SpaSSA: Superpixelwise adaptive SSA for unsupervised spatial-spectral feature extraction in hyperspectral image

Genyun Sun, *Member, IEEE*, Hang Fu, Jinchang Ren, *Member, IEEE*, Aizhu Zhang, Jamie Zabalza, Xiuping Jia, Senior *Member, IEEE* , Huimin Zhao

Abstract—**Singular Spectral Analysis (SSA) has recently been successfully applied to feature extraction in hyperspectral image (HSI), including conventional (1-D) SSA in spectral domain and 2-D SSA in spatial domain. However, there are some drawbacks such as sensitivity to the window size, high computational complexity under a large window and failing to extract joint spectral-spatial features. To tackle these issues, in this paper, we propose SpaSSA, i.e. superpixelwise adaptive SSA for exploiting local spatial information of HSI. Extraction of local (instead of global) features, particularly in HSI, can be more effective for characterizing the objects within an image. In SpaSSA, conventional SSA and 2-D SSA are combined and adaptively applied to each superpixel derived from an oversegmented HSI. According to the size of the derived superpixels, either SSA or 2D-SSA is adaptively applied for feature extraction, where the embedding window in 2D-SSA is also adaptive to the size of the superpixel. Experimental results on the three datasets have shown that the proposed SpaSSA outperforms both SSA and 2D-SSA in terms of classification accuracy and computational complexity. By combining SpaSSA with the principal component analysis (SpaSSA-PCA), the accuracy of land cover analysis can be further improved, outperforming several state-of-the-art approaches.**

Index Terms—**Hyperspectral image (HSI), feature extraction, remote sensing, land cover analysis, singular spectrum analysis (SSA), superpixelwise adaptive SSA (SpaSSA).**

I. INTRODUCTION

HYPERSPECTRAL image (HSI) presents a 3-D hypercube that contains a 2-D spatial structure and 1-D spectral signatures. With numerous and contiguous spectral bands acquired from visible light to (near) infrared, HSI is able to identify minor changes in terms of moisture, temperature, and chemical components of objects within the scene [1, 2]. The rich spectral and spatial information in HSI has enabled a number of emerging applications for object/change detection and condition monitoring – such as agriculture [3], military [4], environmental monitoring [5] and land cover analysis in remote sensing [6].

However, the high spectral resolution of HSI has also inevitably brought new challenges into data analysis. In classification of the HSI, the spectral dimension of the dataset, i.e. the number of bands, are usually much larger than the number of samples in the labelled classes within the of his. This has led to a serious mismatch between the complexity of the models and the number of training samples, namely the Hughes phenomenon [7-9]. In addition, data redundancy in both the spectral and spatial domains and noise caused by environmental factors and sensors need be considered [10]. To this end, effective feature extraction and dimensionality reduction is essential for land cover classification in HSI remote sensing.

In the last decades, a series of feature extraction methods have been developed, especially in the spectral domain. Some classic statistic-based methods, such as principal component analysis (PCA) [2, 11], and linear discriminant analysis (LDA) [12], maximum noise fraction (MNF) [13], have been designed for feature transformation. As one of the most widely used methods, PCA can transform high dimensional data into linearly uncorrelated variables, namely principal components (PCs). In HSI, several PCs images contain main feature information and the dimension is significantly reduced compared to original data. A selected subset of the first several PCs rather than the original spectral vectors are then used for classification; thus the feature dimension can be significantly reduced. Besides, some manifold learning methods are further developed to analyze the intrinsic features of HSI, including Laplacian eigenmaps (LE) [14], locally linear embedding (LLE) [15] and its extension robust local manifold representation (RLMR) [16]. However, these methods only consider the spectral information, leading to some limitations due to lack of spatial analysis.

J. Ren and H. Zhao are with the School of Computer Sciences, Guangdong Polytechnic University, Guangzhou, China. J. Ren is also with the National Subsea Centre, Robert Gordon University, Aberdeen, U.K.

J. Zabalza is with the Dept. of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G11XW, U.K..

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

In order to exploit the spatial characteristics of HSI, researchers have developed a number of approaches. One of the most widely used spatial feature extraction methods is based on morphological profiles (MPs) [17], which applied a series of structuring elements (SEs) to extract spatial geometrical details of images. By extracting MPs from several PCs of HSI [18], the extended MPs (EMPs) achieved a spatial-spectral feature extraction. Considering the effects of associated structuring elements (SEs) of MPs, morphological attribute profiles (APs) were proposed [19], where a sequence of morphological attribute filters were applied to reduced HSI components to obtain its characterized multilevel features for modelling different kinds of the structural information. In [20], extinction profiles (EPs) was introduced to precisely extract spatial features from HSI, and demonstrate their superiority in automatic feature extraction and data classification. In addition, some graph learning based methods have been proposed for spectral-spatial feature extraction in recent years. Luo *et.al* [21] proposed a novel spatial-spectral hypergraph discriminant analysis (SSHGDA) method to represent the complex intrinsic spatial-spectral of HSI. In [22], an enhanced hybrid-graph discriminant learning (EHGDL) method was proposed to reveal the complex high-order relationships of HSI and reduce the data dimension, which achieved the superior classification results with simple classifiers.

Recently, a time-series analysis technique, singular spectrum analysis (SSA), has been introduced for effective feature extraction in HSI [23, 24]. Given a pixel-based spectral vector, SSA can decompose it into several sub-components, where each sub-component has the same size as the original vector. By selecting sub-component(s) to reconstruct the spectral profile, useful information can be enhanced while noise or less representative signals can be effectively suppressed for improving classification accuracy [25]. Compared with PCA, SSA retains more spectral information and has better separability in subsequent task of data classification. Moreover, SSA can be combined with other methods such as Curvelet [10] for HSI classification. The extended 2-D version of SSA (2D-SSA) was proposed for the case of two-dimensional arrays of data, especially in images [26], which is believed to have similar capabilities with SSA [27]. Zabalza *et.al* [28, 29] further applied 2D-SSA to each band image of HSI for effective spatial feature extraction. Unlike conventional spatial features which are sensitive to fixed structuring elements or a small neighborhood (window), 2D-SSA can exploit global spatial correlation of HSI by embedding a Hankel–block–Hankel (HbH) trajectory matrix, outperforming a number of conventional approaches [28].

However, SSA and 2D-SSA still have some limitations in feature extraction of HSI. For conventional SSA, it only extracts the spectral trend features while ignores the abundant spatial context information, leading to misclassification and noise in the classification maps. Although 1.5D-SSA [30] combined some locally similar spectral features, the improvement in classification accuracy was limited. Applying SSA to the spatial domain seems to be more feasible to improve the classification performance. As for 2D-SSA, it has several main drawbacks. Firstly, considering the irregular shapes and inconsistent sizes of the ground objects in remote sensing images, it is inappropriate to extract spatial features with fixed rectangle windows as used in 2D-SSA, which may lead to poor classification results especially for small regions and the object's boundaries [31]. In other words, the regions used for spatial feature extraction should be adaptive to such spatial structures of the HSI. In addition, fix-sized embedding window (or extraction scale) cannot fully exploit the abundant spatial features. How to select the appropriate scale of feature extraction is worth exploring. Finally, conventional 2D-SSA suffers from large computational costs in terms of memory and processing, especially for both large embedding window and huge image.

In order to make full use of the advantages of SSA and 2D-SSA in feature extraction while to tackle their drawbacks, a novel superpixelwise adaptive singular spectrum analysis (SpaSSA) approach is proposed in this paper for effective feature extraction in HSI, aiming to deeply exploit the spatial features in different homogeneous regions. Experiments on the three publicly available datasets are fully validated the superior performance of the proposed SpaSSA approach. The main contributions of this paper can be summarized as follows.

