Multiscale 2-D Singular Spectrum Analysis and Principal Component Analysis for Spatial–Spectral Noise-Robust Feature Extraction and Classification of Hyperspectral Images

Ping Ma, Jinchang Ren, *Senior Member, IEEE*, Huimin Zhao, Genyun Sun, *Member, IEEE*, Paul Murray, and Jiangbin Zheng

***Abstract*—In hyperspectral images (HSI), most feature extraction and data classification methods rely on corrected dataset, in which the noisy and water absorption bands are removed. This can result in not only extra working burden but also information loss from removed bands. To tackle these issues, in this article, we propose a novel spatial–spectral feature extraction framework, multiscale 2-D singular spectrum analysis (2-D-SSA) with principal component analysis (PCA) (2-D-MSSP), for noise-robust feature extraction and data classification of HSI. First, multiscale 2-D-SSA is applied to exploit the multiscale spatial features in each spectral band of HSI via extracting the varying trends within defined windows. Taking the extracted trend signals at each scale level as the input features, the PCA is employed to the spectral domain for dimensionality reduction and spatial–spectral feature extraction. The derived spatial–spectral features in each scale are separately classified and then fused at decision-level for efficacy. As our 2-D-MSSP method can extract features and simultaneously remove noise in both spatial and spectral domains, which ensures it to be noise-robust for classification of HSI, even the uncorrected dataset. Experiments on three publicly available datasets have fully validated the efficacy and robustness of the proposed approach, when benchmarked with 10 state-of-the-art classifiers, including six spatial–spectral methods and four deep learning classifiers. In addition, both quantitative and qualitative assessment has validated the efficacy of our approach in noise-robust classification of HSI even with limited training samples, especially in classifying uncorrected data without filtering noisy bands.**

***Index Terms*—Hyperspectral image (HSI), image classification, multiscale 2-D-SSA, principal component analysis (PCA), spatial–spectral feature extraction.**

Manuscript received September 26, 2020; revised November 12, 2020; accepted November 23, 2020. Date of publication November 25, 2020; date of current version January 6, 2021. This work was supported in part by the Natural and Environmental Research Council (NE/S002545/1), U.K., in part by the Dazhi Scholarship of the Guangdong Polytechnic Normal University, in part by the National Natural Science Foundation of China under Grant 62072122, in part by the Education Department of Guangdong Province under Grant 2019KSYS009, China, in part by the National Key Research and Development Program of China under Grant 2019YFE0126700, in part by the University of Strathclyde JARA Scholarship, and in part by the Ph.D. Scholarship of the China Scholarship Council. *(Corresponding authors: Jinchang Ren; Huimin Zhao.)*

Ping Ma and Paul Murray are with the Department of Electronic and Electrical Engineering, University of Strathclyde, G11XW Glasgow, U.K.

Jinchang Ren is with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou 510665, China, and also with the National Subsea Centre, School of Computing, Robert Gordon University, Aberdeen, University of Strathclyde, G11XQ Glasgow, U.K. (e-mail: jinchang.ren@ieee.com).

Huimin Zhao is with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou 510665, China (e-mail: zhaohuimin@gpnu.edu.cn).

Genyun Sun is with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China.

Jiangbin Zheng is with the School of Software, Northwestern Polytechnical University, Xi'an 710072, China.

Digital Object Identifier 10.1109/JSTARS.2020.3040699

I. Introduction

HYPERSPECTRAL images (HSI) contains rich spectral information in the hundreds of narrow spectral bands, which improves the earth observation even with subtle spectral differences [1], [2]. This makes HSI be applied to many applications, such as target detection and land-cover mapping [3]. In recent years, the supervised HSI classification has become a research hotpot for earth observation, in which each pixel is assigned to a certain class label based on the samples labeled in advance for training [4]. For achieving this, many well-known supervised classifier have been developed, such as the popularly used support vector machines (SVM) [5].

However, in HSI, some bands would lack enough light caused by the limitations of atmosphere and sensors, such as the atmospheric water absorption bands and low signal-to-noise ratio bands [6], which inevitably introduces various noise into the image process [7], [8]. In addition, it is difficult to collect sufficient training samples in practice [9], where limited and unbalanced sampling in different classes can easily lead to the lack of training data [7]. These limitations can degrade the accuracy of target detection and image classification.

In order to alleviate these problems, during the last decades, the feature extraction and dimension reduction in HSI have been increasingly investigated. Some widely used spectral feature extraction approaches include principal component analysis (PCA) [10], nonnegative matrix factorization [11], independent component analysis [12] and singular spectrum analysis (SSA) [13]. Here, PCA can realize dimensionality reduction through a linear transformation and thus speed up the subsequent data classification. It is fast and simple to implement without any

label information on the input data [14]. However, these methods perform worse when deal with the HSI scene with high intraclass disparity and interclass similarity [15]. As presented in [16], the HSI data are supposed to be processed as an image rather than a collection of pixels. Thus, various spatial–spectral feature extraction methods are developed to consider both spectral and contextual information [17]. The Gabor filter [18], wavelets [19], the extended morphological profile [15], Markov random field [20], and sparse representation [21] have been applied to the HSI classification. In [3], a 2-D extension to SSA method (denoted as 2-D-SSA) is developed for effective spatial features extraction of HSI, in which each band image is decomposed into various components and reconstructed using trend and selected oscillation information. The 2-D-SSA can extract the spatial structural features by using the characteristics of neighboring pixels in a specified embedding window. It performs better in suppressing high noises and data classification. In 2-D-SSA, the embedding window size is a key parameter that influences the informativeness and noise level of extracted features. However, it is a challenging task to select an optimal window size for all the spectral bands in different HSI datasets.

Recently, many methods use the superpixel segmentation technique to exploit adaptive spatial structures of image [4], [14]. In [5], a novel spatial–spectral model, adjacent superpixel-based multiscale spatial-spectral kernel (ASMGSSK) is proposed. This method employs the entropy rate superpixel segmentation algorithm for multiscale superpixels and then calculate kernel matrix on spatial–spectral features extracted within adjacent superpixels. The ASMGSSK achieves superior classification performance on HSI datasets with very limited training samples. In [14], a superpixelwise PCA approach is proposed by incorporating the spatial context information generated from superpixel technique into the spectral information with dimensionality reduction by PCA method. With the advantage of providing hidden structure spatial–spectral features, the deep learning models have also been successfully applied for classifying HSI [22]–[24], such as the contextual deep convolution neural network (CD-CNN) [25], the diverse region-based CNN (DR-CNN) [26], the spatial–spectral generative adversarial networks (SS-GAN) [27], and the spectral–spatial cascaded recurrent neural network (SSCasRNN) [28]. However, most of the deep learning methods demand more training samples to realize better classification performance and would perform poor when the training size is limited [5].

The aforementioned spectral and/or spatial feature extraction methods show superiority on the classification of corrected HSI data with image preprocessing where noisy and water absorption bands are removed. However, this can lead to the loss of some information in these noisy bands, which may be useful for the classification of some land cover classes in HSI scene. On the other hand, manually prefilter the unwanted bands inevitably increase the image processing burden. Motivated by these, in this article, we focus on the effective feature extraction and classification on uncorrected HSI dataset.

In recent years, many denoising techniques have been proposed to deal with the uncorrected datasets without removing noisy and water absorption bands [8], [29]. For example, Sun *et al.* [30] present a fast superpixel-based subspace low-rank learning method (termed as FS^{2}LRL) for HSI mixed noise removal. This method factorizes HSI data into two lower rank submatrices in the spectral domain and exploit the local spatial low rankness of superpixel regions in the decomposed subspace. In [31], the low-rank representation (LRR) is combined with superpixel segmentation (denoted as SS-LRR), in which LRR is applied to each superpixel to excavate the spatial–spectral information of HSI. These approaches can effectively remove different types of noise simultaneously on the simulated and real HSI datasets, however, achieve limited accuracy improvements on uncorrected HSI classification.

In this article, a novel multi-scale 2-D-SSA with PCA approach (2D-MSSP) is proposed for noise-robust feature extraction and effective data classification of HSI under limited training size. We aim to fully exploit spectral and spatial features directly on the uncorrected dataset without removing the noisy bands. In detail, a multiscale 2-D-SSA model is first applied to original uncorrected HSI dataset to adaptively extract the main information and spatial structures with multiple window sizes. Then, the PCA is utilized to derive informative spectral features and decrease the computational burden for next image processing. The extracted multiscale spatial–spectral features are classified by SVM classifier and then fused on the classification maps with a decision-level fusion technique, which can further improve the robustness. The main contributions of this article can be highlighted as follows:

1) The proposed 2-D-MSSP model can simultaneously derive the informative spatial–spectral features, remove noisy content, and reduce computational burden, which is an effective and efficient noise-robust model for feature extraction and classification of HSI especially on the uncorrected HSI datasets without removing severely noisy and water absorption bands.
2) The different scales used in 2-D-MSSP help to effectively extract the discriminative features under various noise levels, which is beneficial to the classification of different complicated land cover classes in high dimensions.
3) By combining feature-level and decision-level fusion, our 2-D-MSSP method can achieve effective classification in noise-robust HSI with limited training data. Results on three public datasets have validated its superiority than several state-of-the-art spatial–spectral classifiers and even deep learning models.

The remainder of this article is organized as follows. In Section II, our proposed 2-D-MSSP is presented in detail. Section III describes the experimental setup, including the testing HSI datasets and parameter settings. Experimental results and analysis are discussed in Section IV, with some conclusions are drawn in Section V.

