features MSF , as well as PCA-based spectral features Y_{PCA} , are fused to MSF -PCs features, which can be represented as

$$MSF\text{-PCs} = \{MSF, Y_{\text{PCA}}\} \in \mathfrak{R}^{W \times H \times (nL+P)}. \quad (11)$$

For any pixel vector in MSF -PCs, it not only has the main spectral feature of raw spectral profile but also the multiscale concentrated features containing neighborhood spatial and adjacent band information. Multiscale concentrated features can reduce the misclassification caused by similar features on a single scale and improve the identification ability to determine the corresponding label.

In addition, selecting a suitable classifier is crucial for the performance assessment of the abovementioned features, especially with a small training sample size in hyperspectral images. Among the various pixelwise classifiers, SVM is a widely used supervised statistical learning framework. By using a kernel function, it can map the data to the higher dimensional space via a nonlinear transformation, aiming to

find the optimal hyperplane to separates samples belonging to different classes. Because of robustness to the variation of data dimensions, SVM has shown an outstanding performance in the classification of HSI [46]–[48]. Therefore, we use the fused feature vectors as input to the SVM classifier to evaluate its effectiveness for classification.

IV. EXPERIMENTAL SETUP

In this section, we describe the details of three benchmark HSI data sets and experimental setup. Furthermore, the parameters' sensitivity in the proposed method is analyzed.

A. Data Sets

Four hyperspectral data sets, i.e., Indian Pines, Kennedy Space Center (KSC), Pavia Centre (PaviaC), and DFC2018, are utilized in our experiments.

1) *Indian Pines*: This data set was acquired by the Airborne Visible/Infrared Imaging Spectrometer sensor that has a spectral coverage ranging from 0.4 to 2.5 μm . It covers the Indian Pines study site in Northwest Indiana, USA. This data set has a spatial size of 145×145 pixels with a spatial resolution of 20 m per pixel and 220 spectral bands. In our experiment, the number of bands is reduced to 200 by removing 20 water absorption bands (104–108, 150–163, and 220). The color composite image and the ground-truth map, which contains 16 land cover classes, are shown in Fig. 6(a) and (b).

2) *Kennedy Space Center*: This data set was acquired by the Airborne Visible/Infrared Imaging Spectrometer sensor over the KSC, Florida, on March 23, 1996. This data set contains 512×614 pixels with a spatial resolution of 18 m per pixel and 224 spectral bands in the wavelength range of 0.4–2.5 μm . After removing water absorption and low SNR bands, 176 bands were used for the analysis. The color composite image and its ground-truth image, which contains 13 land cover classes, are presented in Fig. 7(a) and (b).

3) *Pavia Centre*: This data set was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor during a flight campaign over Pavia, Italy, with a spectral coverage ranging from 0.43 to 0.86 μm . This data set contains 103 bands of size 1096×1096 pixels with a spatial resolution of 1.3 m per pixel. In our experiment, a subscene of 510×490 , including all nine land cover classes, is employed; its color composite image and ground-truth image are shown in Fig. 8(a) and (b).

4) *DFC2018*: This data set was distributed by the 2018 IEEE GRSS Data Fusion Contest (DFC). It was acquired by the IRTES CASI-1500 sensor at a GSD of 1 m over the campus of the University of Houston and its surrounding areas, in Houston, TX, USA. It contains 601×2384 pixels with 50 spectral bands sampling the wavelength of between 380 and 1050 nm at intervals of 10 nm. Its color composite image and ground-truth image, which contains 20 land cover classes, are shown in Fig. 8(a) and (b).

B. Parameters Sensitivity Analysis

In order to extract abundant spatial features, the scales of 2-D-SSA are important. As suggested experience in [23]

TABLE I

CLASSIFICATION RESULTS OF DIFFERENT 2-D-SSA SCALES ON THE RAW DATA SET OF THE THREE DATA SETS, IN WHICH 1%, 1%, 0.2%, AND 0.1% SAMPLES PER CLASS ARE USED FOR TRAINING IN CLASSIFICATION, RESPECTIVELY

Datasets	Metrics	Raw	$SF_1(5 \times 5)$	$SF_2(10 \times 10)$	$SF_3(20 \times 20)$	$SF_4(30 \times 30)$	$SF_5(40 \times 40)$	Stacked SF
Indian Pines (1% for training)	OA	58.63	71.41	67.38	69.66	71.57	72.23	77.32
	kappa	0.52	0.67	0.63	0.66	0.68	0.68	0.74
KSC (1% for training)	OA	76.91	79.51	83.84	87.20	86.33	86.47	91.80
	kappa	0.74	0.77	0.82	0.86	0.85	0.85	0.91
PaviaC (0.2% for training)	OA	88.69	92.73	91.69	88.61	82.87	81.57	94.77
	kappa	0.86	0.91	0.90	0.86	0.79	0.78	0.94
Datasets	Metrics	Raw	$SF_1(3 \times 3)$	$SF_2(5 \times 5)$	$SF_3(10 \times 10)$	$SF_4(20 \times 20)$	$SF_5(30 \times 30)$	Stacked SF
DFC2018 (0.1% for training)	OA	69.64	70.75	73.19	73.30	72.02	71.29	76.58
	kappa	0.60	0.61	0.65	0.65	0.64	0.62	0.69

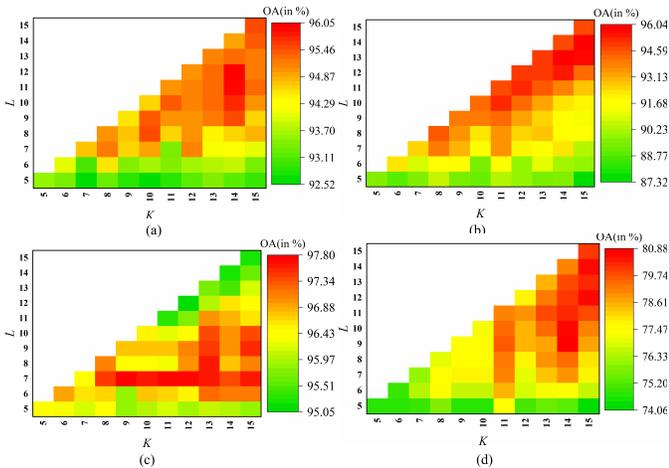


Fig. 3. Influence of the parameters K and L on classification results. OAs obtained on (a) Indian Pines, (b) KSC, (c) PaviaC, and (d) DFC2018 data sets.

and [24] and considering the fact that larger embedding window causes the computation time to grow exponentially, the 2-D-SSA scale sets $u_l \times v_l = \{5 \times 5, 10 \times 10, 20 \times 20, 30 \times 30, 40 \times 40\}$ are adopted for the first three data sets, while $u_l \times v_l = \{3 \times 3, 5 \times 5, 10 \times 10, 20 \times 20, 30 \times 30\}$ for the DFC2018 data set. The reason is that the large image with a big embedding window will cause the trajectory matrix to be huge and exceed the memory limit, which is also the defect of 2-D-SSA itself. In this experiment, the classification results of raw data, SF_l features in five scales above, and stacked SF_l features are compared by the SVM classifier, which is given in Table I. As we can see, the single scale of 2-D-SSA has an improvement in classification to some degree compared to the raw HSI data set. The combination of different scales can further improve the accuracy by about 2%–13%, achieving the best classification performance. The stacked spatial features could fully characterize different land-cover features, which demonstrates the advantages of multiscale analysis.