1) A novel superpixelwise adaptive SSA (SpaSSA) approach is proposed for effective feature extraction in HSI. Whilst 2D-SSA is applied to the most of superpixel regions from each band to extract local spatial features of HSI, 1DSSA is used instead in the cases when the superpixels are too small. Actually, 2D-SSA and 1D-SSA are adaptively applied to local homogeneous regions so as to enhance the difference between different objects and preserve edge characteristics.

2) The embedding window size (or extraction scale) of the associated 2D-SSA is adaptively determined according to the size of the superpixels. In the adaptive criterion of SpaSSA, for most superpixels, the largest extraction scale within a reasonable range is selected to effectively improve the intra-class similarity, enhance the inter-class distinction, and fully separate the noise components. While for the remaining superpixels, small or large, 1DSSA and 2DSSA with relatively small fixed windows are applied respectively for efficiency.

3) As SpaSSA is performed in a superpixelwise way rather than on the full image, the size of the trajectory matrix and the computational complexity for SVD have been significantly reduced. Hence the computational cost is greatly reduced and the efficiency is improved.

The rest of this paper is organized as follows. Section II introduces the methodology of SSA and 2D-SSA. Section III descries the proposed SpaSSA classification method. The experimental results are presented in Section IV. Finally, the main concluding remarks are provided in Section V.

II. Background Introduction

In this section, we introduce conventional SSA and 2D-SSA methods. In terms of notations, we used capital letters to denote matrices, for example, X. A vector was represented as a lowercase letter, for example, x, while the italicized letter is used for denoting scalar, for example, *x*.

A. Singular Spectrum Analysis (SSA)

Originally proposed for time-series analysis, SSA can decompose a given signal into several independent components, i.e. the varying trend, oscillations, or noise [23]. Given a 1-D signal defined by $x = [x_1, x_2, ..., x_N] \in R^N$, the SSA algorithm can be briefly summarized in the following steps.

1) Embedding: Defining a window size $L \in Z$ where $L \in [1, N]$, the trajectory matrix X of the vector x can be constructed by:

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix} = (c_1, c_2, ..., c_K). \quad (1)$$

The columns c_i of the trajectory matrix is the lagged vector. The window size should be chosen properly as the size equals to the number of extracted components. The matrix X has equal values along the anti-diagonals and forms a Hankel matrix by definition.

2) SVD: Let the matrix $S=XX^T$, the Eigenvalues of S and their corresponding Eigenvectors are denoted as $(\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_L)$ and $(u_1, u_2, ..., u_L)$, respectively. After SVD, the trajectory matrix can be written as follows:

$$X = X_1 + X_2 + \cdots + X_L. \quad (2)$$

For simplicity, L is considered equal to the rank of X, hence we will have:

$$X_i = \sqrt{\lambda_i} u_i v_i^T \quad v_i = X^T u_i / \sqrt{\lambda_i}. \quad (3)$$

where X_i is called elementary matrix of rank one; u_i and v_i are the empirical orthogonal functions and the principal components of the trajectory matrix, respectively.

Matrices built by u_i and v_i are as follows:

$$\begin{aligned} U &= (u_1 \ u_2 \cdots u_L) \in R^{L \times L}, \\ V &= (v_1 \ v_2 \cdots v_L) \in R^{K \times L}. \end{aligned} \quad (4)$$

The ratio of each Eigenvalue $\lambda_i / \sum_{i=1}^{L} \lambda_i$ represents its contribution to the matrix X.

3) Grouping: The total set of L individual components is grouped into M independent subsets, denoted as $I = [I_1, I_2, ..., I_M]$. Each subset is composed of one or more elementary matrices X_i. After grouping, the trajectory matrix X becomes:

$$X = X_{I1} + X_{I2} + \cdots + X_{IM}. \quad (5)$$

4) Diagonal Averaging and Projection: The resulting matrices X_{Im}, $m \in [1, M]$ obtained from grouping are not necessarily Hankel type matrix as the original one. Thus, it is necessary to apply a diagonal averaging to each matrix by averaging those values in the matrix anti-diagonals and projecting into a new 1-D signals, a process known as Hankelisation. For a 1-D signal $y_m = [y_{m1}, y_{m2}, \cdots, y_{mN}] \in R^N$ transformed from X_{Im}, its elements after Hankelisation can be obtained as follows, where $a_{j,n-j+1}$ represents the elements of X_{Im}:

$$y_{mn} = \begin{cases} \dfrac{1}{n} \sum_{j=1}^{n} a_{j,n-j+1} & 1 \leq n \leq L \\[2mm] \dfrac{1}{L} \sum_{j=1}^{L} a_{j,n-j+1} & L < n < K \\[2mm] \dfrac{1}{N-n+1} \sum_{j=n-K+1}^{L} a_{j,n-j+1} & K \leq n \leq N \end{cases} \quad (6)$$

By repeating this for each matrix X_{Im}, the original 1-D signal x is reconstructed by:

$$x = y_1 + y_2 + \cdots + y_M = \sum_{m=1}^{M} y_m. \quad (7)$$

In HSI, SSA is used to extract the main trend of a given 1-D pixel vector. In general, the first reconstructed component, corresponding to the maximum Eigenvalue, can roughly replace the original data. This means that the Eigenvalue grouping (EVG) equals to 1, normally showing a good performance in classification [23]. Similarly, in our paper, we fix EVG to 1, and thus the performance of SSA is only dependent on its window size L_{1-D}. In addition, considering the limited performance of SSA in the spectral domain, we apply it in the spatial domain aiming to utilize spatial information especially for small objects, where is detailed in Section III.

B. Two-Dimensional Singular Spectrum Analysis (2D-SSA)

Compared with SSA, 2D-SSA has the same steps in SVD and grouping, yet the difference in embedding and diagonal averaging processes are detailed below.

1) Embedding a 2-D Signal: For an image P^{2D} with a size $N_x \times N_y$, its matrix representation is given as follows:

$$P^{2D} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,Ny} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,Ny} \\ \vdots & \vdots & \ddots & \vdots \\ p_{Nx,1} & p_{Nx,2} & \cdots & p_{Nx,Ny} \end{pmatrix}. \quad (8)$$

A 2-D window W_{ij} of size $L_x \times L_y$, where $L_x \in [1, N_x]$ and $L_y \in [1, N_y]$, is defined to construct the trajectory matrix:

$$W_{i,j} = \begin{pmatrix} p_{i,j} & p_{i,j+1} & \cdots & p_{i,j+Ly-1} \\ p_{i+1,j} & p_{i+1,j+1} & \cdots & p_{i+1,j+Ly-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i+Lx-1,j} & p_{i+Lx-1,j+1} & \cdots & p_{i+Lx-1,j+Ly-1} \end{pmatrix}. \quad (9)$$

This window uses the position of its top-left corner (i, j) as reference point, and the range of the reference point comprises $i \in [1, N_x - L_x + 1]$ and $j \in [1, N_y - L_y + 1]$. For a given reference pixel coordinates (i, j), the corresponding 2-D window is rearranged into a column vector $vcol_{i,j} \in R^{L_x L_y \times 1}$ as

$$vcol_{i,j} = \begin{pmatrix} p_{i,j} & p_{i,j+1} & \cdots & p_{i,j+Ly-1} & p_{i+1,j} & \cdots & p_{i+Lx-1,j+Ly-1} \end{pmatrix}^T \quad (10)$$

As the window $W_{i,j}$ moves from the top left to the bottom right of the image, there are $(N_x - L_x + 1) \times (N_y - L_y + 1)$ possible locations of the window. Therefore, the trajectory matrix $X^{2D} \in R^{L_x L_y \times (N_x - L_x + 1)(N_y - L_y + 1)}$ of the image P can be derived by:

$$X^{2D} = (vcol_{1,1}, vcol_{1,2}, \cdots vcol_{1,Ny-Ly+1}, vcol_{2,1}, \cdots vcol_{Nx-Lx+1,Ny-Ly+1})$$