II. Proposed Method

Fig. 1 shows the flowchart of proposed 2-D-MSSP approach, which has three main stages, i.e., multiscale 2-D-SSA for

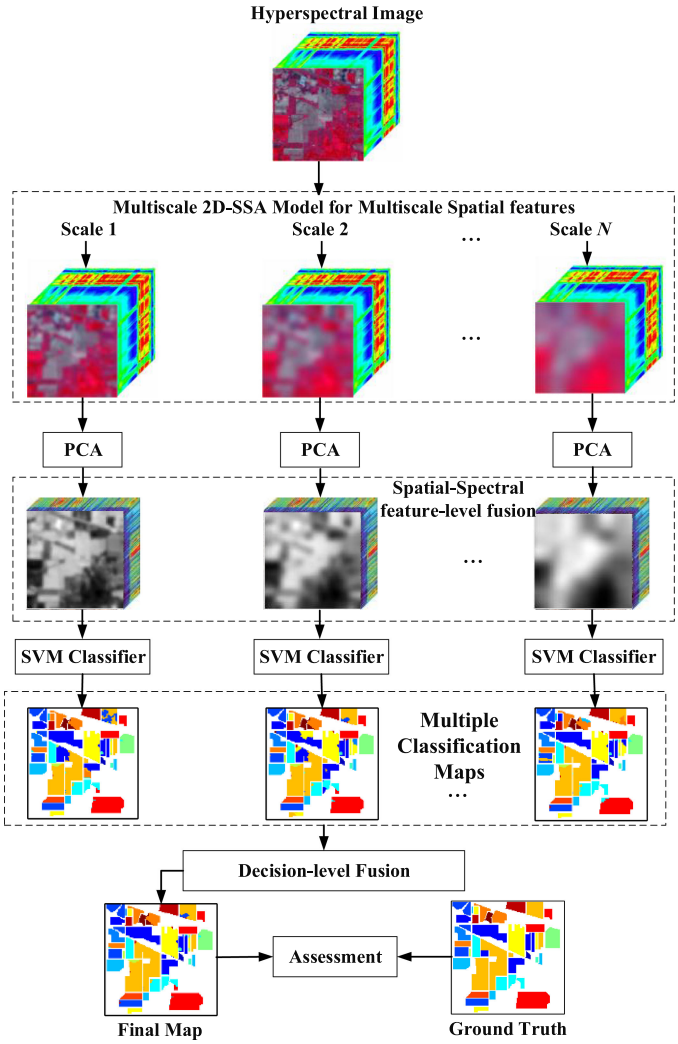

Fig. 1. Flowchart of the proposed SMP method.

spatial feature extraction, PCA for spectral features extraction, HSI classification, and fusion. More detailed implementation of 2-D-MSSP is presented as follows.

A. Spatial–Spectral Features Extraction by 2-D-MSSP

The SSA is an effective tool for the analysis and forecasting of time-series data [32]. It can decompose a series into several components, i.e., trend, oscillations, and noise. When applying SSA to each spectral band of a hypercube, the 2-D scene can be decomposed and then reconstructed using respective main components with the noise information removal. Specifically, for an HSI data $\boldsymbol{I}$ with a size $N_x \times N_y \times D$, where N_x and N_y represent the band image size and D is the dimension size. Then, one band image $I^d (d \in D)$ can be represented by the following matrix:

$$I^d = \begin{pmatrix} I^d_{1,1} & I^d_{1,2} & \cdots & I^d_{1,N_y} \\ I^d_{2,1} & I^d_{2,2} & \cdots & I^d_{2,N_y} \\ \vdots & \vdots & \ddots & \vdots \\ I^d_{N_x,1} & I^d_{N_x,2} & \cdots & I^d_{N_x,N_y} \end{pmatrix}. \tag{1}$$

A 2-D window W with a size $L_x \times L_y$ is then defined as

$$W^d = \begin{pmatrix} I^d_{i,j} & I^d_{i,j+1} & \cdots & I^d_{i,j+L_y-1} \\ I^d_{i+1,j} & I^d_{i+1,j+1} & \cdots & I^d_{i+1,j+L_y-1} \\ \vdots & \vdots & \ddots & \vdots \\ I^d_{i+L_x-1,j} & I^d_{i+L_x-1,j+1} & \cdots & I^d_{i+L_x-1,j+L_y-1} \end{pmatrix} \tag{2}$$

where $L_x \le N_x$, $L_y \le N_y$; (i, j) is the spatial position of one pixel in image I^d. Given a reference position (i, j), the pixels in the corresponding window W can be rearranged into a column vector $A^d_{i,j} \in \Re^{L_x L_y}$ given as follows:

$$A^d_{i,j} = \left(I^d_{i,j},\ I^d_{i,j+1}, \ldots,\ I^d_{i,j+L_y-1}, I^d_{i+1,j}, \ldots, I^d_{i+L_x-1,j+L_y-1}\right)^T. \tag{3}$$

This 2-D window would raw scan image I^d from the top left to bottom right in order to exploit all the possible position over the whole image. Thus, the trajectory matrix X^d of all these possible 2-D window in image I^d with size $L_x L_y \times (N_x - L_x + 1)(N_y - L_y + 1)$ can be derived by

$$X^d = \left(\left(A^d_{1,1}\right)^T, \left(A^d_{1,2}\right)^T, \ldots,\ \left(A^d_{1,N_y-L_y+1}\right)^T \right. \left. \left(A^d_{2,1}\right)^T, \ldots, \left(A^d_{N_x-L_x+1,N_x-L_x+1}\right)^T\right). \tag{4}$$

Note that the trajectory matrix X^d has a structure of Hankel-block-Hankel (HbH). Here, X^d can be represented as

$$X^d = \begin{pmatrix} H^d_1 & H^d_2 & \cdots & H^d_{N_x-L_x+1} \\ H^d_2 & H^d_3 & \cdots & H^d_{N_x-L_x+2} \\ \vdots & \vdots & \ddots & \vdots \\ H^d_{L_x} & H^d_{L_x+1} & \cdots & H^d_{N_x} \end{pmatrix}_{L_y \times (N_y - L_y + 1)} \tag{5}$$

where each submatrix H^d_k is Hankel structure as

$$H^d_k = \begin{pmatrix} I^d_{k,1} & I^d_{k,2} & \cdots & I^d_{k,N_y-L_y+1} \\ I^d_{k,2} & I^d_{k,3} & \cdots & I^d_{k,N_y-L_y+2} \\ \vdots & \vdots & \ddots & \vdots \\ I^d_{k,L_y} & I^d_{k,L_y+1} & \cdots & I^d_{k,N_y} \end{pmatrix}_{L_y \times (N_y - L_y + 1)}. \tag{6}$$

With the obtained trajectory matrix X^d, the singular value decomposition (SVD) is implemented to derive the eigenvalues $(\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{L_x L_y})$ and the corresponding eigenvectors $(U_1, U_2, \ldots, U_{L_x L_y})$ of $X^d (X^d)^T$, and X^d can be rewritten as

$$X^d = X^d_1 + X^d_2 + \ldots + X^d_{L_x L_y} \tag{7}$$

where the rth elementary matrix $X_r = \sqrt{\lambda_r}\, U_r V_r^T$, its principal components (PCs) $V_r = (X^d)^T\, U_r / \sqrt{\lambda_r}$.

Afterwards, the eigenvalue grouping is carried out, in which the total set of $L_x L_y$ individual components in (7) is grouped into m subsets, denoted as $P = [P_1, P_2, \ldots, P_m]$. We can select one or more elementary matrices X_r in each subset to derive the main information of the image without high noisy content. Thus,

trajectory matrix X is transformed as

$$X^d = X^d_{P_1} + X^d_{P_2} + \ldots + X^d_{P_m}. \tag{8}$$

Here, X is not necessarily an HbH type matrix. In order to project it into a 2-D signal, a two-step Hankelization process is applied first within each block (6) and then between blocks (5), by averaging the antidiagonal values in the matrix [3], [32]. As a result, we obtain the reconstructed image $\boldsymbol{I}_{\mathbf{spa}}$, which contain the informative spatial features based on local relevant information in the defined 2-D window.

After extracting the spatial structural content of each band in HSI, PCA is utilized for the main spectral features exploitation and further noise removal as follows:

$$\boldsymbol{I}_{\boldsymbol{spa-spe}} = \mathrm{PCA}(\boldsymbol{I}_{\boldsymbol{spa}})_{N_x \times N_y \times L} \tag{9}$$

where L is the number of PCs derived by PCA technique. In this way, the informative content and noisy information even in noisy and water absorption bands can be processed simultaneously in spatial and spectral domain.

In this article, multiple window sizes $(N_x \times N_y)$ in 2-D-MSSP are utilized. At a given window size, the obtained spatial–spectral features from 2-D-MSSP are taken as the input of SVM classifier to generate the classification map. Then, we can get multiple classification maps by using different window sizes. Here, a decision fusion strategy is utilized to generate the final classification result. Specifically, set T scales on the window size, then there would be T classification maps generated, denoted as $\boldsymbol{S} = \{S_1, S_2, \ldots, S_T\}$. For a given pixel with the spatial position (i, j), its final class label is calculated through majority voting, i.e., the label with the maximum number of appearances within multiple classification maps $\boldsymbol{S}$

$$S_{i,j} = \underset{c=1,\ldots,C}{\arg\max} \sum_{t=1}^{T} F(S^t_{i,j}, c) \tag{10}$$

where c represents the possible class labels in one HSI data, C is the number of class labels, and F is an indicator function given by

$$F(p,q) = \begin{cases} 1 \text{ if } p = q \\ 0 \text{ otherwise.} \end{cases}$$

B. Noise Robustness Analysis on 2-D-MSSP

In HSI, each pixel through hundreds of bands can be characterized as a sequential spectral vector, which contains high noisy information especially in the water absorption bands [30]. By applying 2-D-MSSP, one pixel vector can be decomposed and reconstructed based on the eigenvalues obtained in SVD. In general, the component from the first eigenvalue contains the main information of input data, while noise is usually existed in those small eigenvalues [3]. Therefore, the newly reconstructed HSI data only using more significant components and excluding smaller eigencomponents can be clearer and more discriminative for image analysis and interpretation.