For effective dimension reduction, the combination of the number of subsets K and the number of reduced-dimension L is critical, where $L < K$. In this section, another set of experiments is designed to select the optimal K and L values on the four data sets. Fig. 3 shows the obtained overall accuracy (OA) of classification with different parameters. According to the

experience in [29] and our experiments, the parameters K and L are set to vary in a range from 5 to 15. The training samples are selected randomly, and the numbers used for training are 2%, 2%, 0.2%, and 0.1% of the four data sets, respectively. As can be seen in Fig. 3, the superior accuracy (highlighted in red) is mainly concentrated in the second half of the K and the middle part of the L value on all data sets. With the increase in K , more subsets containing local spectral features can be retained, while the OA is greatly improved. The classification accuracy is optimal at $K = 13$ or $K = 14$ on average, where a larger K will bring a significant increase in running time but limited gain in OA. In addition, preserving most of the subsets, i.e., L is slightly lower than K , usually achieves the best OA except on the PaviaC data set. On the PaviaC data set, a small L usually has better OA, mainly because it concentrates the main spectral discrimination features of the data, removing redundant and useless information. In order to obtain satisfactory classification results and ensure processing efficiency, the parameters K and L are fixed as 11 and 9, respectively, on all the four data sets in the following experiments.

In addition, according to the abovementioned optimal dimension-reduced parameters, we further evaluate the effectiveness of the SPCA in (7) and PCA in (9) (see Table II). In the spectral dimension reduction step in (7), three methods, including PCA, dominant set extraction-based band selector (DSEBS) [49], and FPCA [18], are adopted, in which the number of bands in the dimension-reduced data is fixed as $K = 11$ for a fair comparison. In the feature dimension reduction step in (9), another three methods, LDA [14], locally linear embedding (LLE) [50], and ICA [13], are used for comparison with PCA, and their dimension-reduced bands are all $L = 7$. As shown in Table II, the combination of SPCA and PCA achieves the highest accuracies, which demonstrates that, first, SPCA is effective for the spectral dimension reduction than others, mainly because it concentrates the spectral characteristics of adjacent bands; then, PCA not only further reduces the dimension of spatial features but also highlights the distinction of pixels.

Finally, the parameter P determines the number of PCs in spectral features obtained by PCA. An appropriate number of PCs can provide useful spectral characteristics while avoiding abundant information. Noisy information and high dimension problem would not be introduced into the finally derived spectral-spatial features. The parameter P is fixed as 3 in

TABLE II
COMPARISON OF CLASSIFICATION RESULTS OF DIFFERENT DIMENSION REDUCTION METHODS ON THE INDIAN PINES DATA SET WITH 2% OF SAMPLES PER CLASS FOR TRAINING

Metrics	Spectral dimension reduction in Eq. (7)			Feature dimension reduction in Eq. (9)			SPCA+PCA
	PCA	DSEBS	FPCA	LDA	LLE	ICA	
OA	88.24	93.18	93.65	89.10	90.26	92.80	94.90
kappa	0.87	0.93	0.93	0.88	0.89	0.92	0.94

TABLE III
LIST OF PARAMETERS FOR THE PROPOSED METHOD AND OTHER COMPARED ONES ON THE TEST DATA SETS

Methods	Indian Pines	KSC	PaviaC	DFC2018
Raw	---	---	---	---
IID(2015)			$M=32, Z=4, r=2$	
RLMR(2017)			$d=50, k=80$	
SPCA-GPs(2018)	$K=10, P=8$	$K=10, P=6$	$K=10, P=8$	$K=10, P=4$
MSTV(2019)			$K=20, N=30;$ $(\lambda_1 = 0.003, \sigma_1 = 2), (\lambda_2 = 0.02, \sigma_2 = 1),$ and $(\lambda_3 = 0.01, \sigma_3 = 3)$	
LAJSR(2019)			$M=45, L=200, K=60$	
IAPs(2020)	$n_s=3, r=[2, 4, 6], m=[0, 1, 2, 3], d=30,$ and $n_g=5$	$n_s=3, r=[2, 4, 6], m=[0, 1, 2, 3], d=30,$ and $n_g=4$	$n_s=3, r=[2, 4, 6], m=[0, 1, 2, 3], d=40,$ and $n_g=5$	$n_s=3, r=[2, 4, 6], m=[0, 1, 2], d=30,$ and $n_g=4$
GSLs(2020)	$d=20, v=150, q=4, t=9$	$d=30, v=100, q=4, t=9$	$d=20, v=150, q=3, t=9$	$d=20, v=100, q=3, t=9$
Proposed			Scale: $\{5 \times 5, 10 \times 10, 20 \times 20, 30 \times 30, 40 \times 40\}; K=11, L=7, P=3$	Scale: $\{3 \times 3, 5 \times 5, 10 \times 10, 20 \times 20, 30 \times 30\}; K=11, L=7, P=3$

this article, and more detailed results are given in Section S.I in the Supplementary Material.

C. Experimental Setup

In order to evaluate the performance of our proposed method, seven state-of-the-art HSI feature extraction approaches are adopted as compared methods. These include IID [28], robust local manifold representation (RLMR) [15], SPCA-GPs [29], MSTV [30], local adaptive joint sparse representation (LAJSR) [51], invariant attribute profiles (IAPs) [31], and GSLs [32] while the raw HSI data set as a baseline method (abbreviated as ‘‘Raw’’). Note that these methods use the optimal parameters according to different data sets. For our proposed method, the parameters are the same as described in Section IV-B. The optimal parameters of all involved methods are summarized in Table III. The SVM classifier is used in the stage of classification, which is implemented by the LIBSVM library [52] using an RBF kernel with fivefold cross validation.

For avoiding systematic errors and reducing random discrepancies, all experiments were independently carried out ten times. The training and testing samples sets were sampling randomly without any overlapping each time, using an equal sample rate or sample quantity for each class. In addition, different numbers of training samples are used in experiments to fully investigate the performance of the involved methods. In experiments, we mainly concern the case of small sample size, and the number of selected training samples varies within $\{1\%, 1.5\%, 2\%, 2.5\%, 3\%\}$ for Indian Pines, $\{3, 5, 7, 9, 11, 13\}$ for both KSC and PaviaC data sets per class, and $\{0.05\%, 0.1\%, 0.15\%, 0.2\%, 0.25\%, 0.3\%\}$ for the DFC2018 data set. In addition, five objective quality indexes, i.e., the OA, the average accuracy (AA), the kappa coefficient, class-by-class

accuracy, and running time, are utilized to evaluate the performance of the image classification. All the experiments are performed using MATLAB 2017a on a computer with a 3.5-GHz CPU and 8-GB memory.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the effectiveness of the fused features is validated using four HSI data sets discussed in Section IV, in a number of experiments, presenting results that demonstrate the related benefits in HSI classification.

A. Classification Results and Analysis

In this section, the qualitative and quantitative results of the proposed method and the other eight compared methods are presented, using four different HSI data sets. The details are given as follows.

First, the classification results with a different number of training samples on three data sets are compared and shown in Fig. 4. It can be seen that the increase in the labeled sample amount for training has a positive promotion on the classification performance. In almost all cases, our proposed method has achieved the highest classification accuracies even in the small number of samples, demonstrating the robustness of the proposed method.

In order to quantitatively evaluate the superiority of the proposed method, the classification results from the four data sets are given in Tables IV–VII. In addition, the classification maps yielded by all compared methods and the proposed method on four data sets are shown in Figs. 5–8, respectively.