(11)

Note that the trajectory matrix X^{2D} has a structure called HbH, i.e. Hankel by Hankel, which can be represented as

$$\mathbf{X}^{2D} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_{Nx-Lx+1} \\ \mathbf{H}_2 & \mathbf{H}_3 & \cdots & \mathbf{H}_{Nx-Lx+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{Lx} & \mathbf{H}_{Lx+1} & \cdots & \mathbf{H}_{Nx} \end{pmatrix}_{Lx \times (Nx-Lx+1)}, \quad (12)$$

$$\mathbf{H}_r = \begin{pmatrix} p_{r,1} & p_{r,2} & \cdots & p_{r,Ny-Ly+1} \\ p_{r,2} & p_{r,3} & \cdots & p_{r,Ny-Ly+2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r,Ly} & p_{r,Ly+1} & \cdots & p_{r,Ny} \end{pmatrix}_{Ly \times (Ny-Ly+1)}. \quad (13)$$

In simple terms, the HbH matrix ($\mathbf{X}^{2D}$) is Hankel in block terms, with each of the blocks ($\mathbf{H}_r$) being a Hankel matrix by itself.

2) *SVD and Grouping:* These two stages are the same as those in SSA. However, the respective dimensions of the different matrices change to 2-D accordingly. Specifically, $K_{2D} = (N_x - L_x + 1)(N_y - L_y + 1)$, and $L_{2D} = L_x \times L_y$.

3) *Diagonal Averaging:* Similar to the 1-D case, the resulting matrices $\mathbf{X}_m^{2D}$ in 2D-SSA are not necessarily HbH type. Therefore, it is necessary to transform them to HbH matrices by means of a two-step diagonal averaging process as given in (6), i.e., first applied within each block and then applied between blocks.

Let the 2-D signal $Z_m^{2D} \in R^{N_x \times N_y}$ be transformed from the group matrix $\mathbf{X}_m^{2D}$, it can be expressed as follows:

$$Z_m^{2D} = \begin{pmatrix} z_{m1,1} & z_{m1,2} & \cdots & z_{m1,Ny} \\ z_{m2,1} & z_{m2,2} & \cdots & z_{m2,Ny} \\ \vdots & \vdots & \ddots & \vdots \\ z_{mNx,1} & z_{mNx,2} & \cdots & z_{mNx,Ny} \end{pmatrix}. \quad (14)$$

Afterwards, the original 2-D image can be reconstructed by:

$$P^{2D} = Z_1^{2D} + Z_2^{2D} + \cdots + Z_M^{2D} = \sum_{m=1}^{M} Z_m^{2D}. \quad (15)$$

In 2D-SSA, the first decomposed component contains the main spatial trend information and thus is used to replace the original image for classification [28, 29]. Similar to SSA, the number of components used in representing the original image is fixed as 1 (EVG=1) and the window size $L_x \times L_y$ in embedding is the only parameter that affects performance.

III. THE PROPOSED SPASSA METHOD

The flowchart of the proposed SpaSSA method is presented in Fig. 1, which is composed of the following major parts: (i) generation of a superpixel map, (ii) superpixel-based adaptive SSA and (iii) Classification, corresponding to the three sections of this chapter. The details are described below.

A. Generation of a Superpixel Map

Superpixel segmentation provides a manner to compute local image features that can group pixels in local regions as clusters [32]. Superpixel algorithms generally consist of graph-based and clustering-based segmentation methods. Typical approaches incluse ERS [33] and simple linear iterative clustering (SLIC) [34], which are widely used in pre-processing or post-processing of HSI.

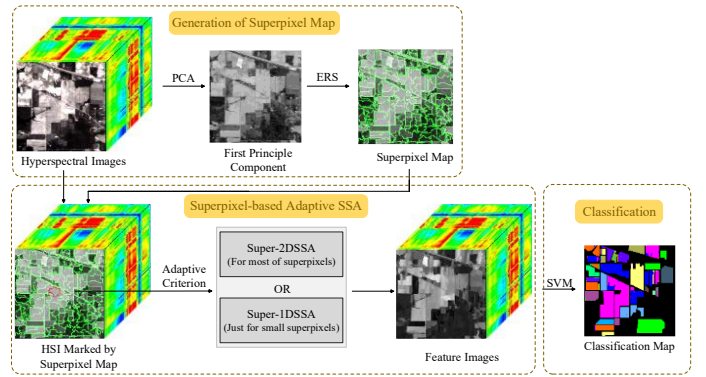In ERS, the image is firstly mapped to a graph, in which each

Fig. 1. Schematic of the proposed SpaSSA for HSI classification framework.

pixel is considered as a vertex and the pairwise similarities are defined as edge weights. The superpixel segmentation is defined by graph topology maximization, using an objective function to yield compact and homogeneous superpixels of similar sizes [35]. The SLIC method has the same promising performance as ERS. It can obtain regular superpixels that adhere well and efficiently to boundaries, but it only considers the similarity of pixels and, therefore, global image properties cannot be captured, which may result in under segmentation errors [34]. For this reason, in our method, we adopt ERS to obtain the segmented regions.

In order to reduce the computational cost, we firstly apply PCA on HSI and obtain the first PCs. Since the first PCs of HSI correspond to the largest Eigenvalue, it contains the major spatial information and can be used as the base image for the segmentation [36]. To this end, we choose an appropriate number of superpixel and perform ERS on the first PC to obtain the superpixel region map.

B. Superpixel-based Adaptive SSA (SpaSSA)

After obtaining the segmented superpixels, the proposed SpaSSA is applied on each superpixel for extracting the spatial features as detailed below.

1) Super-2DSSA

We denote superpixel and 2D-SSA based feature extraction as Super-2DSSA. As 2D-SSA can only deal with regular image matrices in a rectangle shape, adaptation is needed for processing irregularly shaped superpixel.

Actually, as shown in Fig. 2, Super-2DSSA is applied to each band on the extracted map of superpixels. For each superpixel, some adjacent pixels that may belong to other superpixels are included to form a rectangle region, and 2D-SSA is applied to this regular pixel region, where the results are the extracted superpixel from this region after processing. In 2D-SSA, an appropriate embedding window is used to extract spatial features of these local regions, so that it can obtain spatial trend
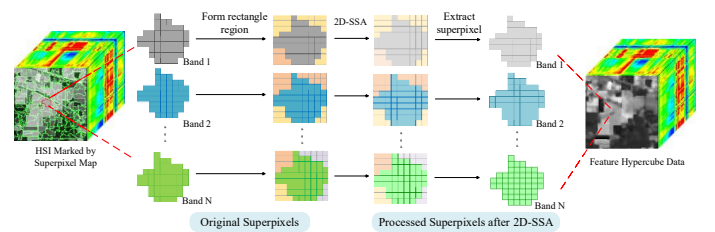

Fig. 2. Schematic of Super-2DSSA.

information and eliminate noise for improved feature extraction. After applying this process to all band images, we can obtain the features for the whole hypercube that can be used to replace the original data for classification.

2) Super-1DSSA

Although 2D-SSA can extract spatial information from superpixels, it fails to cope with small superpixels, when the embedding window becomes too small to extract sufficient components of spatial information. Therefore, a 1-D embedding window is used for such small superpixels. In other words, we replace 2D-SSA by 1D-SSA to deal with the cases of small superpixels, which forms a new scheme, i.e. Super-1DSSA. The way Super-1DSSA works is different from Super-2DSSA.