In 2-D-MSSP, the 2-D embedding window size is a key parameter which affects the informativeness and noise level of reconstructed HSI. This parameter directly decides the total number of eigenvalues (components) available. A large window size leads to more components generated. For example, when the window size is $L_x = L_y = 5$, the 2-D embedding window would provide 25 components for each spectral scene. The grouping of these components denoted as a reconstruction. Thus, when only used the basic trend information (the first eigenvalue component) for image reconstruction, the reconstructed data with small window size will contain more information, while the large one could produce more smoothed results with most of noisy content removed [33].

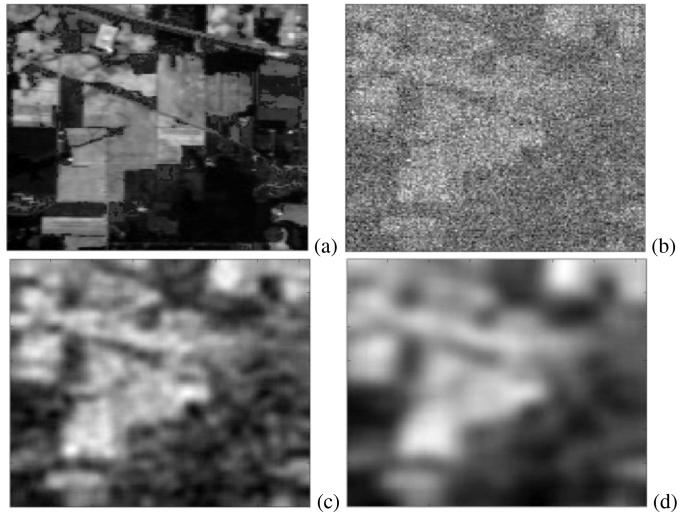

Fig. 2. Application of 2-D-MSSP to a scene in HSI. (a) Original scene at 667 nm. (b) Original scene at 2499 nm (a water absorption band); and reconstructed scenes from (b) with $L_x = L_y = 5$ (c) and $L_x = L_y = 10$ (d).

An example taken the 220th spectral band (at 2499 nm), which is a water absorption band, from the uncorrected Indian Pines dataset [3], Fig. 2 shows the reconstructed scene using first component with different sized 2-D embedding window, along with a clear spectral band (at 667 nm) for comparison. As seen by comparing Fig. 2(a) and (b), it is clear that the water absorption band contains much highly noisy content with less detailed information, and that is why many methods remove noisy and water absorption bands before processing HSI dataset. However, these bands may still contain some useful information. As shown in Fig. 2(c) and (d), when applying 2-D-SSA to the 220th spectral band, detailed local spatial structures are preserved, and noisy information is reduced. Besides, the 2-D-SSA with a larger 2-D embedding window size generates more smoothed scene with noise and details much reduced. In essence, for HSI data with a high noise level, larger window sizes can remove most of the noisy content. However, if the window size is too large, some important details of the spectral scene may be lost. On the other hand, small window sizes may result in the noise kept which degrades the further image classification. Overall, it shows that our proposed 2-D-MSSP can derive the main information and spatial structures of HSI even on the water absorption bands, which demonstrates its robustness to noisy content.

In real cases, the noise level varies throughout the hyperspectral bands [8], [34], which also verified in Fig. 2(a) and (b). Thus,

it is challenging to set an optimal window size for all the bands. The proposed multiscale 2-D-SSA in this article can better solve this problem and further improve the classification performance, which will be further validated in Section III.

III. Experimental Setup

In this article, we evaluated the performance of the proposed method on three publicly available HSI datasets. The details of these three datasets, experimental settings, and related ablation study are described in this section.

A. Real Datasets Description

1) *Indian Pines:* The dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor [35] over the Indian Pines study site in Northwest Indiana in 1992. The whole scene has 16 types of land-cover classes and 145 × 145 pixels with a low spatial resolution of 20 m/pixel. Each pixel contains 220 spectral bands covering the wavelengths from $0.2-2.4\ \mu\mathrm{m}$. The spectral bands used in most image analysis methods are usually reduced to 200 by discarding 20 water absorption bands (104–108, 150–163, and 220) [36].
2) *SalinasA:* We used a subscene from SalinasA image, denoted SalinasA. It was also collected via AVIRIS in the Salinas Valley in California, USA. This subscene was composed of 86 × 83 pixels and six land-cover classes with a geometric resolution of 3.7 m. Similarly, this scene can be corrected by removing 20 water absorption bands (108–112, 154–167, and 224) from 224 spectral bands [37].
3) *Pavia University:* The dataset is named as PaviaU in this article. It was acquired by the Reflective Optics System Imaging Spectrometer sensor over Pavia, Northern Italy [38], in 2001. This scene consists of 610 × 340 pixels and eight dominant classes of land covers with a spatial resolution of 1.3 m. Each pixel includes 114 spectral bands in a spectral range of $0.43-0.86\ \mu\mathrm{m}$. For avoiding the effect of water absorption, this scene can be corrected by reducing the available number of bands to 103. Further details including the sampling size of each class in these three datasets can be found in [37], [39].

B. Experimental Settings

To investigate the robustness and classification performance of the proposed method, several state-of-the-art spatial–spectral classification approaches are employed to test on the uncorrected HSI dataset (with no bands removed), and the corrected dataset (with removing the noisy and water-absorption bands). These benchmarks include 2-D-SSA [3], MSuperPCA [14], ASMGSSK [5], along with two advanced denoising methods, SS-LRR [31], and FS^{2}LRL [30], which show efficacy in classifying uncorrected HSI data. In addition, we directly apply the SVM on raw spectral profiles of HSI as the baseline for classification with spectral-only information (denoted as SVM-spe).

TABLE I
List of Parameter Settings for Our Proposed SMP Framework and Other Benchmarking Approaches

Method	Indian Pines, SalinasA	PaviaU
SVM-spe	N/A	N/A
2D-SSA [3]	Window size: 10×10 EV Grouping (EVG): 1st	Window size: 5×5 EV Grouping (EVG): 1-2nd
MSuperPCA [14]	fundamental superpixel number:100 Scale number: 4	fundamental superpixel number:20 Scale number: 6
ASMGSSK [5]	Filtering degree:0.1 Superpixel number: {100, 200, 400, 800, 1600, 3200}	Filtering degree:0.1 Superpixel number: {200, 400, 800, 1600, 3200, 6400}
SS-LRR [31]	superpixel number: 40	superpixel number: 190
FS^{2}LRL [30]	Subspace dimension: 10 superpixel number: 100 sparse parameter: 0.13 fidelity parameter: 0.040	Subspace dimension: 10 superpixel number: 100 sparse parameter: 0.13 fidelity parameter: 0.040
2D-MSSP	Window size: {5 × 5, 10 × 10, 20 × 20,40 × 40,60 × 60} NPCs: 40	Window size: {5 × 5, 10 × 10, 20 × 20,40 × 40,60 × 60} NPCs: 40

The key parameters of these approaches are set according to their recommended default values [13] as illustrated in Table I. As for our proposed 2-D-MSSP, the sizes of 2-D embedding window are set to $\{5\times5,\ 10\times10,\ 20\times20,\ 40\times40,\ 60\times60\}$ with five scales. In [3], the 2-D-SSA was proposed for feature extraction and data classification in HSI, where different window sizes including $5\times5,\ 10\times10,\ 20\times20,\ 40\times40$, and 60×60 were tested. For Indian Pines and Pavia University datasets, the optimal window size is found to be 10×10 and 5×5 to produce the highest classification accuracy. As for the Salinas dataset, the window sizes of 40×40 and 60×60 achieves the best classification. As the optimal window size varies in different datasets, the multiscale strategy is adopted in the proposed method, aiming to improve the generalization ability on different datasets. The size number of PCs selected in the PCA is set to 40 according to the ablation study in Section III-C. In this article, the SVM is adopted as the default classifier for all the involved methods, which are implemented using LIBSVM toolbox [40]. Here, the base kernel function is Gaussian radial basis function with the kernel factor and penalty parameter set to 0.125 and 1024 based on the grid search, respectively. Such parameter settings of SVM are kept same in all comparing experiments to ensure a fair comparison.

In order to evaluate the efficacy of the proposed method under various limited and unbalanced training sizes, we set the training size varies within $\{1\%, 3\%, 5\%, 10\%\}$ samples per class in the training set using a stratified random scheme from ground truth data. The remaining samples are left for test set. Note that there are some minority classes named *alfalfa*, *grass/pasture-mowed,* and *oats* in Indian Pines dataset [41]. The experiments in all datasets are independently repeated 10 times, where the averaged results in terms of the overall accuracy (OA), Kappa coefficients (κ) and average accuracy (AA) are employed as quantitative evaluation metrics.

All experiments in this article were implemented using the MATLAB 2018a platform on the computer with an Intel (R) Core (TM) i7-8700 CPU (3.20 GHz) and 16.0 GB of memory.

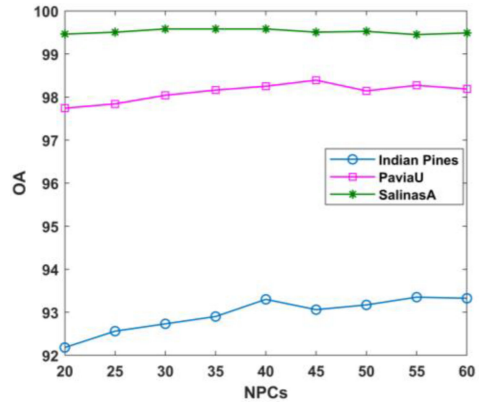

Fig. 3. OA of 2-D-MSSP with various numbers of PCs for test datasets.