1) *Indian Pines*: As shown in Table IV, the proposed method obtains the highest values in terms of three metrics and most of the classes, including 100% accuracy in two classes and over 89% accuracy in all classes. Compared with the raw

TABLE IV

CLASSIFICATION ACCURACY (IN %) OF ALL COMPARED METHODS FOR THE INDIAN PINES DATA SET WITH 2% TRAINING SAMPLES PER CLASS

Class	Samples	Raw	IID	RLMR	SPCA-GPs	MSTV	LAJSR	IAPs	GSLs	Proposed
1	46	6.67(2.23)	97.78(0.00)	100.0(0.00)	96.30(1.28)	83.70(16.8)	88.89(9.69)	100.0(0.00)	96.30(4.63)	94.08(4.36)
2	1428	58.54(10.1)	86.97(2.72)	70.50(4.76)	87.33(5.37)	90.02(3.52)	89.97(1.46)	75.58(3.12)	88.94(3.02)	96.54(1.75)
3	830	45.47(7.59)	88.27(1.12)	79.58(2.38)	88.56(5.36)	89.79(4.93)	82.21(2.43)	71.57(1.06)	92.09(1.88)	94.06(0.47)
4	237	16.09(4.81)	87.65(13.2)	67.18(11.4)	82.47(15.4)	82.33(15.8)	91.24(0.66)	88.06(10.1)	90.23(7.22)	97.13(3.87)
5	483	67.58(1.05)	87.45(3.17)	96.51(2.69)	91.05(8.15)	88.16(19.9)	88.51(5.50)	93.88(7.22)	88.94(5.24)	95.42(1.22)
6	730	88.16(5.73)	98.51(1.86)	84.97(6.86)	99.21(1.03)	98.97(1.66)	99.86(0.14)	89.03(3.16)	96.08(1.19)	99.02(0.42)
7	28	30.86(27.8)	98.77(2.14)	95.48(0.50)	96.30(6.41)	98.77(2.14)	93.83(5.66)	100.0(0.00)	91.36(15.0)	100.0(0.00)
8	478	90.17(3.15)	100.0(0.00)	99.22(0.32)	100.0(0.00)	99.79(0.22)	99.93(0.1)	92.79(0.89)	96.44(2.89)	100.0(0.00)
9	20	17.54(13.3)	92.98(12.2)	87.77(6.45)	94.74(9.12)	100.0(0.00)	82.46(13.3)	100.0(0.00)	85.97(12.2)	96.49(6.08)
10	972	57.25(11.5)	90.16(1.42)	69.67(5.98)	86.21(9.36)	95.76(0.76)	88.10(3.22)	76.01(11.7)	86.38(6.59)	90.27(3.89)
11	2455	71.02(9.94)	90.02(0.31)	76.14(3.54)	95.70(3.29)	93.79(1.51)	84.34(1.31)	73.24(1.86)	94.17(1.56)	97.55(0.17)
12	593	28.97(3.74)	85.83(1.10)	83.48(5.91)	85.37(5.52)	87.09(10.5)	79.98(1.95)	74.30(18.0)	80.49(2.55)	89.50(7.69)
13	205	92.33(3.18)	99.33(0.29)	88.92(8.56)	98.00(0.87)	99.50(0.00)	99.00(0.00)	99.83(0.29)	96.17(4.16)	99.00(0.87)
14	1265	94.49(1.53)	96.53(5.46)	92.79(2.05)	100.0(0.00)	99.71(0.05)	97.44(1.58)	89.65(2.58)	99.30(0.41)	99.95(0.09)
15	386	20.90(1.65)	79.72(13.5)	74.24(9.90)	85.36(7.74)	92.15(8.25)	85.45(6.68)	82.89(7.17)	86.68(10.0)	89.51(11.4)
16	93	45.42(6.72)	90.11(7.20)	96.59(4.04)	97.07(3.66)	78.76(6.97)	98.53(1.68)	94.98(2.18)	98.17(2.29)	99.63(0.64)
OA		64.86(1.11)	90.78(0.66)	80.03(0.57)	92.58(1.50)	93.48(1.76)	89.37(0.78)	79.80(0.59)	92.06(0.05)	96.11(0.65)
AA		51.97(1.64)	91.88(1.53)	81.06(1.58)	92.73(0.92)	92.39(2.09)	90.61(1.47)	71.11(1.16)	91.73(1.04)	96.13(0.07)
kappa		0.596(0.009)	0.895(0.008)	0.771(0.007)	0.916(0.017)	0.926(0.020)	0.879(0.009)	0.767(0.006)	0.909(0.001)	0.956(0.007)
time(s)		5.61	80.02	60.58	1.83	12.95	193.86	6.49	152.36	12.73

TABLE V

CLASSIFICATION ACCURACY (IN %) OF ALL COMPARED METHODS FOR THE KSC DATA SET WITH SEVEN TRAINING SAMPLES PER CLASS

Class	Samples	Raw	IID	RLMR	SPCA-GPs	MSTV	LAJSR	IAPs	GSLs	Proposed
1	761	74.98(11.0)	80.28(10.8)	89.57(6.35)	82.71(11.4)	72.81(0.00)	95.15(4.95)	86.35(0.65)	91.91(7.73)	87.54(12.1)
2	243	70.48(12.7)	99.15(0.85)	86.46(4.53)	89.97(6.90)	90.96(7.93)	91.88(9.13)	69.16(1.03)	76.98(3.91)	99.72(0.49)
3	256	86.75(4.53)	97.86(0.61)	82.17(4.40)	96.65(3.80)	89.02(18.3)	98.01(1.05)	53.96(4.91)	93.44(1.81)	96.79(2.44)
4	252	54.29(10.0)	93.88(7.08)	63.80(9.76)	76.74(4.96)	100.0(0.00)	70.72(15.9)	20.63(4.80)	62.18(18.7)	94.42(3.79)
5	161	71.00(6.94)	88.10(6.82)	52.65(3.80)	99.35(0.65)	81.17(0.00)	92.73(8.17)	46.32(2.84)	63.42(6.51)	90.48(9.42)
6	229	55.10(17.0)	99.25(1.30)	50.85(9.26)	92.79(1.62)	100.0(0.00)	90.03(6.93)	68.71(3.54)	89.49(8.66)	94.45(5.79)
7	105	91.16(6.80)	100.0(0.00)	66.12(4.81)	96.60(5.03)	100.0(0.00)	100.0(0.00)	80.97(3.89)	87.07(7.24)	97.96(3.53)
8	431	74.13(19.9)	96.46(1.87)	89.57(2.10)	84.98(12.5)	90.96(7.83)	81.54(17.9)	68.26(1.79)	87.27(12.4)	96.93(3.97)
9	520	80.25(3.32)	78.04(24.8)	84.66(4.39)	89.54(17.1)	89.09(9.45)	94.56(5.06)	87.78(3.68)	89.73(10.1)	100.0(0.00)
10	404	69.69(11.5)	80.60(16.4)	89.52(6.80)	79.60(13.4)	89.17(18.8)	95.49(0.75)	85.46(3.12)	78.17(11.6)	74.48(15.8)
11	419	94.98(2.80)	95.87(4.54)	97.26(2.30)	92.96(5.08)	100.0(0.00)	96.30(6.21)	97.94(0.76)	96.68(5.11)	99.84(0.14)
12	503	79.10(3.06)	89.18(9.09)	94.78(5.05)	90.86(6.94)	96.04(5.69)	82.39(5.62)	82.61(1.06)	91.46(13.9)	100.0(0.00)
13	927	99.20(0.41)	84.78(22.8)	98.95(0.95)	97.72(1.54)	100.0(0.00)	100.0(0.00)	86.37(2.61)	99.89(0.19)	99.06(1.36)
OA		80.09(0.90)	88.16(2.25)	84.44(1.28)	89.59(2.47)	91.41(1.98)	92.33(1.78)	75.88(0.14)	88.93(0.49)	94.74(0.69)
AA		77.01(1.05)	91.04(0.55)	82.51(0.66)	90.04(1.81)	92.25(1.97)	91.44(1.38)	72.86(0.20)	85.21(0.51)	94.74(1.00)
kappa		0.779(0.010)	0.869(0.024)	0.827(0.014)	0.884(0.027)	0.905(0.022)	0.915(0.020)	0.732(0.001)	0.877(0.006)	0.942(0.008)
time(s)		13.71	1444.15	33.42	10.96	181.92	79.48	99.09	1027.30	135.89