As shown in Fig. 3, all pixels in a small superpixel region are stretched into a 1-D spatial vector, and a 1-D embedding window is applied on this vector. After SSA, these processed pixels are returned to their original positions. This processing can effectively reduce differences of small homogeneous regions. In addition, as far as we know, this is the first time SSA is applied along a spatial direction, which can explore the spatial correlation of adjacent pixels in a local image.

3) Adaptive criterion of SpaSSA

Superpixelwise adaptive SSA is further employed to combine Super-2DSSA and Super-1DSSA for effective feature extraction. The adaptation of this approach is two-fold, i.e. adaptive selection of Super-2DSSA and Super-1DSSA methods for a given superpixel region, and adaptive determining the size of the embedding window. In general, Super-2DSSA is used for the large superpixels while Super-1DSSA is used for small ones. According to our experience and also as suggested in [27, 28], the optimal 2D-SSA embedding window varies according to the shape and size of the superpixels. To this end, we propose criteria for adaptively selecting the SSA methods and related parameters as detailed below.

Let S_{sp} be a parameter to measure the size of a superpixel region, where *col* and *row* are the width and height of the superpixel, respectively:

$$S_{sp} = \min\ (col, row).\qquad(16)$$

Considering its symmetric implementation, the range of the embedding window in 2D-SSA ($L_x \times L_y$ with $L_x = L_y$ for simplicity) is [1, N_x / 2] and [1, N_y / 2] in an image of size $N_x \times N_y$. Thus, the range of Super-2DSSA embedding window size L_x (or L_y) is [1, S_{sp} /2].

Two thresholds (T_1 and T_2, where $T_2 > T_1$) are used to decide the size of the embedding window for adaptive SSA as follows:

$$L_{SA-SSA} = \begin{cases} L_{1-D} & S_{sp}/2 < T_1 \\ L_x = L_y = floor(S_{sp}/2) & T_1 \le S_{sp}/2 < T_2 \\ L_x = L_y = T_2 & S_{sp}/2 \ge T_2 \end{cases}\quad(17)$$

where L_{SpaSSA} represents the window size, L_{1-D} denotes the length of 1-D embedding window of Super-1DSSA, and $L_x \times L_y$ is the 2-D embedding window of Super-2DSSA; *floor(S_{sp} /2)* represents the floor operation that calculates the largest integer not greater than S_{sp} /2.
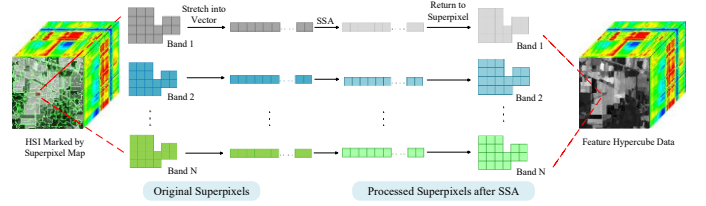

Fig. 3. Schematic of Super-1DSSA.

As shown in (17), for a given superpixel region, if the region size is smaller than threshold T_1, Super-1DSSA is chosen for this region with a 1-D embedding window size of L_{1-D}. On the other hand, if the region size is larger than T_1, Super-2DSSA is selected. For a given region, the embedding window of Super-2DSSA is defined as *floor(S_{sp} / 2)*. If the superpixel region is larger than T_2, a fixed window is used to replace the adaptive window, mainly because the adaptive window will lead to too much computation time not necessarily achieving better results.

Note that for a segmented image, the majority of superpixel regions are processed by Super-2DSSA with adaptively determined embedding windows for effective mining of spatial information in local regions. On the other hand, small superpixels are inevitable in most cases, for which Super-1DSSA is used to as a complementary method to improve the extraction of spatial regions.

C. Classification

After feature extraction, the SVM classifier is employed for pixel based classification of HSI. Based on margin criterion, SVM shows high robustness to dimensional problems. In fact, SVM has been widely used in HSI for classification [37, 38]. In addition, there are several available software tools and libraries for the implementation of SVM for fast and accurate data modeling.

Moreover, SVM supports different types of kernel functions, including linear, polynomial, and Gaussian radial basis functions (RBF), thus it can be flexibly applied in a wide range of data classification problems. As such, SVM is also used in our paper for evaluating the efficacy of the extracted features. Following the suggestions from others [39, 40], the RBF kernel is also used in our SVM classifier.

IV. EXPERIMENTS AND ANALYSIS

A. Datasets and Experimental Setup

Three publicly available hyperspectral datasets with available ground truth are adopted in our experiments for performance evaluation of the proposed approach.

The first dataset is Indian Pines, which was acquired by the AVIRIS sensor [41] over an agricultural study site in Northwest Indiana, USA. With a spatial size of 145×145 pixels, it contains 224 contiguous bands with the wavelength ranging from 400nm to 2500 nm. In our experiments, following the same procedure as others [23, 28], 24 water absorption or noisy bands are removed hence only 200 bands are remained for classification. In Fig.4, the false color image of dataset and its ground reference map are presented in (a) and (b), respectively. There

are 16 land-cover classes in this dataset, including agriculture, forest and vegetation, as shown in Fig. 4(d) by a color legend.
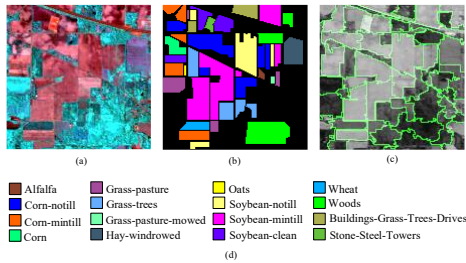

Fig. 4. (a) False color image of the Indian Pines dataset. (b) Ground truth map. (c) Segmented map (*Ns*=50). (d) Class names and legend of the different land-cover classes.

The second dataset is Pavia University, which was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor. This dataset has a spatial size of 610×340 pixels and 115 bands with a wavelength range of 430–860 nm. In our experiments, the number of bands is reduced to 103 by removing 12 noisy bands. Fig. 5 (a) and (b) present the false color image and related ground reference map with 9 different land-cover classes as defined in Fig. 5(d).
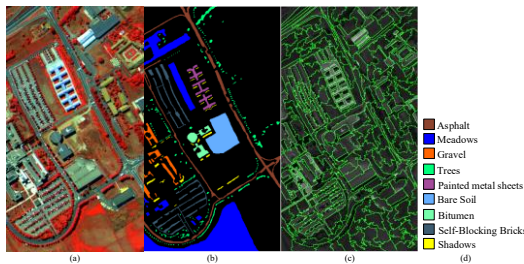

Fig. 5. (a) False color image of the Pavia University dataset. (b) Ground truth map. (c) Segmented map (*Ns*=300). (d) Class names and legend of the different land-cover classes.

The third dataset is Salinas, which was also acquired by the AVIRIS sensor over an agricultural area of Salinas Valley in California, USA. With a spatial size of 512×217 pixels, this dataset contains 224 bands, presenting a spatial resolution of 3.7 m. Again 20 water absorption bands are removed and the remaining 204 bands are used for classification. In Fig. 6, the false color image and the ground reference map are shown in (a) and (b), where the defined 16 different land-cover classes are shown in (d).
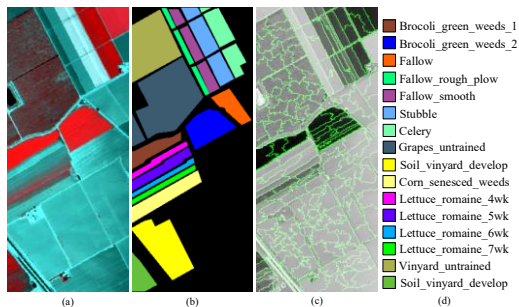

Fig. 6. (a) False color image of the Salinas dataset. (b) Ground truth map. (c) Segmented map (*Ns*=200). (d) Class names and legend of the different land-cover classes.