C. Parameter Analysis

The number of PCs (NPCs) in PCA directly affects the informativeness in the spatial–spectral features, which further influence the final classification results. In this subsection, we discuss the effect of NPCs on the performance of our proposed framework using three corrected datasets. Here, the value of NPCs ranges from 20 to 60 with 5 as the interval. The numbers of training samples were set to 1% per class and the remaining for testing. Meanwhile, all tests are conducted for 10 times and the corresponding averaged results are recorded to reduce possible biases induced by random sampling.

Fig. 3 shows the OA results of the proposed 2-D-MSSP by using different values of NPCs on the three datasets. As seen, when using 1% training samples per class, the best NPCs settings on Indian Pines is 40 and 55, while the optimal values on SalinasA fall in the range (30, 40). As for the PaviaU dataset, 2-D-MSSP performs best when the NPCs equals to 45. Herein, it is clear that the Indian Pines dataset is more sensitive to the NPCs than other two datasets. This may be due to the fact that the spatial resolution of Indian Pines is much lower and thus its corresponding noises are relatively higher. Therefore, the effect of PCA in noise reduction is more significant on the Indian Pines dataset. Overall, we set NPCs to 40 in all experiments as a tradeoff between the efficiency and accuracy in this article.

IV. Experimental Results and Analysis

In this section, to investigate the robustness of our proposed approach, the experiments are conducted on, respectively, the uncorrected HSI dataset (without removing any bands), and the corrected dataset (with the noisy and water-absorption bands removed). Besides, the effectiveness of different strategies in 2-D-MSSP is also assessed, i.e., the multiscale 2-D-SSA, and the PCA technique. To achieve this, we consider two cases: 1) 2-D-MSSP with only the multiscale 2-D-SSA but not PCA, denoted as 2-D-MSSA, and 2) the proposed 2-D-MSSP. Here, the baseline SVM-spe is denoted as the method without the above two strategies. Furthermore, our method is also compared with several advanced deep learning approaches on three datasets.

A. Results From the Indian Pines Dataset

The quantitative and qualitative performances are evaluated on the corrected and uncorrected Indian Pines datasets. First, the quantitative evaluation in terms of the OA and the κ is conducted between 2-D-MSSP and SVM-spe, 2-D-SSA [3], MSuperPCA [14], ASMGSSK [5], SS-LRR [31], FS^{2}LRL [30], and 2-D-MSSA. Then, three best performing methods are selected to assess the detailed visual and class-based performance comparing with the proposed 2-D-MSSP. Note that the best results in tables are labeled in bold face for comparison. As for the visual assessment, the correct or incorrect classification results of different methods are marked in black and magenta circles, respectively.

1) Quantitative Evaluation: From Table II, it is clear that the larger training size leads to higher classification accuracies for all involved methods. Even on these limited and unbalanced training samples, our proposed 2-D-MSSP achieve the best performance in terms of OA and κ in all cases especially when the training size is small. For instance, 2-D-MSSP yields about 6.5% better than the second-best method, MSuperPCA, in OA when using only 1% training samples per class. This validates the superiority of our method in dealing with the Hughes effect. Besides, by comparing SVM_spe with other methods, it is obvious that the spatial feature introduction (2-D-SSA, MSuperPCA, ASMGSSK, SS-LRR, FS^{2}LRL, 2-D-MSSA, and 2-D-MSSP) can improve the classification performance by only using spectral information. Herein, the SS-LRR and FS^{2}LRL are the denoising methods, which also exhibit promising results on HSI classification. This demonstrates that the noisy content would degrade the image quality and affect the performance of image processing.

Through analyzing the results obtained by SVM-spe, 2-D-SSA, 2-D-MSSA, and 2-D-MSSP, we can find that the multiscale 2-D-SSA and PCA has significantly enhanced the classification results on the corrected and uncorrected Indian Pines dataset. Specifically, the 2-D-MSSP using multiple window sizes has outperformed 2-D-SSA which only uses one fixed window size, with a gain up to 15% on uncorrected data and 18% on the corrected one. This is mainly because that the different window sizes can exploit multiscale local spatial features, which are beneficial to model the different sized land cover classes. In addition, by comparing 2-D-MSSA and 2-D-MSSP, we find that the PCA enables the accuracy improvement up to 3.5%. This may owe to the combination of spatial and spectral features.

As seen in Table II, the involved comparing methods perform differently on the corrected and uncorrected Indian Pines datasets. Concretely, SVM-spe, FS^{2}LRL, MSuperPCA, and ASMGSSK perform better on corrected dataset. On the contrary, the 2-D-SSA can mitigate the image degradation from noisy and water absorption bands and produce superior performance on uncorrected Indian Pines dataset in most cases. With regard to our proposed 2-D-MSSP, by only using multiscale 2-D-SSA, 2-D-MSSA is able to achieve better results on the uncorrected datasets, which avoids the noise effects in all cases. Furthermore, the 2-D-MSSP obtains superior accuracy on uncorrected dataset to those from the corrected one. This validates the efficacy of

TABLE II
OA (%) AND κ (IN PARENTHESES) OF DIFFERENT METHODS ON CORRECTED AND UNCORRECTED INDIAN PINES DATASETS UNDER VARIOUS TRAINING SIZES

Samples	Dataset	SVM-spe	2D-SSA	SS-LRR	FS^{2}LRL	MSuperPCA	ASMGSSK	2D-MSSA	2D-MSSP
1%	Uncorrected	55.10 (0.48)	78.96 (0.76)	76.83 (0.74)	70.18 (0.66)	87.83 (0.86)	87.76 (0.86)	90.88 (0.90)	**94.32 (0.94)**
	Corrected	59.14 (0.53)	75.07 (0.72)	76.17 (0.73)	70.72 (0.67)	89.12 (0.87)	87.98 (0.86)	89.54 (0.88)	**93.08 (0.92)**
3%	Uncorrected	64.68 (0.60)	91.44 (0.90)	89.04 (0.88)	83.02 (0.81)	95.05 (0.94)	94.61 (0.94)	96.55 (0.96)	**98.42 (0.98)**
	Corrected	70.59 (0.66)	89.08 (0.88)	89.25 (0.88)	83.53 (0.81)	95.24 (0.95)	94.63 (0.94)	96.16 (0.96)	**98.23 (0.98)**
5%	Uncorrected	69.46 (0.65)	94.76 (0.94)	91.71 (0.91)	87.76 (0.86)	96.42 (0.96)	96.02 (0.95)	98.03 (0.98)	**99.35 (0.99)**
	Corrected	74.90 (0.71)	93.55 (0.93)	92.00 (0.91)	88.16 (0.87)	96.56 (0.96)	96.18 (0.96)	97.57 (0.97)	**99.04 (0.99)**
10%	Uncorrected	76.13 (0.73)	97.78 (0.97)	94.25 (0.93)	91.80 (0.91)	97.45 (0.97)	97.27 (0.97)	99.29 (0.99)	**99.77 (1.00)**
	Corrected	80.98 (0.78)	97.15 (0.97)	94.68 (0.94)	91.96 (0.91)	97.58 (0.97)	97.56 (0.97)	99.06 (0.99)	**99.67 (1.00)**

TABLE III
CLASSIFICATION ACCURACY (%) OF COMPETITIVE METHODS FOR THE INDIAN PINES DATA WITH 1% TRAINING SAMPLES PER CLASS

Class	Uncorrected				Corrected			
	2D-SSA	MSuper PCA	ASM GSSK	2D-MSSP	2D-SSA	MSuper PCA	ASM GSSK	2D-MSSP
1	52.59	**100.00**	83.04	**100.00**	5.41	**100.00**	60.96	**100.00**
2	73.71	84.46	83.57	**90.23**	67.74	85.54	84.74	**95.85**
3	74.59	79.23	83.94	**92.75**	69.37	78.60	83.35	**92.83**
4	78.33	52.88	85.46	**99.15**	71.61	60.57	84.50	**97.91**
5	69.74	80.43	81.63	**82.49**	68.79	**88.31**	82.43	81.78
6	86.36	93.99	87.49	**98.63**	84.37	97.63	88.99	**99.18**
7	92.72	93.09	92.96	**96.30**	91.85	93.09	95.19	**100.00**
8	93.28	97.05	96.17	**100.00**	91.26	96.75	98.33	**98.73**
9	64.04	**100.00**	70.35	**100.00**	55.79	**100.00**	76.14	84.21
10	68.42	84.76	82.80	**91.90**	65.26	86.51	81.05	**91.11**
11	80.79	94.63	95.07	**96.54**	77.73	**96.11**	94.60	94.41
12	61.83	75.32	77.79	**87.34**	53.05	69.64	76.80	**89.91**
13	97.81	**99.50**	86.68	94.11	91.35	**99.52**	91.82	90.79
14	91.99	95.42	94.39	**99.38**	90.71	**96.12**	94.45	95.24
15	76.25	86.29	82.16	**93.27**	68.32	83.22	83.97	**88.04**
16	92.50	47.93	45.11	**97.93**	89.28	73.88	57.86	**96.74**
OA	78.96	87.83	87.76	**94.30**	75.07	89.12	87.98	**93.08**
κ	0.76	0.86	0.86	**0.94**	0.72	0.88	0.86	**0.92**
AA	78.43	85.31	83.04	**95.00**	71.37	87.84	83.45	**93.64**

2-D-MSSP in noise-robust feature extraction. The operations from multiscale 2-D-SSA and PCA can extract discriminative features and reduce noises on both spatial and spectral domains. As a result, 2-D-MSSP can be used as an efficient and effective tool for feature extraction and HIS classification without removing unwanted bands.