TABLE VI

CLASSIFICATION ACCURACY (IN %) OF ALL COMPARED METHODS FOR THE PAVIAC DATA SET WITH SEVEN TRAINING SAMPLES PER CLASS

Class	Samples	Raw	IID	RLMR	SPCA-GPs	MSTV	LAJSR	IAPs	GSLs	Proposed
1	8092	100.0(0.00)	99.84(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)	99.88(0.06)	100.0(0.00)	99.84(0.22)
2	2758	81.15(5.61)	85.68(6.41)	87.57(4.44)	92.98(5.20)	92.79(1.37)	84.97(3.39)	97.14(1.30)	93.04(4.15)	93.86(4.18)
3	1065	94.52(2.85)	94.08(9.44)	85.35(11.6)	95.84(5.99)	87.52(2.34)	94.77(2.09)	70.93(6.14)	84.34(7.51)	95.56(2.29)
4	2140	81.70(15.5)	99.78(0.22)	96.73(2.16)	99.26(0.73)	99.84(0.15)	78.53(6.37)	83.35(9.37)	94.81(3.04)	100.0(0.00)
5	1737	73.06(8.08)	97.98(3.40)	87.48(14.2)	99.94(0.10)	96.57(5.45)	93.32(6.14)	93.89(3.06)	94.84(4.87)	99.88(0.20)
6	3694	87.17(6.00)	93.82(1.99)	89.79(7.36)	95.33(7.37)	97.11(0.37)	94.34(2.96)	89.81(3.22)	95.63(3.05)	98.03(2.27)
7	7288	65.21(6.16)	84.80(3.13)	75.86(2.69)	88.29(5.08)	83.17(2.91)	91.69(5.98)	99.13(0.72)	83.56(6.29)	96.61(2.75)
8	1123	95.73(3.61)	97.01(3.49)	99.43(0.83)	100.0(0.00)	94.92(8.56)	99.91(0.16)	98.31(0.59)	95.16(6.20)	99.97(0.05)
9	1767	99.92(0.03)	95.45(1.19)	99.92(0.03)	95.61(3.08)	96.21(1.43)	91.53(3.09)	96.82(1.47)	99.68(0.41)	96.86(1.69)
OA		84.84(3.74)	93.39(0.48)	90.12(1.45)	95.42(0.71)	93.76(0.27)	93.22(1.08)	94.67(1.40)	93.33(1.49)	97.95(0.91)
AA		86.50(4.66)	94.27(0.90)	91.35(1.44)	96.36(0.15)	94.24(1.01)	92.12(0.29)	94.88(0.52)	93.45(0.24)	97.85(0.71)
kappa		0.819(0.045)	0.921(0.006)	0.881(0.018)	0.945(0.008)	0.925(0.003)	0.918(0.013)	0.936(0.017)	0.920(0.018)	0.975(0.011)
time(s)		4.87	753.76	190.24	5.42	180.71	369.67	70.58	854.71	124.75

HSI data, the proposed MSF-PCs features can significantly improve the OA from 64.86% to 96.11%, which is about 3% higher than the other two multiscale feature extraction methods: SPCA-GPs and MSTV. The results also demonstrate our features are superior to the intrinsic image features in IID and spatial invariant features in IAPs. Similar to GSLs, our method also considers the global and local spectral and spatial information in HSI, while our method obtains better classification results in almost all classes.

As shown in Fig. 5, the proposed method shows satisfactory performances from classification maps. The raw data plus

SVM method appear obvious misclassification inside the land covers, which is manifested as classification noise. The RLMR and LAJSR methods reduce this spot-like misclassification by using neighborhood spatial information while still exist small lump or strip misclassification. These misclassifications are also exist in the IID method due to its single spatial feature. In contrast, the multiscale methods, including SPCA-GPs, MSTV, and IAPs, can effectively reduce those misclassifications, but they still suffer from incorrectly distinguishing the boundaries of land covers. As can be seen, our method has effectively solved the abovementioned problems because

TABLE VII

CLASSIFICATION ACCURACY (IN %) OF ALL COMPARED METHODS FOR THE DFC2018 DATA SET WITH 0.1% TRAINING SAMPLES PER CLASS

Class	Samples	Raw	IID	RLMR	SPCA-GPs	MSTV	LAJSR	IAPs	GSLs	Proposed
1	9495	72.48(16.8)	80.50(2.67)	76.96(11.7)	72.68(14.8)	69.37(6.20)	63.69(12.3)	51.01(10.2)	51.24(4.12)	69.53(9.31)
2	31930	86.65(0.66)	85.72(2.71)	86.38(0.45)	85.02(3.03)	84.20(6.02)	83.75(4.11)	91.98(1.57)	73.53(5.59)	85.43(1.67)
3	644	100.0(0.00)	94.61(9.20)	99.84(0.16)	68.07(16.2)	100.0(0.00)	100.0(0.00)	11.04(5.14)	100.0(0.00)	73.66(6.92)
4	13410	81.97(4.82)	88.94(6.26)	82.74(4.19)	88.52(3.34)	87.95(8.32)	82.25(15.4)	74.98(2.89)	71.84(6.99)	77.26(4.17)
5	4928	36.41(12.4)	42.40(8.66)	37.78(5.31)	44.52(7.26)	55.51(16.7)	36.94(3.88)	4.35(0.56)	30.95(8.76)	44.47(10.2)
6	4483	74.65(7.98)	70.28(3.40)	68.09(7.81)	92.01(3.94)	75.38(12.7)	72.79(20.8)	25.89(6.88)	96.64(3.82)	77.43(14.7)
7	238	67.79(0.48)	55.41(0.65)	68.07(0.65)	66.24(2.54)	59.78(7.81)	68.78(0.00)	22.08(3.68)	60.20(6.70)	55.42(1.22)
8	39505	71.76(2.96)	84.90(4.58)	72.25(3.68)	77.30(4.39)	87.04(4.27)	77.96(3.01)	66.92(0.77)	91.50(4.54)	87.34(5.53)
9	223996	85.58(0.34)	92.09(1.83)	86.95(0.40)	85.42(2.54)	93.71(0.75)	89.06(1.59)	98.91(0.37)	92.83(0.89)	92.74(0.74)
10	44944	42.65(2.34)	49.86(4.28)	43.74(4.23)	46.82(4.39)	57.50(6.36)	50.15(3.48)	33.89(2.74)	49.75(4.18)	59.04(8.02)
11	32133	31.10(2.97)	33.38(5.34)	35.24(0.63)	34.08(1.79)	41.12(2.15)	29.99(1.96)	20.73(2.22)	35.65(3.66)	43.14(5.04)
12	1517	5.26(3.46)	11.11(6.17)	2.68(1.07)	9.46(3.05)	11.07(10.7)	9.81(3.40)	5.55(0.16)	12.06(7.17)	7.41(2.25)
13	46783	50.49(4.38)	67.76(4.97)	51.67(1.61)	58.54(8.29)	73.45(4.71)	55.54(1.01)	59.68(1.85)	73.31(5.51)	78.28(4.63)
14	9882	53.79(8.11)	71.88(4.18)	58.83(5.14)	72.40(6.26)	76.63(5.14)	66.21(5.69)	25.90(3.99)	90.18(5.52)	77.93(9.35)
15	6927	66.14(5.97)	94.82(7.78)	76.47(2.59)	71.74(10.2)	87.50(12.2)	69.69(3.67)	67.66(2.47)	65.09(13.5)	85.96(3.43)
16	11437	60.23(5.90)	81.82(11.8)	62.97(6.52)	67.92(3.68)	71.76(3.47)	74.33(3.88)	71.25(6.60)	65.03(2.93)	78.70(8.99)
17	139	94.20(5.08)	91.06(15.5)	87.92(16.6)	100.0(0.00)	98.55(1.92)	100.0(0.00)	21.98(10.9)	86.23(19.1)	100.0(0.00)
18	6605	36.03(8.80)	68.79(8.56)	31.34(7.31)	46.31(7.00)	76.52(3.41)	30.80(7.17)	5.76(1.26)	39.93(11.7)	73.97(9.41)
19	5229	38.64(1.84)	74.06(3.01)	39.22(7.41)	68.75(12.9)	65.03(8.67)	62.56(7.04)	6.58(2.59)	61.93(5.36)	83.30(5.31)
20	6790	68.47(9.15)	86.38(6.56)	72.95(4.82)	71.45(15.1)	84.71(5.14)	82.23(5.03)	18.58(3.16)	98.67(2.31)	87.72(6.68)
OA		69.93(0.91)	78.88(0.47)	71.40(0.75)	73.01(0.56)	80.42(0.31)	73.69(0.91)	71.98(0.61)	77.95(1.01)	81.41(0.96)
AA		61.21(1.96)	71.29(1.11)	62.10(1.71)	66.36(2.51)	72.84(1.29)	65.33(1.41)	48.74(1.00)	67.33(1.00)	72.94(2.26)
kappa		0.607(0.012)	0.724(0.005)	0.627(0.009)	0.650(0.010)	0.753(0.005)	0.659(0.012)	0.597(0.008)	0.712(0.014)	0.768(0.013)
time(s)		30.526	6427.4686	3911.72	67.635	293.044	7994.65	533.07	6807.54	829.77