In the stage of classification, the SVM classifier [37] is implemented by the LIBSVM library [42] using the RBF kernel with five-fold 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, which account for 10% (Indian Pines), 1% (Pavia University) and 1% (Salinas) of the total samples in each class, respectively. In addition, three objective quality indexes, i.e., overall accuracy (OA), average accuracy (AA), kappa coefficient and class-by-class accuracy are utilized to evaluate the performance of image classification.

B. Parameter Analysis

The proposed approach has several key parameters, which include: (i) number of superpixels Ns; (ii) the 1-D embedding window L_{1-D} within Super-1DSSA; and (iii) two thresholds T_1 and T_2. The sensitivity analysis of the parameters is below.

Firstly, the number of superpixels Ns determines the degree of oversegmentation when producing the superpixels, which is expected to have a significant influence on the extracted features. Considering the spatial resolutions and sizes of the three datasets are different, the optimal Ns values for these datasets are actually different. In our method, Ns is set to 50, 300 and 200 for Indian Pines, Pavia University, and Salinas datasets, respectively, where the detailed parameter tuning results are given in Fig.7. For the size of superpixels, it is determined by the number of them. The larger the total number of segmented superpixels, the smaller the corresponding size. The determined Ns in our experiment can make the segmentation areas of different datasets as uniform as possible, as shown in Figs. 4-6 (c). In addition, different superpixel regions tend to have similar numbers of pixels, which enables as many superpixels as possible to be processed by 2DSSA with adaptive embedding windows.

On the contrary, the influence of L_{1-D} on the classification accuracy is insignificant, as the Super-1DSSA is only used to deal with a few small superpixels. These conclusions are drawn mainly from a separate experiment, where relevant results from the three dataset are shown in Fig. 8. In this experiment, the values of Ns are set to 400, 1000 and 300 for the three datasets respectively to have sufficient numbers of small superpixels to be processed by Super-1DSSA, where we have T_1=3 and T_2=11. As can be seen, the classification accuracy obtained by from different values of L_{1-D} varies in a small range, where the running time also changes slightly. Similar to conventional SSA, a large embedding window brings extra computing costs. Considering both the efficiency and efficacy, we fix the L_{1-D} to 10 in Super-1DSSA.

Finally, the thresholds T_1 and T_2 mainly determine the embedding window of the SpaSSA, affecting directly the extracted spatial information. In order to take more advantage of the combination of Super-1DSSA and Super-2DSSA, another experiment was conducted to analyze the impact of the two parameters. In the experiment, the two thresholds are selected from a sensible range of 3-15, and other parameters have their optimal values determined as mentioned above.
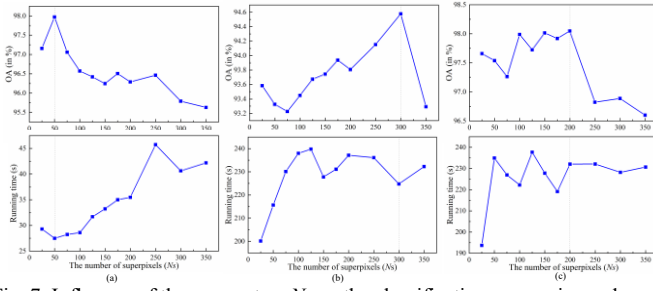
Fig. 7. Influence of the parameters Ns on the classification accuracies and running time of the SpaSSA the datasets of Indian Pines (a), Pavia University (b) and Salinas (c).
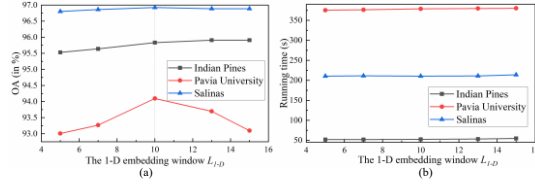

Fig. 8. Influence of the parameter L_{1-D} for the three datasets in terms in terms of OA (a) and running time (b).
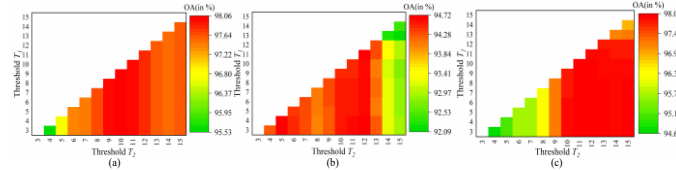

Fig. 9. Influence of the parameters T_1 and T_2 on the classification accuracies of the SpaSSA method for the datasets of Indian Pines (a), Pavia University (b) and Salinas (c).

According to the classification results shown in Fig. 9, the threshold T_2 is the key parameter that affects the performance of the proposed method compared with T_1. The optimal T_2 is in a relatively concentrated range of 9-13, which indicates that a suitable window size improves the accuracy in Super-2DSSA. In addition, for a fixed T_2, the increase of T_1 leads to a degraded accuracy. This is because of too many superpixels being processed by Super-1DSSA, which has a weaker feature extraction effect than Super-2DSSA. In this paper, the thresholds T_1 and T_2 are fixed to 3 and 11, respectively, for a balance between high classification accuracy and fast implementation.

C. Classification Accuracy Comparison

In this section, the efficacy of the features extracted from the proposed SpaSSA are compared with those from i) the original hyperspectral data, ii) 1D-SSA, iii) Super-1DSSA, iv) 2D-SSA, v) Super-2DSSA, where ours are numbered vi) for comparison. The classification accuracy from SVM on the three datasets is used for performance assessment.

For the SSA method, the window size L is selected as 10. For 2D-SSA, the embedding window is fixed as 5×5 and 10×10 as suggested in [28]. In Super-1DSSA, the parameter L_{1-D} is also 10. The embedding window of Super-2DSSA is 5×5 to ensure that all sizes of superpixels can be processed. For the proposed SpaSSA method, the parameters including L_{1-D}=10, T_1=3, T_2=11 and the optimal Ns value corresponding to three different datasets are used. Moreover, the related parameters C and γ of the Gaussian RBF kernel in the SVM classifier are optimally

obtained by a five-fold cross validation. Specifically, γ=0.125 for all three datasets; whilst C=1000 for the Indian Pines and Pavia University datasets, and C=10000 for Salinas. The classification maps and quantitative results of the proposed approach, in comparison to five benchmarking ones, are shown in Figs. 10–12 and Tables I–III, respectively.
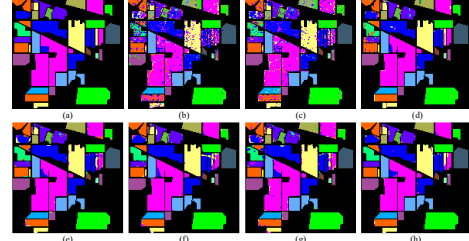

Fig. 10. Classification results for the Indian Pines dataset (10% training samples). (a) Ground Truth. (b) SVM (OA=79.75%). (c) SSA (OA=84.02%). (d) Super-1DSSA (OA=95.13%). (e) 2D-SSA in 5×5 window (OA=95.76%). (f) 2D-SSA in 10 × 10 window (OA=96.00%). (g) Super-2DSSA (OA=93.79%). (h) SpaSSA (OA=97.97%).