2) Visual and Class-Based Performance Assessment: In order to analyze the classification performance of each class on corrected and uncorrected datasets, our method is benchmarked with three best performing methods when using 1% training samples per class as shown in Table II, i.e., 2-D-SSA, MSuperPCA, and ASMGSSK.

From Table III, it is clear that the 2-D-MSSP has obtained the best classification accuracy no matter if the noisy and water absorption bands removed or not, where 2-D-MSSP produced best results in 15 out of 16 classes on uncorrected dataset while it achieved 11 out of 16 classes on corrected Indian Pines dataset. The performance of 2-D-MSSP is about 6.5% better than second-best on uncorrected Indian Pines. Especially, 2-D-MSSP significantly suppress its peers especially in the classes 2, 3, 4, 10, 12, 15, and 16 (*Corn-no till*, *Corn-min till, Corn*, *Soybeans-no till*, *Soybeans-clean till, Bldg-Grass-Tree-Drives*, *Stone-steel towers*). In addition, 2-D-SSA and our proposed 2-D-MSSP perform better on uncorrected dataset, while MSuperPCA and ASMGSSK produce better results on the corrected one. This validates the noise-robustness of 2-D-SSA and 2-D-MSSP when training size is 1%. In Indian Pines dataset, there are some minor classes *alfalfa*, *grass/pasture-mowed,* and *oats*, which have the small number of samples for training. When using 1% training size, there are only 5, 3, and 2 samples for training in classes *alfalfa*, *grass/pasture-mowed,* and *oats*, respectively. From Table III, we can find that 2-D-MSSP and MSuperPCA achieve promising performance in classifying *grass/pasture-mowed* and even realize correct identification of all corresponding pixels of *alfalfa* and *oats.* Overall, 2-D-MSSP can effectively identify the large homogeneous regions and some small objects simultaneously.

Fig. 4 shows the visual result comparisons between four approaches on uncorrected and corrected datasets using 1% training samples per class, respectively. As shown, these four methods with spatial features have generated more accurate and smoother results. However, there still exist many misclassifications on maps. Specifically, as marked in magenta circles in Fig. 4, the classes *Corn-no till* and *Soybeans-min till, Soybeans-no till* and *Corn-no till* are severely misclassified with each other. As for 2-D-SSA, it better identifies these classes on uncorrected datasets as seen in Fig. 4(c). Even so, on both uncorrected and corrected data, 2-D-SSA yields wrong classifications at the boundaries or small objects. This mainly because that there is only one type of spatial features are extracted in 2-D-SSA without considering boundary information. On the contrary, MSuperPCA and ASMGSSK give much better uniformity and boundary location results as illustrated in Fig. 4(e)–(h). This may be due to that they employ the superpixel segmentation techniques which segment the land cover classes into different objects with adaptive shapes and sizes. Though the geometrical distortions in Fig. 4(c) and (d) are much corrected by MSuperPCA and ASMGSSK, there are still several small uncorrected classifications Fig. 4(e)–(h). To be specific, MSuperPCA produces better visual performance on corrected Indian Pines dataset where misclassified regions of *Soybeans-no till* and *Corn-no till* are reduced. However, the classes *Corn-no till* and *Soybeans-min till* are still wrongly discriminated in some areas as highlighted in magenta circles in Fig. 4(e) and (f). Besides, MSuperPCA gains promising performance in identifying Woods on uncorrected data while it better classifies Grass/pasture on the corrected one as marked in black circle in Fig. 4(e) and (f). The ASMGSSK obtain comparably similar results with MSuperPCA. It performs better on corrected data, where classes *Grass/trees* and *Corn-no till* are seriously misclassified on the uncorrected data. With

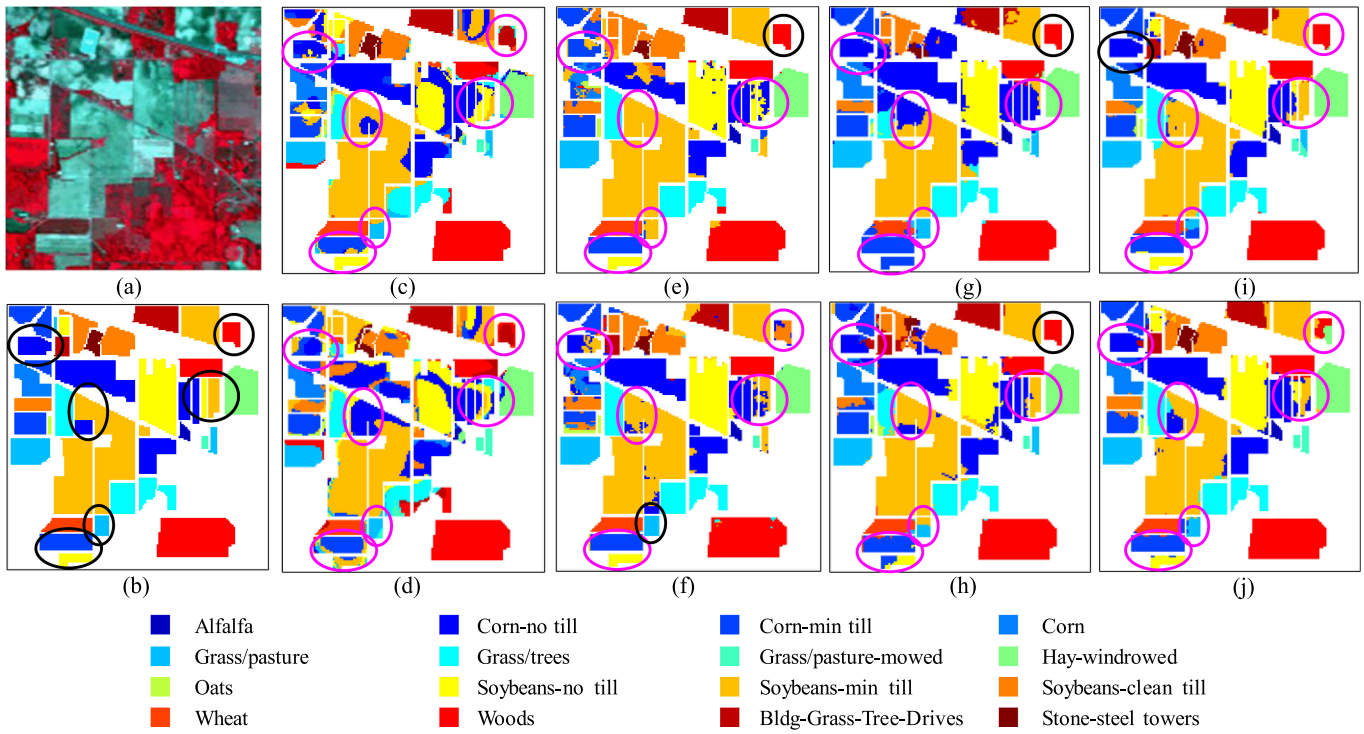

Fig. 4. Classification maps on the Indian Pines data (1% samples per class). (a) False color image (R: 831 nm, G: 657 nm, B: 557 nm). (b) Ground truth in 16 classes. (c) 2-D-SSA on uncorrected data (OA = 78.96%). (d) 2-D-SSA on corrected data (OA = 75.07%). (e) MSuperPCA on uncorrected data (OA = 87.83%). (f) MSuperPCA on corrected data (OA = 89.12%). (g) ASMGSSK on uncorrected data (OA = 87.76%). (h) ASMGSSK on corrected data (OA = 87.98%). (i) 2-D-MSSP on uncorrected data (OA = 94.30%). (j) 2-D-MSSP on corrected data (OA = 93.08%).

TABLE IV
OA (%) AND κ (IN PARENTHESES) OF DIFFERENT METHODS ON THE CORRECTED AND UNCORRECTED SALINAS DATASETS UNDER VARIOUS TRAINING SIZES

Samples	Dataset	SVM-spe	2D-SSA	SS-LRR	FS^{2}LRL	MSuperPCA	ASMGSSK	2D-MSSA	2D-MSSP
1%	Uncorrected	97.49 (0.97)	99.04 (0.99)	97.67 (0.97)	98.38 (0.98)	88.95 (0.86)	97.29 (0.97)	99.53 (0.99)	**99.72 (1.00)**
	Corrected	97.12 (0.96)	98.86 (0.99)	97.79 (0.97)	97.88 (0.97)	91.81 (0.90)	97.37 (0.97)	99.57 (0.99)	**99.59 (0.99)**
3%	Uncorrected	98.17 (0.98)	99.57 (0.99)	98.67 (0.93)	99.04 (0.99)	97.15 (0.96)	98.56 (0.98)	99.85 (1.00)	**99.96 (1.00)**
	Corrected	98.34 (0.98)	99.60 (0.99)	98.73 (0.98)	98.91 (0.99)	97.46 (0.97)	98.57 (0.98)	99.91 (1.00)	**99.94 (1.00)**
5%	Uncorrected	98.72 (0.98)	99.67 (1.00)	99.04 (0.99)	99.41 (0.99)	98.40 (0.98)	99.10 (0.99)	99.98 (1.00)	**100.00 (1.00)**
	Corrected	98.80 (0.99)	99.71 (1.00)	99.04 (0.99)	99.34 (0.99)	98.54 (0.98)	99.20 (0.99)	99.94 (1.00)	**100.00 (1.00)**
10%	Uncorrected	99.26 (0.99)	99.83 (1.00)	99.42 (0.99)	99.66 (1.00)	99.27 (0.99)	99.56 (0.99)	100.00 (1.00)	**100.00 (1.00)**
	Corrected	99.34 (0.99)	99.78 (1.00)	99.50 (0.99)	99.69 (1.00)	99.37 (0.99)	99.59 (0.99)	100.00 (1.00)	**100.00 (1.00)**

regard to our proposed 2-D-MSSP, as shown in Fig. 4(i) and (j), it achieves better visual performance than others with improved object boundaries and spatial consistency. As highlighted in the circles, the results on uncorrected dataset are superior to those on corrected one. The classes *Corn-no till*, *Soybeans-min till, Grass/pasture, Soybeans-no till* are better identified by 2-D-MSSP by exploiting more information from uncorrected data. These visual results are consistent with the class-based results in Table III, which again verify the efficacy of our method in noise-robust feature extraction and HSI classification.