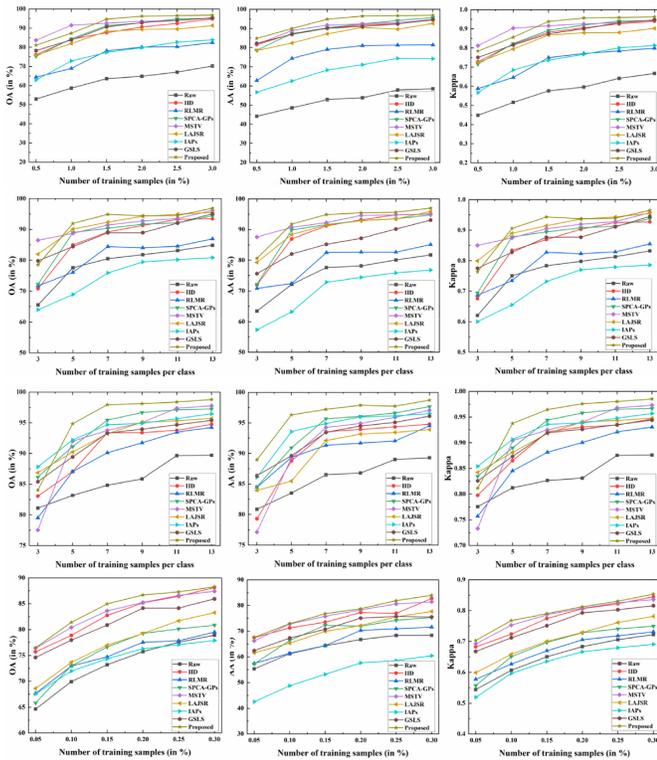


Fig. 4. Effect of various percentages (top/Indian Pines and bottom/DFC2018) or numbers (middle/KSC and PaviaC) of training samples per class on the classification results in terms of (Left) OA, (Middle) AA, and (Right) Kappa.

of the reasonable utilization of spectral and multiscale spatial features.

2) *KSC and PaviaC*: For the rest two data sets, there are some small and irregular land covers, such as trees, marshes, and roads, which bring difficulties for classification. From Tables V and VI, it can be seen that the classification accuracy of our method is still the highest among all compared methods in terms of OA, AA, and the kappa coefficient, with decent

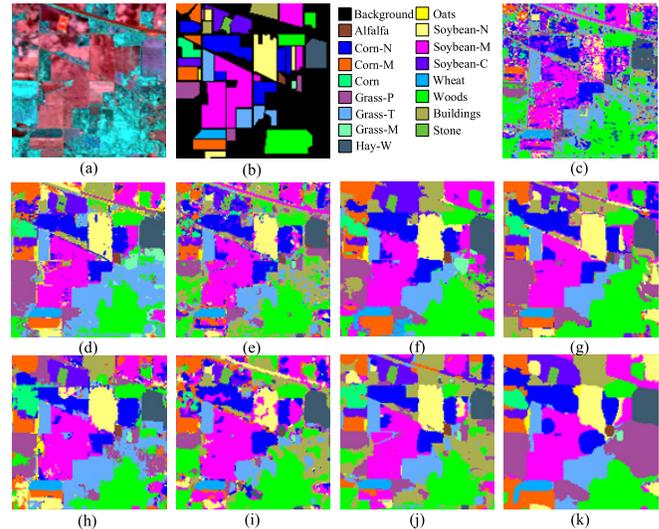


Fig. 5. Classification maps obtained by different methods on Indian Pines with 2% training samples per class. (a) Color composite image. (b) Ground-truth image. (c) Raw. (d) IID. (e) RLNR. (f) SPCA-GPs. (g) MSTV. (h) LAJSR. (i) IAPs. (j) GSLs. (k) Proposed method.

processing efficiency. The OA of our method improves about 14% and 13% compared with the baseline method on KSC and PaviaC, respectively. All the compared methods have different degrees of improvement on accuracies than baseline. However, the IAPs method has difficulty classifying hommock, oak, and slash pine, leading to lower classification accuracy than raw data.

Figs. 6 and 7 show the classification maps yielded by the compared methods and our method, in which our proposed method has the best performance on both two data sets. IID cannot distinguish meadows and trees well. The RLNR and LAJSR methods have spot and blocking artifacts in the resulting maps because of the limited utilization of spatial information. MSTV ignores some small or detailed land covers, resulting in a too smooth classification map. IAPs yielded

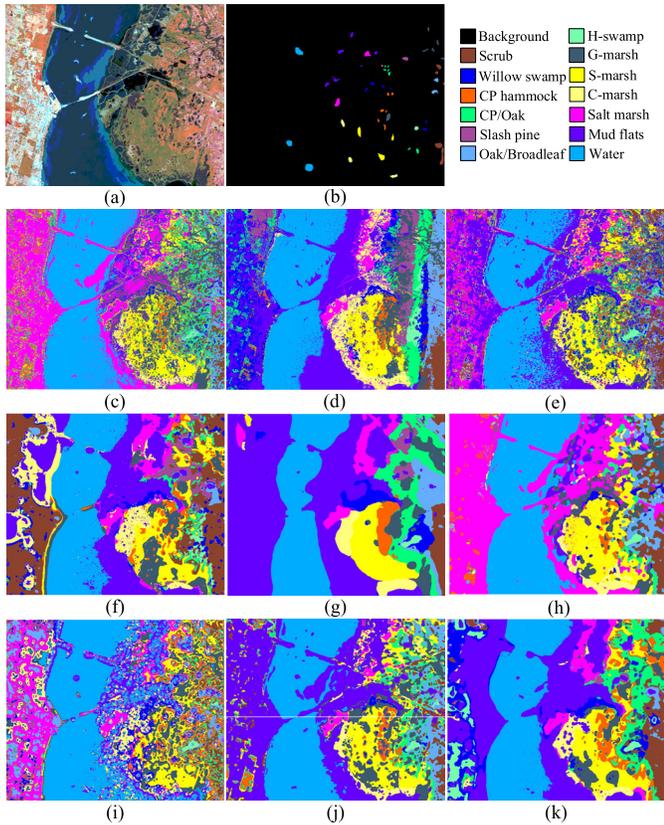


Fig. 6. Classification maps obtained by different methods on KSC with seven training samples per class. (a) Color composite image. (b) Ground-truth image. (c) Raw. (d) IID. (e) RLMR. (f) SPCA-GPs. (g) MSTV. (h) LAJSR. (i) IAPs. (j) GSLS. (k) Proposed method.