According to the quantitative results in Tables I–III, the proposed SpaSSA method always obtains the highest classification accuracy on three datasets. For Indian Pines, the OA can be significantly increased from 79.75% to 97.97% compared to the SVM that performs on the raw HSI data. The SpaSSA method also beats the 2D-SSA method and improves the classification accuracy by about 2%. In addition, Super-1DSSA achieves superior classification accuracy, even higher than 2D-SSA in some ground classes. Due to the small embedding window, the Super-2DSSA did not effectively take the advantages of 2DSSA and superpixel to extract spatial features, leading to lower accuracy. Our SpaSSA method integrated their advantages and achieved the best accuracy on most of the ground classes. For Pavia University, the OA of our proposed method can increase from 89.21% to 94.58% compared with the raw data. Among all other compared methods, the Super-2DSSA method has best performance, and 2D-SSA with 10×10 window achieves the highest accuracy in meadows, painted metal sheets, bitumen and self-blocking bricks classes. Although only two classes achieved the highest accuracy, the SpaSSA can obtain higher accuracy than 2D-SSA and Super-2DSSA in terms of OA and kappa. For Salinas, the SpaSSA still maintained its effectiveness and outperformed all other compared methods in most of classes and three indexes. These have clearly demonstrated the efficacy of the proposed feature extraction approach for HSI classification, especially with a small number of training samples.

From Figs. 10-12, the proposed SpaSSA has yielded the best classification maps from three datasets in comparison to other approaches. Although SSA can improve the misclassification from the raw data, it performed especially not well on the Pavia University dataset. 2D-SSA can smooth the ground object and significantly improve the classification results. However, there are still misclassification within large ground objects when using a small processing window (5×5), whilst small ground objects tend to be misclassified or missed when a large processing window (10×10) is used, especially for the Pavia

TABLE I
CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES DATASET (10% TRAINING)

Class	Train	Test	SVM	SSA (L=10)	Super-1DSSA (L_{1-D}=10)	2DSSA (5×5)	2DSSA (10×10)	Super-2DSSA (3×3)	SpaSSA
Alfalfa	5	41	14.63	73.17	**97.56**	70.73	78.05	60.98	85.37
Corn-notill	143	1285	75.33	81.87	91.05	92.45	91.67	91.52	**96.11**
Corn-mintill	83	747	66.67	77.38	92.10	94.11	96.52	92.10	**96.79**
Corn	24	213	46.01	70.42	76.06	95.31	98.59	87.79	**99.53**
Grass-pasture	49	434	85.71	94.47	**99.08**	97.70	98.16	96.77	97.70
Grass-trees	73	657	92.39	92.54	99.09	98.63	97.26	98.17	**99.24**
Grass-pasture-mowed	3	25	68.00	80.00	72.00	88.00	96.00	80.00	**100.0**
Hay-windrowed	48	430	99.07	99.77	99.77	98.37	98.60	98.84	**100.0**
Oats	2	18	33.33	55.56	94.44	**100.0**	72.22	88.89	**100.0**
Soybean-notill	98	874	75.06	84.21	**95.31**	92.91	93.59	90.85	94.97
Soybean-mintill	246	2209	84.56	84.93	97.92	96.42	97.28	94.21	**98.87**
Soybean-clean	60	533	61.54	74.30	85.18	92.87	90.99	84.43	**97.56**
Wheat	21	184	91.85	90.76	99.46	98.91	98.37	97.28	**99.46**
Woods	127	1138	93.76	91.39	99.12	98.77	98.95	98.77	**99.38**
Buildings	39	347	57.35	49.57	93.08	98.85	98.85	95.97	**99.71**
Stone-Steel-Towers	10	83	79.52	85.54	96.39	**100.0**	96.39	**100.0**	**100.0**
OA			79.75	84.02	95.13	95.76	96.00	93.79	**97.97**
AA			70.30	80.37	92.97	94.63	93.84	91.03	**97.79**
kappa×100			76.79	81.78	94.44	95.16	95.43	92.92	**97.69**

TABLE II
CLASSIFICATION ACCURAY (%) FOR THE PAVIA UNIVERSITY DATASET (1% TRAINING)

Class	Train	Test	SVM	SSA (L=10)	Super-1DSSA (L_{1-D}=10)	2DSSA (5×5)	2DSSA (10×10)	Super-2DSSA (3×3)	SpaSSA
Asphalt	67	6564	88.88	87.22	88.86	**95.17**	83.82	92.66	92.78
Meadows	187	18462	96.13	94.83	97.62	98.54	**99.60**	97.94	99.44
Gravel	21	2078	78.73	80.08	72.28	84.70	70.84	86.62	**87.63**
Trees	31	3033	**87.11**	86.61	81.73	82.95	79.56	86.05	82.86
Painted metal sheets	14	1331	98.87	98.65	98.72	99.77	**100.0**	99.62	99.77
Bare Soil	51	4978	78.81	70.89	90.14	91.10	96.75	91.58	**97.29**
Bitumen	14	1316	75.30	79.79	83.51	83.05	**91.57**	80.47	89.36
Self-Blocking Bricks	37	3645	75.47	71.28	78.63	86.06	**88.70**	87.65	87.46
Shadows	10	937	**99.68**	**99.68**	83.67	88.37	61.69	99.15	77.91
OA			89.21	87.26	90.65	93.60	91.96	93.62	**94.58**
AA			86.55	85.45	86.13	89.97	85.83	**91.31**	90.50
kappa×100			85.60	83.00	87.56	91.47	89.30	91.50	**92.80**

TABLE III
CLASSIFICATION ACCURACY (%) FOR THE SALINAS DATASET (1% TRAINING)

Class	Train	Test	SVM	SSA (L=10)	Super-1DSSA (L_{1-D}=10)	2DSSA (5×5)	2DSSA (10×10)	Super-2DSSA (3×3)	SpaSSA
Brocoli_green_weeds_1	21	1988	98.14	98.99	**100.0**	93.91	93.01	98.94	97.64
Brocoli_green_weeds_2	38	3688	98.43	97.80	99.76	99.40	99.38	99.70	**99.92**
Fallow	20	1956	99.03	99.34	**99.90**	99.34	99.44	99.64	98.93
Fallow_rough_plow	14	1380	97.54	97.90	90.36	93.41	90.43	97.03	**98.84**
Fallow_smooth	27	2651	97.74	98.45	97.43	93.89	99.21	95.66	**99.96**
Stubble	40	3919	99.46	99.44	99.92	**100.0**	**100.0**	99.64	99.64
Celery	36	3543	99.44	99.41	98.70	99.41	99.24	98.65	**99.72**
Grapes_untrained	113	11158	76.41	79.55	93.90	92.36	97.10	90.93	**97.29**
Soil_vinyard_develop	63	6140	98.89	98.53	**99.63**	99.38	98.83	99.20	98.99
Corn_senesced_green	33	3245	85.95	86.29	92.94	93.68	95.13	93.00	**95.25**
Lettuce_romaine_4wk	11	1057	90.35	89.88	95.65	91.01	94.80	89.88	**98.39**
Lettuce_romaine_5wk	20	1907	98.95	98.95	99.84	99.63	95.07	**100.0**	99.42
Lettuce_romaine_6wk	10	906	98.34	98.01	97.68	**99.78**	97.46	99.34	99.23
Lettuce_romaine_7wk	11	1059	88.67	88.48	85.27	92.26	**98.96**	88.20	96.60
Vinyard_untrained	73	7195	63.95	68.51	81.86	87.80	**95.91**	78.67	**95.91**
Vinyard_vertical_trellis	19	1788	92.39	91.95	96.48	**97.82**	97.37	97.48	97.15
OA			88.07	89.32	94.80	95.09	97.33	93.72	**98.05**
AA			92.73	93.22	95.58	95.82	96.96	95.37	**98.31**
kappa×100			86.71	88.10	94.20	94.53	97.03	93.00	**97.83**

University and Salinas datasets. The other two superpixel-based SSA methods, Super-1DSSA and Super-2DSSA have different degrees of classification performance on different datasets, in other words, the performance of the two methods alone are limited. In contrast, the proposed SpaSSA method can effectively overcome these drawbacks and obtain the highest classification accuracy.