B. Results From the Salinas A Dataset

Similarly, the quantitative and qualitative performance are also assessed on SalinasA dataset with or without noisy and water absorption bands, denoted as uncorrected SalinasA and corrected SalinasA, respectively. The comparisons on these two types of data between SVM-spe, 2-D-SSA, SS-LRR, FS^{2}LRL, MSuperPCA, ASMGSSK, 2-D-MSSA, and 2-D-MSSP in terms of OA and κ are summarized in Table IV. Furthermore, the visual- and class-based results of top four methods also described in detail.

1) Quantitative Evaluation: As shown in Table IV, all approaches achieve much higher classification results than those on Indian Pines datasets. This mainly due to the high spatial resolution and simple structured land covers in SalinasA image, which leads to low noise level and less spectral and spatial confusion in the scene. As a result, land cover classes are more easily to identify. Overall, our proposed method 2-D-MSSP still ranks first on both uncorrected and corrected datasets with the highest OA and κ in six cases. This again demonstrates the efficacy of 2-D-MSSP under various limited training samples. Besides, 2-D-MSSP generates better results on uncorrected dataset when using 1% and 3% training samples per class, and correctly

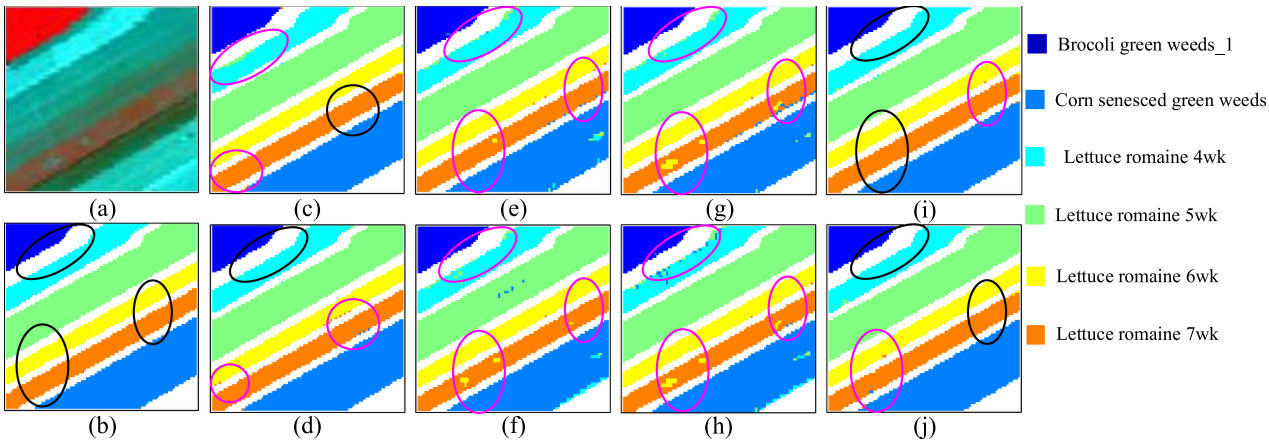

Fig. 5. Classification maps on the SalinasA dataset (1% samples per class). (a) False color image (R: 831 nm, G: 657 nm, B: 557 nm). (b) Ground truth in 16 classes. (c) 2-D-SSA on uncorrected data (OA = 99.04%). (d) 2-D-SSA on corrected data (OA = 98.86%). (e) SS-LRR on uncorrected data (OA = 97.67%). (f) SS-LRR on corrected data (OA = 97.79%). (g) FS^{2}LRL on uncorrected data (OA = 98.38%). (h) FS^{2}LRL on corrected data (OA = 97.88%). (i) 2-D-MSSP on uncorrected data (OA = 99.72%). (j) 2-D-MSSP on corrected data (OA = 99.59%).

TABLE V
CLASSIFICATION ACCURACY (%) OF COMPETITIVE METHODS ON THE SALINASA DATA WITH 1% TRAINING SAMPLES PER CLASS

	Uncorrected				Corrected			
Class	2D-SSA	SS-LRR	FS^{2}LRL	2D-MSSP	2D-SSA	SS-LRR	FS^{2}LRL	2D-MSSP
1	99.50	99.53	99.74	**100.00**	99.49	99.53	99.74	**100.00**
2	96.44	96.39	97.85	**99.93**	95.22	96.68	97.55	**99.85**
3	93.48	93.41	94.98	**99.36**	91.85	93.28	93.77	**98.23**
4	99.01	99.62	**99.95**	99.74	99.48	99.76	99.55	**100.00**
5	99.75	**99.76**	99.57	99.55	99.65	99.68	99.52	**99.70**
6	96.57	96.71	97.21	**99.63**	96.60	96.92	96.11	**99.12**
OA	97.49	97.67	98.38	**99.72**	97.12	97.79	97.88	**99.59**
κ	0.97	0.97	0.98	**1.00**	0.96	0.97	0.97	**0.99**
AA	97.42	97.57	98.24	**99.70**	97.06	97.64	97.73	**99.48**

classifies uncorrected and corrected data when the training size is larger than 5%. This validates the noise-robust property of 2-D-MSSP, where the superior results are mainly owing to the comprehensive spatial–spectral features even in noisy and water absorption bands.

As for other methods, SVM-spe, SS-LRR, MSuperPCA, and ASMGSSK produce better results on corrected SalinasA in most cases. The 2-D-SSA performs inconsistently on the two kinds of data, where it only achieves better classification performance on uncorrected data when 1% and 10% training samples are used. The denoising method FS^{2}LR yields higher OA and κ on uncorrected SalinasA in most cases. This has validated its efficacy in noise removal and image recovery on the SalinasA HSI data. Meanwhile, 2-D-MSSP performs better on corrected data when the training size is larger than 5% per class.

Through the comparisons between SVM-spe, 2-D-SSA, 2-D-MSSA, and 2-D-MSSP, we can find that 2-D-SSA has the potential in noise-robust feature extraction especially when training size is 1% and 10%, and the incorporation of multiscale window size (2-D-MSSA) improves its ability in suppressing noise. Consequently, by combing the 2-D-MSSA and PCA, our approach 2-D-MSSP shows strong noise robustness and high classification accuracy on SalinasA image.

2) Visual and Class-Based Performance Assessment: Here, we select three best performing methods on SalinasA, i.e., 2-D-SSA, SS-LRR, and FS^{2}LRL to compare with our 2-D-MSSP in terms of class-based accuracy and classification maps with 1% of training samples per class used. As shown in Table V, our 2-D-MSSP still produces the best results on the SalinasA dataset, especially in four out of six classes on the uncorrected data, and generates the highest accuracy in all classes on the corrected dataset. Especially, 2-D-MSSP produces significant better classification accuracy than its peers on uncorrected SalinasA data in the classes #2, #3, and #6 (*Corn senesced green weeds*, *Lettuce romaine 4wk,* and *Lettuce romaine 7wk*). With 1% of training samples per class, only the SS-LRR performs better on the corrected data, while 2-D-SSA, FS^{2}LRL, and 2-D-MSSP all have better results on the uncorrected one.

The visual results of these four approaches are depicted in Fig. 5. As seen, our method 2-D-MSSP generates most accurate classification maps with better uniformity and boundary location on uncorrected SalinasA image. From these classification maps in Fig. 5, it is clear that the classes *Broccoli green weeds_1*, *Lettuce romaine 4wk,* and *Lettuce romaine 5wk*, and the classes *Lettuce romaine 6wk* and *Lettuce romaine 7wk* are easily misclassified with each other. Specifically, SS-LRR and FS^{2}LRL fail to correctly identify them on either uncorrected or corrected data as marked in magenta circles in Fig. 5(e)–(h). The 2-D-SSA on corrected dataset [Fig. 5(d)], 2-D-MSSP on two kinds of datasets [Fig. 5(i) and (j)] can better classify the *Broccoli green weeds_1* and *Lettuce romaine 4wk* as highlighted in the black circles. However, all involved methods fail to correctly classify the classes *Lettuce romaine 6wk* and *Lettuce romaine 7wk.* Fortunately, as depicted in black circles in Fig. 5(i) and (j), the proposed 2-D-MSSP can significantly improve the classification performance on these two classes especially on uncorrected datasets. This superiority thanks mainly to the combination of multiscale spatial and spectral features in 2-D-MSSP. Overall, 2-D-SSA, FS^{2}LRL, and 2-D-MSSP yield better classification

TABLE VI
OA (%) AND κ (IN PARENTHESES) OF DIFFERENT METHODS ON THE CORRECTED PAVIAU DATASET UNDER VARIOUS TRAINING SIZES

Samples	1%	3%	5%	10%
SVM-spe	88.75 (0.85)	92.12 (0.89)	92.93 (0.91)	93.84 (0.92)
2D-SSA	93.30 (0.91)	96.19 (0.95)	96.87 (0.96)	97.70 (0.97)
SS-LRR	92.12 (0.89)	95.10 (0.93)	95.92 (0.95)	96.71 (0.96)
FS^2LRL	93.40 (0.91)	95.86 (0.94)	96.58 (0.95)	97.12 (0.96)
MSuperPCA	95.43 (0.94)	97.66 (0.97)	98.18 (0.98)	98.78 (0.98)
ASMGSSK	98.17 (0.98)	99.11 (0.99)	99.32 (0.99)	99.57 (0.99)
2D-MSSA	95.85 (0.94)	97.06 (0.96)	97.92 (0.97)	98.48 (0.98)
2D-MSSP	**98.29 (0.98)**	**99.20 (0.99)**	**99.42 (0.99)**	**99.73 (1.00)**

maps on uncorrected SalinasA image, which is consistent with the class-based results.