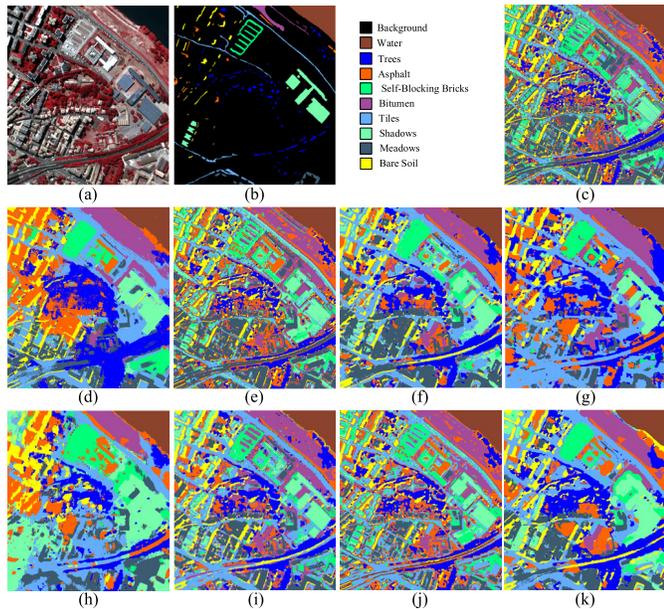


Fig. 7. Classification maps obtained by different methods on PaviaC with seven training samples per class. (a) Color composite image. (b) Ground-truth image. (c) Raw. (d) IID. (e) RLMR. (f) SPCA-GPs. (g) MSTV. (h) LAJSR. (i) IAPs. (j) GSLS. (k) Proposed method.

poor classification results containing irregular blocking artifacts on KSC, mainly because its extracted semantic features from complex land covers cannot correctly identify by the

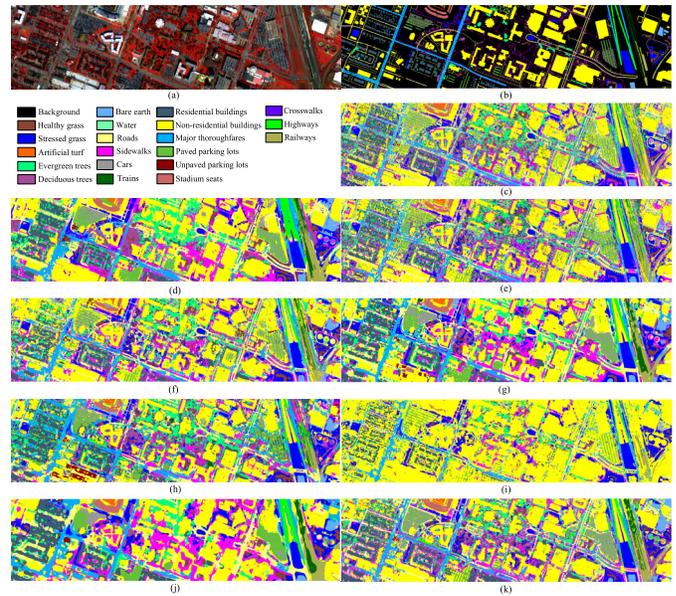


Fig. 8. Classification maps obtained by different methods on DFC2018 with 0.1% training samples per class. (a) Color composite image. (b) Ground-truth image. (c) Raw. (d) IID. (e) RLMR. (f) SPCA-GPs. (g) MSTV. (h) LAJSR. (i) IAPs. (j) GSLS. (k) Proposed method.

simple spectral-based classifier. GSLS does not distinguish well between oak and slash pine. In addition, most of these methods have edges misclassification problems between land covers. In contrast, the proposed method not only effectively eliminates the misclassification within land covers but also successfully preserves the boundary of different areas.

3) *DFC2018*: The quantitative results and corresponding classification maps of different compared methods on the DFC2018 data set are given in Table VII and Fig. 8, respectively. As observed in Table VII, the same conclusions can be drawn that the proposed method has good robustness and achieves the highest classification accuracies in challenging scenes. More specifically, the baseline method only holds a 69.93% result with the raw data due to the lack of spatial information. On the basis of spectral features, the RLMR and LAJSR combine local spatial similarity information and improve the OA to 71.40% and 73.69%, respectively. However, considering the local and ignoring global spatial characteristics can only bring finite accuracy improvement while increasing the processing time. IID fully excavates the intrinsic spatial features of complex land covers but takes an enormous amount of processing time. The accuracy of SPCA-GPs on this data is poor since it does not consider the global spectral and local spatial characteristics. The GSLS method takes into account the features of SPCA-GPs that are ignored and achieves 77.95% accuracy, but it is still limited by processing efficiency. MSTV utilizes multiscale spatial features, as well as global and local spectral features, and obtains good classification results (80.42%), but it is still lower than our method (81.41%).

Fig. 8 highlights the superiority of the proposed method by means of classification maps. Generally speaking, our method effectively eliminates the effects of salt-and-pepper and blocking noise from classification maps and preserves the shape and semantic structure of different objects, especially

TABLE VIII
CLASSIFICATION RESULTS (IN %) WITH EXTRACTED FEATURES UNDER DIFFERENT CLASSIFIERS ON THE THREE DATA SETS

Datasets	Metrics	Raw				MSF-PCs features			
		MLR	SRC	KNN	SVM	MLR	SRC	KNN	SVM
Indian Pines (2% training samples)	OA	52.24	60.25	59.32	65.54	71.57	81.36	94.12	96.24
	Kappa	0.46	0.55	0.53	0.60	0.68	0.79	0.93	0.96
KSC (7 training samples)	OA	49.55	78.40	70.82	80.51	83.85	88.36	94.92	94.88
	Kappa	0.43	0.76	0.68	0.78	0.82	0.87	0.95	0.94
PaviaC (7 training samples)	OA	71.46	81.77	84.34	85.46	91.19	87.33	91.00	97.06
	Kappa	0.65	0.78	0.81	0.83	0.89	0.85	0.93	0.96
DFC2018 (0.1% training samples)	OA	41.37	61.59	61.44	69.64	51.59	66.61	76.75	81.41
	Kappa	0.35	0.50	0.49	0.60	0.44	0.57	0.70	0.78

TABLE IX
COMPARISON OF CLASSIFICATION ACCURACY (IN %) OF OUR PROPOSED METHOD WITH DEEP LEARNING METHODS ON THE INDIAN PINES DATA SET WITH 10% TRAINING SAMPLES PER CLASS

Metrics	3D-CNN	ResNet	SSRN	MPRN	DFFN	MSSN	Proposed
OA	90.38(0.89)	96.92(0.59)	98.22(0.25)	98.71(0.21)	99.06(0.16)	99.12(0.15)	99.49(0.03)
AA	88.39(1.46)	96.73(0.85)	98.08(0.59)	98.34(0.60)	98.24(0.50)	99.23(0.60)	99.57(0.17)
Kappa	0.890(0.010)	0.965(0.006)	0.980(0.009)	0.983(0.005)	0.989(0.002)	0.990(0.002)	0.994(0.001)

for the roads and buildings. It validates that the extracted feature contained in MSF-PCs can improve the identifiability and discrimination of different pixels.

In terms of processing efficiency, our method is not the most efficient but still faster than most of the compared methods. The efficiency of our method is only lower than SPCA-GPS and IAPs on three data sets, while, on the KSC data set, the time cost of RLMR and LAJSR is less than our method because they only execute on the labeled pixels, while the latter processes all pixels in HSI. In addition, the computational complexity of the proposed method is also analyzed, and the details are given in Section S.IV in the Supplementary Material.