In addition, we further evaluate how the number of training samples may affect the classification accuracy, and the results are shown in Fig. 13. For the Indian Pines dataset, the number of training samples is ranged from 1% to 10%, while for the Pavia University and Salinas datasets, the number of training samples is increased from 0.1% to 1%. Some observations from Fig. 11 can be summarized as follows. First, increasing number of training samples can obviously improve the classification accuracy on all three datasets. For the SpaSSA method, for example, the OA on the Indian Pines dataset has been improved from 78.27% to 98.06% when the training samples increases

from 1% to 10%. In almost all the cases, the proposed SpaSSA has obtained the highest accuracy.

In Table IV, we further compare the running time of our approach and those from compared methods. For SSA, its computing time shows a linear increment as the image size, number of pixels, increases. While Super-1DSSA can save more calculation time than SSA, because it processes each superpixel region instead of pixels. As 2D-SSA is applied to each band image rather than each pixel as SSA does, the number of iterations has been significantly reduced. However, as the size of the embedding window increases, the computation time increases exponentially, due mainly to the large HbH matrix and the following on SVD decomposition. Although the embedding window of Super-2DSSA is small, the number of superpixels increases the processing times of 2DSSA, leading to the rise of computational cost inevitably. SpaSSA takes more time than Super-2DSSA. How to further improve the efficiency of SpaSSA will be further investigated.
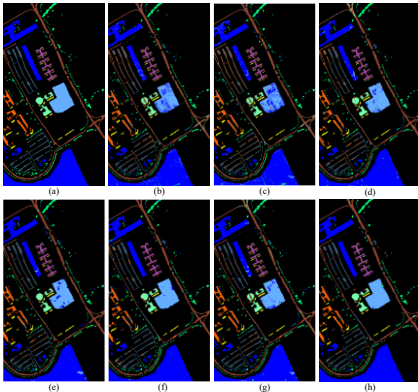
Fig. 11. Classification results for the Pavia University dataset (1% training samples). (a) Ground Truth. (b) SVM (OA=89.21%). (c) SSA (OA=87.26%). (d) Super-1DSSA (OA=90.65%). (e) 2D-SSA in 5×5 window (OA=93.60%). (f) 2D-SSA in 10 × 10 window (OA=91.96%). (g) Super-2DSSA (OA=93.62%). (h) SpaSSA (OA=94.58%).
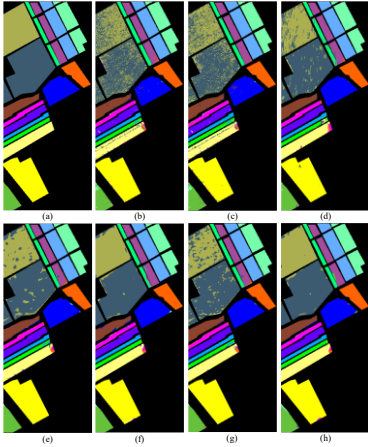

Fig. 12. Classification results for the Salinas dataset (1% training samples). (a) Ground Truth. (b) SVM (OA=88.07%). (c) SSA (OA=89.32%). (d) Super-1DSSA (OA=94.80%). (e) 2DSSA in 5×5 window (OA=95.09%). (f) 2D-SSA in 10 × 10 window (OA=97.33%). (g) Super-2DSSA (OA=93.72%). (h) SpaSSA (OA=98.11%).

D. Comparison with other State-of-the-Art Approaches

Considering the great potential in combining 2D-SSA with PCA [28], the proposed SpaSSA method is also combined with
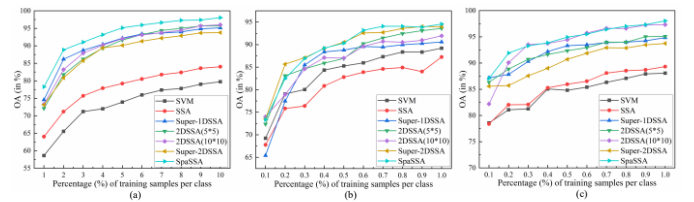

Fig. 13. Effect of the number of training samples on the proposed SpaSSA method and the compared methods for the (a) Indian Pines, (b) Pavia University and (c) Salinas datasets.

TABLE IV
COMPUTATION TIME FOR DIFFERENT SSA METHODS ON THREE DATASETS

Method	Parameters	Time(s)		
		Indian Pines	Pavia University	Salinas
SSA	10	8.59	112.80	39.33
Super-1DSSA	10	**4.20**	79.52	**34.44**
2D-SSA	5×5	8.25	**40.50**	44.11
	10×10	14.41	80.69	85.25
Super-2DSSA	3×3	18.51	195.82	154.57
SpaSSA	Optimal parameter	27.96	229.02	215.85

PCA, aiming to achieve more effective extraction of spectral-spatial features. Accordingly, several spectral-spatial feature extraction and classification approaches are used for benchmarking, which are divided into two groups. In the first group, the original hypercube is used for classification without dimensionality reduction, including 3-D convolutional neural network (3DCNN) [43], spectral-spatial residual network (SSRN) [44], double-branch multi-attention mechanism network (DBMA) [45], local adaptive joint sparse representation (LAJSR) [46], and our SpaSSA. In the second group, dimension reduction methods are focused including PCA, robust local manifold representation (RLMR) [16], superpixelwise PCA (Super-PCA) [47] and 2DSSA-PCA [28], as well as SpaSSA combined with PCA (SpaSSA-PCA). The best classification results from each method on the three datasets are given in Tables V–VII for comparison.

For experimental settings, the methods in the first group select the optimal parameters according to [43-46]. For the dimension reduction methods, the dimension is reduced to 20 for the four methods, where 2DSSA and SpaSSA use the same

TABLE V
CLASSIFICATION ACCURACY (%) COMPARISON WITH THE STATE-OF-THE-ARTS FOR THE INDIAN PINES DATASET (ALMOST 10% TRAINING)

	Original Data of Full Dimension (200).					Reduced Dimension of Data (20).				
	3DCNN	SSRN	DBMA	LAJSR	SpaSSA	PCA	RLMR	Super-PCA	2DSSA-PCA	SpaSSA-PCA
OA	90.38	98.22	**98.65**	96.71	98.00	72.49	84.30	95.91	96.03	**98.34**
AA	88.39	98.08	96.98	95.09	**98.52**	67.67	86.63	96.05	95.48	**98.45**
kappa	0.890	**0.980**	0.976	0.949	0.978	0.686	0.821	0.951	0.953	**0.982**

TABLE VI
CLASSIFICATION ACCURACY (%) COMPARISON WITH THE STATE-OF-THE-ARTS FOR THE PAVIA UNIVERSITY DATASET (ALMOST 1% TRAINING)

	Original Data of Full Dimension (103).					Reduced Dimension of Data (20).				
	3DCNN	SSRN	DBMA	LAJSR	SpaSSA	PCA	RLMR	Super-PCA	2DSSA-PCA	SpaSSA-PCA
OA	89.30	96.26	**98.28**	93.11	94.72	82.73	85.45	90.83	95.02	**95.86**
AA	83.97	94.99	**98.35**	90.51	92.50	78.41	87.35	86.16	92.53	**92.84**
kappa	0.858	0.950	**0.981**	0.904	0.938	0.769	0.834	0.878	0.937	**0.944**

TABLE VII
CLASSIFICATION ACCURACY (%) COMPARISON WITH THE STATE-OF-THE-ARTS FOR THE SALINAS DATASET (ALMOST 1% TRAINING)

	Original Data of Full Dimension (204).					Reduced Dimension of Data (20).				
	3DCNN	SSRN	DBMA	LAJSR	SpaSSA	PCA	RLMR	Super-PCA	2DSSA-PCA	SpaSSA-PCA
OA	89.35	95.56	98.04	94.68	**98.09**	90.45	90.45	**99.30**	97.50	99.23
AA	92.34	96.54	**98.85**	92.05	97.93	93.71	89.32	98.91	98.06	**99.29**
kappa	0.881	0.951	**0.978**	0.914	0.977	0.894	0.891	0.989	0.973	**0.990**

configurations as given in Section IV.C and Super-PCA uses the optimal number of superpixels as in [47]. The same three datasets are used, yet the training percentage is a little different from previous experiments. Considering the training mode of CNN, the number of selected training samples is relatively even. The percentage of training samples is about 10%, 1% and 1% for the Indian Pines, Pavia University and Salinas datasets, respectively.