C. Results From the PaviaU Dataset

We further analyze the quantitative, visual-, and class-based results on the PaviaU dataset. Here we only test on the corrected data since the uncorrected one is unavailable. The best results in each row in tables are marked in bold and the correctly or incorrectly classified regions in the classification maps are highlighted with black and magenta circles, respectively.

1) Quantitative Evaluation: Table VI summaries the OA and κ of comparing methods using different training sizes on corrected PaviaU dataset. This image has higher spatial resolution and lower dimension size, which reduces the Hughes phenomenon to some extent [4]. As shown, the classification performances of many approaches with very limited training samples are relatively better than those on Indian Pines dataset. However, there are various types of land cover classes in PaviaU leading it to be a complicated scene. As listed in Table VI, our proposed 2-D-MSSP still achieves the high classification accuracy and ranks first among these advanced methods in all cases. In addition, the superiority of 2-D-MSSP is more significant when the number of training samples is smaller. This again validates the efficacy of proposed strategies in the HSI classification with limited and unbalanced training samples. By using multiscale superpixel segmentation, ASMGSSK and MSuperPCA significantly improve the classification accuracy, which just follows 2-D-MSSP. The two denoising methods also exhibit promising performance though they pay more attention on noise removal rather than HSI classification.

In order to investigate the effectiveness of different strategies in our methods, we further compare the results generated by SVM-spe, 2-D-SSA, 2-D-MSSA, and 2-D-MSSP. From Table VI, it is clear that the spatial features extracted from 2-D-SSA can improve the accuracy by only using spectral information in SVM-spe with a gain up to 4.6%. Then through using multiscale spatial feature, the 2-D-MSSA further enhances the classification performance with a maximum gain of 2.6%. Finally, by combining multiscale spatial–spectral features from multiscale 2-D-SSA and PCA, our proposed method 2-D-MSSP achieve high classification accuracy in the PaviaU data.

2) Visual and Class-Based Performance Assessment: Here, the best performing methods, 2-D-SSA, MSuperPCA, and ASMGSSK are selected to compare with 2-D-MSSP in terms of visual- and class-based performance. As shown in Table VII, the proposed 2-D-MSSP achieves best results in terms of OA, κ, and AA. As for the class-based performance, it ranks first in six out of nine classes, which is followed by ASMGSSK that exhibits superiority in three out of nine classes. The 2-D-SSA and MSuperPCA show potential on the identification of class *Shadows*, though it generates inferior results of OA, κ, and AA.

TABLE VII
CLASSIFICATION ACCURACY (%) OF COMPETITIVE METHODS FOR THE CORRECTED PAVIAU DATA WITH 1% TRAINING SAMPLES PER CLASS

Class	2D-SSA	MSuperPCA	ASMGSSK	2D-MSSP
1	93.45	94.06	99.06	**99.75**
2	98.82	99.00	99.82	**100.00**
3	72.95	95.94	99.78	**100.00**
4	90.29	72.25	**93.09**	84.44
5	99.29	97.26	98.49	**100.00**
6	87.18	95.05	**100.00**	**100.00**
7	83.09	97.33	**99.84**	99.77
8	87.76	96.91	98.50	**99.08**
9	98.16	**99.73**	58.61	78.57
OA	93.30	95.43	98.17	**98.29**
κ	0.91	0.94	**0.98**	**0.98**
AA	93.45	94.06	94.13	**95.73**

The classification maps of ASMGSSK, MSuperPCA, 2-D-SSA, and 2-D-MSSP are shown in Fig. 6. As seen, the classes *Meadows*, *Trees,* and *Bare Soil* are easily confused with each other. From Fig. 6(c)–(f), it is clear that the 2-D-SSA fails to correctly discriminate the *Meadows* and *Trees* as marked by magenta circles in Fig. 6(c). Most of these misclassifications are further corrected by the MSuperPCA, ASMGSSK, and 2-D-MSSP as shown in black circles in Fig. 6(d) and (e). However, the 2-D-SSA, MSuperPCA, and ASMGSSK all have difficulty in correctly classifying the *Meadows* and *Bare Soil*. This mainly because that the visually sparse *Meadows* regions are easily to be spectrally confused with *Bare Soil*. As depicted in Fig. 6(c)–(e), many small misclassified areas spread all over *Bare Soil* regions in the classification maps of 2-D-SSA and ASMGSSK. With regard to the identification of class *Gravel*, the 2-D-SSA seriously misclassifies it with class *Self-Blocking Bricks*, while MSuperPCA, ASMGSSK, and 2-D-MSSP can better identify it. However, 2-D-SSA yields superior classification performance on discrimination of classes *Asphalt* and *Trees* than the other three methods. Overall, our 2-D-MSSP yields promising visual performance in PaviaU dataset with better region uniformity and boundary outlining. This validates the efficacy of 2-D-MSSP in HSI classification under limited training samples.

D. Comparing With Deep Learning Methods

In this subsection, our 2-D-MSSP is further compared with four state-of-the-art deep learning methods, including CD-CNN [25], DR-CNN [26], SS-GAN [27], and SSCasRNN [28] on the corrected Indian Pines and PaviaU datasets with quantitative results summarized in VIII. Note that the exhibited results of our 2-D-MSSP using 1% training samples (equaling to 6 or 47 samples per class for Indian Pines and PaviaU, respectively), in comparison to 10% training samples used in SSCasRNN (equaling to 65 and 475 samples per class for two datasets).

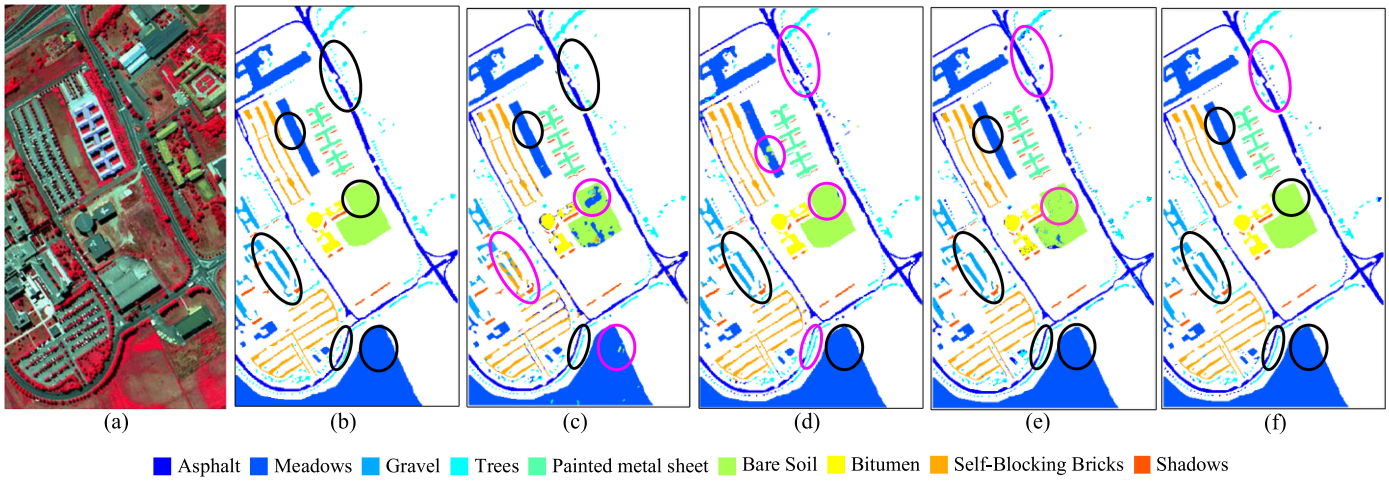

Fig. 6. Classification maps on the corrected PaviaU data (1% samples per class). (a) False color image (R: 834 nm, G: 650 nm, B: 550 nm), (b) Ground truth with nine classes. (c) 2-D-SSA (OA = 93.30%). (d) MSuperPCA (OA = 95.43%). (e) ASMGSSK (OA = 98.17%). (f) 2-D-MSSP (OA = 98.29%).

TABLE VIII
OA (%) OF 2-D-MSSP AND FOUR DEEP LEANING METHODS ON TWO DATASETS WITH DIFFERENT TRAINING SIZES (IN PARENTHESES)

Datasets	CD-CNN	DR-CNN	SSCas-RNN	SS-GAN	2D-MSSP
Indian Pines	84.43 (50)	88.87 (50)	91.79 (65)	90.28 (19)	**93.08 (6)**
PaviaU	92.19 (50)	96.91 (50)	90.30 (475)	97.61 (33)	**98.29 (47)**

The CD-CNN and DR-CNN use 50 training samples per class in each dataset, while the SS-GAN uses in total 300 training samples, which equals to around 19 and 33 samples per class in corrected Indian Pines and PaviaU datasets, respectively.

As seen in Table VIII, our 2-D-MSSP outperforms the deep learning approaches with smaller training size in the two datasets. Deep learning methods usually need more training samples to reach promising performance. Hence, we can conclude that when the training size is limited, the proposed 2-D-MSSP is more advantageous than these state-of-the-art deep learning methods. This experiment further demonstrates the superiority of our method in HSI classification.

V. CONCLUSION

In this article, we have proposed a novel spatial–spectral method for noise-robust feature extraction and effective HSI classification under limited and unbalanced training size, even without filtering out the unwanted noisy and water absorption bands. First, the 2-D-SSA with multiscale embedding window sizes are applied to exploit the multiscale local spatial features at different noise levels. Then, in each scale, the PCA is used to further extract spectral features and remove noisy information. The comprehensive combination of multiscale spatial–spectral features can effectively characterize image structures and remove noisy content, which can directly deal with the uncorrected HSI data even with very limited training samples. The experimental results on three HSI datasets have validated that our 2-D-MSSP model outperforms the other state-of-the-art approaches including several advanced deep learning methods with much larger training size. Especially, the superior classification results on uncorrected datasets provide potential to automatically interpret HSI without prefiltering noisy bands.