B. Evaluation With Different Classifiers

Although the SVM classifier is utilized in the proposed method, the relatively weak classifier can verify the efficacy of the extracted features as hidden in the results [23]. To this end, another three classifiers, KNN [8], MLR [7], and SRC [9], are adopted to verify the efficacy of MSF-PCs features from the proposed model. With default parameter settings as suggested in [7]–[9], the classification results of these three classifiers and SVM performed on the raw data and MSF-PCs features are produced, which are shown in Table VIII for comparison.

As seen in Table VIII, it is obvious that the classification results obtained by SVM are superior to those obtained by the other three classifiers, and the extracted MSF-PCs features achieve a significant improvement in the classification accuracy than the raw data of HSI under all four classifiers. Take the kappa metric as an example, under the SVM classifier, the accuracies of extracted features are improved by 0.36 for Indian Pines, 0.16 for KSC, 0.13 for PaviaC, and 0.18 for DFC2018. While the accuracy improvement are 0.4, 0.27, 0.12, and 0.21 under the KNN, 0.2, 0.39, 0.24, and 0.09 under the MLR, and 0.24, 0.17, 0.07, and 0.07 under the SRC, for the four data sets, respectively. It can be seen that, under weak classifiers, the improvement of kappa from raw data to the MSF-PCs features is higher than the SVM classifier on

some data sets. As the generalization capacity of the weak classifier is limited, the significant accuracy improvement can be mainly attributed to the superior discrimination ability of the MSF-PCs features. These results again highlighted the effectiveness of the extracted features for classification.

C. Comparison With Deep Learning Approaches

In this section, several CNN-based spectral–spatial feature extraction methods are chosen for comparison, which includes 3-D-CNN [37], ResNet [38], SSRN [39], MPRN [40], DFFN [41], and MSSN [42]. This experiment is carried out on the Indian Pines, where 10% of samples per class are selected for full training of the deep network. The classification accuracies of all methods are listed in Table IX. As we can see, compared with conventional 3-D-CNN and ResNet that simply use limited spectral and spatial information, DFFN and MSSN further utilize multiscale and multilevel feature extraction layers, obtaining higher classification accuracies. Our proposed method excavates the multiscale features of HSI with a shallow structure, and it has achieved the best results in terms of three metrics. It can be concluded that the MSF-PCs features using fewer parameters are comparable and superior to those mined by deep networks. Section S.II in the Supplementary Material gives more details.

D. Classification With Spatially Disjoint Samples

Considering the fact that some works [53]–[55] have indicated that traditional random sampling strategy usually leads to an improper assessment of different methods, in this section, we use spatially disjoint samples to acquire more realistic classification results and more accurate assessment of the models. The classification results on two sampling strategies of all compared methods in Section IV-C are summarized in Table X. As can be seen, the proposed method still achieves superior accuracies with region samples, which is especially significant in small training samples. Besides, the accuracy differences between random sampling and region sampling of our method are relatively small among all spectral–spatial

TABLE X
COMPARISON (IN TERMS OF OA) BETWEEN DIFFERENT HSI
CLASSIFICATION METHODS TRAINED USING SAMPLES
FROM RANDOM SAMPLING AND REGION SAMPLING

Methods	Random sampling		Region sampling		Diff	
	2%	10%	2%	10%	2%	10%
SVM	64.47(1.72)	79.59(0.79)	54.34(4.34)	67.82(3.34)	10.13	11.77
2DSSA	81.43(0.92)	96.26(0.42)	58.94(1.00)	67.34(2.90)	22.49	28.92
IID	90.78(0.66)	98.08(0.15)	71.54(1.88)	78.38(2.01)	19.24	19.70
RLMR	80.03(0.57)	87.04(0.19)	52.23(2.78)	58.46(1.35)	27.80	28.58
SPCA-GPs	92.58(1.50)	98.50(0.48)	70.57(1.76)	76.04(2.97)	22.01	22.46
MSTV	93.48(1.76)	98.19(0.17)	78.63(0.56)	83.92(0.12)	14.85	14.27
LAJSR	89.37(0.78)	97.57(0.14)	72.69(2.53)	80.42(1.52)	16.68	17.15
IAPs	79.80(0.59)	93.79(0.37)	58.10(2.41)	73.28(1.23)	21.70	20.51
GSLS	92.06(0.05)	98.17(0.20)	74.06(3.02)	79.25(2.25)	18.00	18.92
Proposed	96.11(0.65)	99.49(0.03)	79.54(1.41)	83.04(1.40)	16.57	16.45

methods, which has demonstrated the validity and robustness of features extracted by our method. Section S.III in the Supplementary Material gives more details.

E. Discussion

In HSI, the multiscale spectral–spatial features’ extraction of complex scenes can improve the feature representation capability, which affects the classification performance. Traditional methods only extract the superficial features of HSI, which constrains the classification performance. In contrast, our proposed method fully mines representative features in multi scales of various land covers.

First, 2-D-SSA extracts spatial trend features, which considers the correlation information of local and global spatial contexts. Multiscale 2-D-SSA features provide complementary yet correlated information for classification. In addition, combining SPCA and PCA postprocessing, the spatial features are further merged on the concentrated spectral features. The local spectral–spatial and global spatial features at different scales are obtained to characterize the land-cover differences at different scales. Finally, the main spectral discrimination feature obtained by PCA further improves the feature classification performance.

The fused MSF-PCs feature contains abundant spectral–spatial information and mines the intrinsic correlation of the HSI data. It can be viewed as the representative features in HSI because of its excellent feature representation capability.

VI. CONCLUSION

In this article, a novel multiscale feature fusion method based on PCA and 2-D-SSA is proposed to extract abundant spectral–spatial features from HSI for data classification. SPCA and multiscale 2-D-SSA are integrated to extract local spectral–spatial features, where multiscale 2-D-SSA helps to derive spatial features at different scales. These features are further fused with the global spectral discrimination features obtained by PCA, aiming for improved classification performance.

Compared with existing spectral–spatial methods, results on three real HSI data sets have demonstrated that the proposed method provides better performance when only a small number of training samples are available. It shows

that the extracted abundant features are effective for distinguishing different land-cover classes. Furthermore, compared with state-of-the-art CNN-based deep learning approaches, the proposed method can still achieve higher classification accuracy, which has validated again its superiority in feature extraction.

Although multiscale 2-D-SSA has improved the classification accuracy, a large number of embedded windows have inevitably increased the computational complexity. In future work, superpixel-alike segmentation may be applied to further improve the efficiency of the proposed approach.

REFERENCES

- [1] M. Imani and H. Ghassemian, “An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges,” *Inf. Fusion*, vol. 59, pp. 59–83, Jul. 2020.
- [2] N. Fox, A. Parbhakar-Fox, J. Moltzen, S. Feig, K. Goemann, and J. Huntington, “Applications of hyperspectral mineralogy for geo-environmental characterisation,” *Minerals Eng.*, vol. 107, pp. 63–77, Jun. 2017.
- [3] S. D. Fabyi et al., “Varietal classification of rice seeds using RGB and hyperspectral images,” *IEEE Access*, vol. 8, pp. 22493–22505, 2020.
- [4] A. W. Bitar, L.-F. Cheong, and J.-P. Ovarlez, “Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5239–5251, Aug. 2019.
- [5] Y. Chen, K. Zhu, L. Zhu, X. He, P. Ghamisi, and J. A. Benediktsson, “Automatic design of convolutional neural network for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7048–7066, Sep. 2019.
- [6] 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.
- [7] J. Li, J. M. Bioucas-Dias, and A. Plaza, “Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
- [8] Y. Guo, H. Cao, S. Han, Y. Sun, and Y. Bai, “Spectral–spatial hyperspectral image classification with K-Nearest neighbor and guided filter,” *IEEE Access*, vol. 6, pp. 18582–18591, 2018.
- [9] 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.
- [10] G. Hughes, “On the mean accuracy of statistical pattern recognizers,” *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
- [11] T. Qiao et al., “Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 119–133, Jan. 2017.
- [12] G. Licciardi, P. R. Marpu, J. Chanussot, and J. A. Benediktsson, “Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles,” *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 447–451, May 2012.
- [13] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, “Hyperspectral image classification with independent component discriminant analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.
- [14] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, “Classification of hyperspectral images with regularized linear discriminant analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.
- [15] D. Hong, N. Yokoya, and X. X. Zhu, “Learning a robust local manifold representation for hyperspectral dimensionality reduction,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2960–2975, Jun. 2017.
- [16] J. Zabalza, J. Ren, J. Ren, Z. Liu, and S. Marshall, “Structured covariance principal component analysis for real-time onsite feature extraction and dimensionality reduction in hyperspectral imaging,” *Appl. Opt.*, vol. 53, no. 20, pp. 4440–4449, Jul. 10 2014.
- [17] X. Jia and J. A. Richards, “Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, 1999.

- [18] J. Zabalza *et al.*, "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 112–122, Jul. 2014.
- [19] R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J. C.-W. Chan, and F. Canters, "Improved classification of VHR images of urban areas using directional morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2803–2813, Oct. 2008.
- [20] R. Bao, J. Xia, M. Dalla Mura, P. Du, J. Chanussot, and J. Ren, "Combining morphological attribute profiles via an ensemble method for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 359–363, Mar. 2016.
- [21] Y. Gu, T. Liu, X. Jia, J. A. Benediktsson, and J. Chanussot, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.
- [22] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, Sep. 2015.
- [23] J. Zabalza *et al.*, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418–4433, Aug. 2015.
- [24] J. Zabalza, C. Qing, P. Yuen, G. Sun, H. Zhao, and J. Ren, "Fast implementation of two-dimensional singular spectrum analysis for effective data classification in hyperspectral imaging," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1733–1751, Mar. 2018.
- [25] J. Zabalza, J. Ren, Z. Wang, S. Marshall, and J. Wang, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1886–1890, Nov. 2014.
- [26] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.
- [27] 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, Feb. 2019.
- [28] X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image decomposition for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2241–2253, Apr. 2015.
- [29] S. Li, Q. Hao, X. Kang, and J. A. Benediktsson, "Gaussian pyramid based multiscale feature fusion for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3312–3324, Sep. 2018.
- [30] P. Duan, X. Kang, S. Li, and P. Ghamisi, "Noise-robust hyperspectral image classification via multi-scale total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1948–1962, Jun. 2019.
- [31] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.
- [32] H. Yu *et al.*, "Global spatial and local spectral similarity-based manifold learning group sparse representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3043–3056, May 2020.
- [33] X. Ma, H. Wang, and J. Geng, "Spectral–spatial classification of hyperspectral image based on deep auto-encoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4073–4085, Sep. 2016.
- [34] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
- [35] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.
- [36] J. Yue, W. Zhao, S. Mao, and H. Liu, "Spectral–spatial classification of hyperspectral images using deep convolutional neural networks," *Remote Sens. Lett.*, vol. 6, no. 6, pp. 468–477, Jun. 2015.
- [37] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, p. 67, Jan. 2017.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," 2016, *arXiv:1603.05027*. [Online]. Available: <http://arxiv.org/abs/1603.05027>
- [39] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
- [40] Z. Meng, L. Li, X. Tang, Z. Feng, L. Jiao, and M. Liang, "Multipath residual network for spectral-spatial hyperspectral image classification," *Remote Sens.*, vol. 11, no. 16, p. 1896, Aug. 2019.
- [41] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.
- [42] Mu, Guo, and Liu, "A multi-scale and multi-level spectral-spatial feature fusion network for hyperspectral image classification," *Remote Sens.*, vol. 12, no. 1, p. 125, Jan. 2020.
- [43] D. Danilov and A. Zhigljavskii, "Principal components of time series: The 'Caterpillar' method," (in Russian), St. Petersburg Univ., Saint Petersburg, Russia, Tech. Rep., 1997. [Online]. Available: <http://www.gistatgroup.com/cat/books.html>
- [44] N. Golyandina and K. Usevich, "2D-extension of singular spectrum analysis: Algorithm and elements of theory," in *Matrix Methods: Theory, Algorithms and Applications*, V. Olshevsky and E. Tyrtyshnikov, Eds. 2010, pp. 449–473, doi: 10.1142/9789812836021_0029.
- [45] L. Rodríguez-Aragón and A. Zhigljavsky, "Singular spectrum analysis for image processing," *Stat. Interface*, vol. 3, no. 3, pp. 419–426, 2010.
- [46] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 674–677, Oct. 2007.
- [47] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1519–1531, Mar. 2016.
- [48] H. Yu, L. Gao, W. Liao, B. Zhang, A. Pizurica, and W. Philips, "Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2142–2146, Nov. 2017.
- [49] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 227–239, Jan. 2016.
- [50] Y. Fang *et al.*, "Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1712–1716, Oct. 2014.
- [51] J. Peng, X. Jiang, N. Chen, and H. Fu, "Local adaptive joint sparse representation for hyperspectral image classification," *Neurocomputing*, vol. 334, pp. 239–248, Mar. 2019.
- [52] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
- [53] R. Hansch, A. Ley, and O. Hellwich, "Correct and still wrong: The relationship between sampling strategies and the estimation of the generalization error," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 3672–3675.
- [54] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral–spatial methods in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 862–880, Feb. 2017.
- [55] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019.



Hang Fu received the B.Sc. degree from the China University of Petroleum (East China), Qingdao, China, in 2019, where he is pursuing the M.Sc. degree in geomatics engineering.

His research interests are related to hyperspectral image feature extraction and classification.



Genyun Sun (Member, IEEE) received the B.Sc. degree from Wuhan University, Wuhan, China, in 2003, and the Ph.D. degree from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2008.

He is a Professor with the China University of Petroleum (East China), Qingdao, China. His research interests include remote sensing image processing, hyperspectral remote sensing, high-resolution remote sensing, and intelligent optimization algorithms.



Jinchang Ren (Senior Member, IEEE) received the Ph.D. degree in electronic imaging and media communication from Bradford University, Bradford, U.K., in 2009.

He is a Reader with the Centre for Excellence for Signal and Image Processing and the Deputy Director of the Strathclyde Hyperspectral Imaging Centre, University of Strathclyde, Glasgow, U.K. His research interests include visual computing and multimedia signal processing, especially on image processing, machine learning, big data analytics and hyperspectral imaging.



Aizhu Zhang (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees from the China University of Petroleum (East China), Qingdao, China, in 2011, 2014, and 2017, respectively.

She is a Lecturer with the China University of Petroleum (East China). Her research interests are in artificial intelligence, pattern recognition, city remote sensing, and wetland remote sensing.



Xiuping Jia (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of New South Wales, Sydney, NSW, Australia, in 1996.

She has been with the School of Information Technology and Electrical Engineering, University of New South Wales at Canberra, Canberra, ACT, Australia, since 1988, where she is a Senior Lecturer. She is also a Guest Professor with Harbin Engineering University, Harbin, China, and an Adjunct Researcher with the China National Engineering

Research Center for Information Technology in Agriculture, Beijing, China. She has coauthored the remote sensing textbook entitled *Remote Sensing Digital Image Analysis* (Springer-Verlag, third edition, 1999, and fourth edition, 2006). Her research interests include remote sensing and imaging spectrometry.

Dr. Jia is also an Editor of the *Annals of GIS* and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.