In the future, we will focus on the adaptive window size selection in 2-D-SSA to automatically assign the optimal window size to each pixel for more efficient HSI classification.

REFERENCES

[1] X. Cao, Y. Ji, L. Wang, B. Ji, L. Jiao, and J. Han, "Fast hyperspectral band selection based on spatial feature extraction," *J. Real-Time Image Process.*, vol. 15, no. 3, pp. 555–564, Apr. 2018.
[2] X. Cao, C. Wei, J. Han, and L. Jiao, "Hyperspectral band selection using improved classification map," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2147–2151, Oct. 2017.
[3] 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, Feb. 2015.
[4] H. Yu, L. Gao, W. Liao, B. Zhang, A. Pižurica, 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, Oct. 2017.
[5] L. Sun, C. Ma, Y. Chen, H. J. Shim, Z. Wu, and B. Jeon, "Adjacent superpixel-based multiscale spatial-spectral kernel for hyperspectral classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1905–1919, May 2019.
[6] L. Ji, L. Wang, and X. Geng, "An automatic bad band pre-removal method for hyperspectral imagery," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4985–4994, Dec. 2019.
[7] Y. Fu, A. Lam, I. Sato, and Y. Sato, "Adaptive spatial-spectral dictionary learning for hyperspectral image restoration," *Int. J. Comput. Vis.*, vol. 122, no. 2, pp. 228–245, Jun. 2017.
[8] F. Zhang, G. Yang, and J. Xue, "Hyperspectral image denoising based on low-rank coefficients and orthonormal dictionary," *Signal Process.*, vol. 6, Aug. 2020, Art. no. 107738.
[9] S. Jia, Y. Xie, G. Tang, and J. Zhu, "Spatial-spectral-combined sparse representation-based classification for hyperspectral imagery," *Soft. Comput.*, vol. 20, no. 12, pp. 4659–4668, Dec. 2016.
[10] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. Roy. Soc. A: Math., Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202.
[11] S. Mirzaei and S. Khosravani, "Hyperspectral image classification using non-negative tensor factorization and 3D convolutional neural networks," *Signal Process. Imag.*, vol. 76, pp. 178–185, Aug. 2019.

[12] R. Zaatour, S. Bouzidi, and E. Zagrouba, "Independent component analysis-based band selection techniques for hyperspectral images analysis," *J. Appl. Remote. Sens.*, vol. 11, no. 2, Apr. 2017, Art. no. 026006.
[13] J. Zabalza, R. Jinchang, W. Zheng, S. Marshall, and W. Jun, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1886–1890, Apr. 2014.
[14] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Jun. 2018.
[15] A. Zhang *et al.*, "Coastal wetland mapping with Sentinel-2 MSI imagery based on gravitational optimized multilayer perceptron and morphological attribute profiles," *Remote Sens.*, vol. 11, no. 8, Jan. 2019, Art. no. 952.
[16] J. Feng, L. Liu, X. Cao, L. Jiao, T. Sun, and X. Zhang, "Marginal stacked autoencoder with adaptively-spatial regularization for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3297–3311, Sep. 2018.
[17] X. Jia, B. C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.
[18] T. Zhang, P. Zhang, W. Zhong, Z. Yang, and F. Yang, "JL-GFDN: A novel gabor filter-based deep network using joint spectral-spatial local binary pattern for hyperspectral image classification," *Remote Sens.*, vol. 12, no. 12, Jan. 2020, Art. no. 2016.
[19] C. Shi and C. M. Pun, "3D multi-resolution wavelet convolutional neural networks for hyperspectral image classification," *Inf. Sci.*, vol. 420, pp. 49–65, Dec. 2017.
[20] X. Zhang, Z. Gao, L. Jiao, and H. Zhou, "Multifeature hyperspectral image classification with local and nonlocal spatial information via Markov random field in semantic space," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1409–1424, Nov. 2017.
[21] B. Tu, S. Huang, L. Fang, G. Zhang, J. Wang, and B. Zheng, "Hyperspectral image classification via weighted joint nearest neighbor and sparse representation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4063–4075, Nov. 2018.
[22] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020.
[23] A. Sellami, A. B. Abbes, V. Barra, and I. R. Farah, "Fused 3-D spectral-spatial deep neural networks and spectral clustering for hyperspectral image classification," *Pattern Recognit. Lett.*, vol. 138, pp. 594–600, Oct. 2020.
[24] G. Ding, Y. Guo, K. Chen, C. Chu, J. Han, and Q. Dai, "DECODE: Deep confidence network for robust image classification," *IEEE Trans. Image. Process.*, vol. 28, no. 8, pp. 3752–3765, Feb. 2019.
[25] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Trans. Image. Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.
[26] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE. Trans. Image. Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.
[27] Z. Zhong, J. Li, D. A. Clausi, and A. Wong, "Generative adversarial networks and conditional random fields for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Apr. 2018.
[28] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Mar. 2019.
[29] Y. Liu, C. Shan, Q. Gao, X. Gao, J. Han, and R. Cui, "Hyperspectral image denoising via minimizing the partial sum of singular values and superpixel segmentation," *Neurocomputing*, vol. 330, pp. 465–482, Feb. 2019.
[30] L. Sun, B. Jeon, B. N. Soomro, Y. Zheng, Z. Wu, and L. Xiao, "Fast superpixel based subspace low rank learning method for hyperspectral denoising," *IEEE Access*, vol. 6, pp. 12031–12043, Feb. 2018.
[31] F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Inf. Sci.*, vol. 397, pp. 48–68, Aug. 2017.
[32] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*. Berlin, Germany: Springer Science & Business Media, 2013.
[33] V. Florinsky Igor, *Digital Terrain Analysis in Soil Science and Geology*. Amsterdam, the Netherlands: Academic Press/Elsevier, 2012.
[34] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, "Noise reduction in hyperspectral imagery: Overview and application," *Remote Sens.*, vol. 10, no. 3, Mar. 2018, Art. no. 482.
[35] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.
[36] F. Xie, C. Lei, C. Jin, and N. An, "A novel spectral–spatial classification method for hyperspectral image at superpixel level," *Appl. Sci.*, vol. 10, no. 2, pp. 463, Jan. 2020.
[37] F. Li, P. Zhang, and H. Lu, "Unsupervised band selection of hyperspectral images via multi-dictionary sparse representation," *IEEE Access*, vol. 6, pp. 71632–71643, Nov. 2018.
[38] S. Holzwarth *et al.*, "HySens-DAIS 7915/ROSIS imaging spectrometers at DLR," in *Proc. 3rd EARSeL Workshop Imag. Spectr.*, May 2003, pp. 3–14.
[39] 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, Jul. 2015.
[40] A. Ajay, K. D. M. Dixon, V. Sowmya, and K. P. Soman, "Aerial image classification using GURLS and LIBSVM," in *Proc. Int. Conf. Commun. Signal Process.*, Apr. 2016, pp. 0396–0401.
[41] M. Wong, W. Abeysinghe, and C. C. Hung, "A massive self-organizing map for hyperspectral image classification," in *Proc. 10th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens.*, vol. , Sep. 2019, pp. 1–5.

Ping Ma received the B.Eng. and M.Eng degrees in surveying and mapping from the China University of Petroleum, Qingdao, China, in 2015 and 2018, respectively. She is currently working toward the Ph.D. degree at the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K.

Her research interests include multimodal remote sensing, digital signal processing, and intelligent optimization, including hyperspectral imaging, high spatial resolution data, LiDAR, and signal processing applications.

Jinchang Ren (Senior Member, IEEE) received the B.E. degree in computer software, in 1992, M.Eng. degree in image processing, in 1997, D.Eng. degree in computer vision, in 2000, all from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree in electronic imaging & media communication, in 2009, from the University of Bradford, Bradford, U.K.

He is currently a Professor of Computing Science, Robert Gordon University, Aberdeen, U.K. He has authored/coauthored more than 300 peer-reviewed journal/conferences papers, and acts as an Associate Editor for several international journals including IEEE JSTARS and *Journal of the Franklin Institute*. His research interests include hyperspectral imaging, image processing, computer vision, big data analytics, and machine learning.

Huimin Zhao was born in Shaanxi, China, in 1966. He received the B.Sc. and M.Sc. degrees in signal processing from NWPU, Xi'an, China, in 1992 and 1997, respectively, and the Ph.D. degree in electrical engineering from Sun Yat-sen University, Guangzhou, China, in 2001.

He is currently a Professor and Dean with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China. His research interests include image/video and information security technologies, and applications.

Genyun Sun (Member, IEEE) received the B.S. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2003, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2008.

He is currently an Associate Professor with the China University of Petroleum, Qingdao, China. His research interests include remote sensing image processing, hyperspectral remote sensing, high resolution remote sensing, and intelligent optimization algorithm.

Paul Murray received the M.Eng. degree from the University of Strathclyde, Glasgow, U.K., in 2008, and the Ph.D. degree from the University of Strathclyde, in 2012, both in electronic and electrical engineering.

He is currently a Senior Lecturer with the University of Strathclyde. His research interests include image processing, hyperspectral imaging and analysis, feature extraction, and machine learning.

Jiangbin Zheng received the Ph.D. degree in computer application from the School of Computer Science, NWPU, Xi'an, China, in 2002.

He is currently a Professor and Dean with the School of Software, NWPU. His research interests include stereo vision, 3-D reconstruction, forgery detection, and image watermarking.