As shown in Tables V-VII, the classification accuracy of SpaSSA is superior to 3DCNN and LAJSR, while lower than SSRN and DBMA. For the Indian Pines and Salinas datasets, the performance of SpaSSA is very close to SSRN and DBMA, and even better than SSRN on the latter dataset. However, SpaSSA achieved relatively low accuracy on the Pavia University dataset compared with SSRN and DBMA, mainly because the trend features extracted by SpaSSA ignore some small ground features. Besides, the SpaSSA-PCA further improves the accuracy of SpaSSA, outperforming all other dimension-reduction based methods, despite a slightly lower OA than Super-PCA on Salinas. The performance of SpaSSA-PCA is comparable to state-of-the-art SSRN and DBMA, and even better than the latter on Indian Pines and Salinas. In conclusion, the proposed SpaSSA method has superior classification performance, and the extended SpaSSA-PCA can indeed further improve the efficacy of spatial-spectral features extraction for improved classification.

V. Conclusion

Feature extraction is an essential stage in hyperspectral data classification, which can deeply exploit data characteristics in both the spectral and spatial domains. Many spectral and spatial feature extraction techniques have been proposed in recent years, such as SSA and its 2-D version (2D-SSA). However, 2D-SSA only considers the global spatial information yet ignores the local spatial context of the pixels, i.e. object-alike regions, which has limited its applicability and efficiency. To this end, a novel superpixelwise adaptive SSA (SpaSSA) is proposed to tackle these issues.

Superpixel segmentation can group pixels into homogeneous regions namely object-alike superpixels. The adaptation of the SSA methods with appropriate parameters setting has been proved effective for extracting spatial features, allowing improved discrimination ability. The proposed SpaSSA approach can effectively reduce the intra-class variance within superpixels and enhance the discrimination among different superpixels, leading to improved classification accuracy.

Experiments carried out on three publicly available hyperspectral datasets have fully validated the superior performance of the proposed SpaSSA approach over existing SSA and 2D-SSA. In addition, the combination of SpaSSA with principal component analysis (SpaSSA-PCA) can further improve the classification accuracy and outperform several state-of-the-art approaches.

Future works will focus on more efficient implementation of SpaSSA and its new applications in remote sensing image analysis, which include i) weakly supervised learning [48, 49] and context-aware learning [50, 51], ii) various deep learning including rotation-invariant CNN [52], autoencoders [53] and deep fusion [54], and iii) saliency detection [55] even with selected bands [56, 57].

References

[1] L. He, J. Li, C. Liu, and S. Li, "Recent Advances on Spectral–Spatial Hyperspectral Image Classification: An Overview and New Guidelines," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, no. 3, pp. 1579-1597, 2018.

[2] J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, and J. Han, "Novel Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 93, pp. 112-122, 2014.

[3] S. D. Fabiyi, H. Vu, C. Tachtatzis, P. Murray, D. Harle, T. K. Dao, I. Andonovic, J. Ren, and S. Marshall, "Varietal Classification of Rice Seeds Using RGB and Hyperspectral Images," *IEEE Access,* vol. 8, pp. 22493-22505, 2020.

[4] C. Zhao, X. Li, J. Ren, and S. Marshall, "Improved sparse representation using adaptive spatial support for effective target detection in hyperspectral imagery," *International Journal of Remote Sensing,* vol. 34, no. 24, pp. 8669-8684, 2013.

[5] J. Pontius, M. Martin, L. Plourde, and R. Hallett, "Ash decline assessment in emerald ash borer-infested regions: A test of tree-level, hyperspectral technologies," *Remote Sensing of Environment,* vol. 112, no. 5, pp. 2665-2676, 2008/05/15/, 2008.

[6] T. Qiao, Z. Yang, J. Ren, P. Yuen, H. Zhao, G. Sun, S. Marshall, and J. A. Benediktsson, "Joint bilateral filtering and spectral similarity-based sparse representation: A generic framework for effective feature extraction and data classification in hyperspectral imaging," *Pattern Recognition,* vol. 77, pp. 316-328, 2018.

[7] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory,* vol. 14, no. 1, pp. 55-63, 1968.

[8] F. Luo, H. Huang, Y. Duan, J. Liu, and Y. Liao, "Local Geometric Structure Feature for Dimensionality Reduction of Hyperspectral Imagery," *Remote Sensing,* vol. 9, pp. 790, 08/01, 2017.

[9] F. Luo, L. Zhang, X. Zhou, T. Guo, Y. Cheng, and T. Yin, "Sparse-Adaptive Hypergraph Discriminant Analysis for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters,* vol. 17, no. 6, pp. 1082-1086, 2020.

[10] T. Qiao, J. Ren, Z. Wang, J. Zabalza, M. Sun, H. Zhao, S. Li, J. A. Benediktsson, Q. Dai, and S. Marshall, "Effective Denoising and Classification of Hyperspectral Images Using Curvelet Transform and Singular Spectrum Analysis," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 1, pp. 119-133, 2017.

[11] S. Prasad, and L. M. Bruce, "Limitations of Principal Components Analysis for Hyperspectral Target Recognition," *IEEE Geoscience and Remote Sensing Letters,* vol. 5, no. 4, pp. 625-629, 2008.

[12] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of Hyperspectral Images With Regularized Linear Discriminant Analysis," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 47, no. 3, pp. 862-873, 2009.

[13] G. Lixin, X. Weixin, and J. Pei, "Segmented minimum noise fraction transformation for efficient feature extraction of hyperspectral images," *Pattern Recognition,* vol. 48, 04/01, 2015.

[14] B. Li, Y.-R. Li, and X.-L. Zhang, "A survey on Laplacian eigenmaps based manifold learning methods," *Neurocomputing,* vol. 335, pp. 336-351, 2019/03/28/, 2019.

[15] Y. Fang, H. Li, Y. Ma, K. Liang, Y. Hu, S. Zhang, and H. Wang, "Dimensionality Reduction of Hyperspectral Images Based on Robust Spatial Information Using Locally Linear Embedding," *IEEE Geoscience and Remote Sensing Letters,* vol. 11, no. 10, pp. 1712-1716, 2014.

[16] D. Hong, N. Yokoya, and X. X. Zhu, "Learning a Robust Local Manifold Representation for Hyperspectral Dimensionality Reduction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 10, no. 6, pp. 2960-2975, 2017.

[17] R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J. C. Chan, and F. Canters, "Improved Classification of VHR Images of Urban Areas Using Directional Morphological Profiles," *IEEE*

Transactions on Geoscience and Remote Sensing, vol. 46, no. 10, pp. 2803-2813, 2008.

[18] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 43, no. 3, pp. 480-491, 2005.

[19] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological Attribute Profiles for the Analysis of Very High Resolution Images," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 48, no. 10, pp. 3747-3762, 2010.

[20] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. A. Lotufo, "Extinction Profiles for the Classification of Remote Sensing Data," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 54, no. 10, pp. 5631-5645, 2016.

[21] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature Learning Using Spatial-Spectral Hypergraph Discriminant Analysis for Hyperspectral Image," *IEEE Transactions on Cybernetics,* vol. 49, no. 7, pp. 2406-2419, 2019.

[22] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality Reduction With Enhanced Hybrid-Graph Discriminant Learning for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 58, no. 8, pp. 5336-5353, 2020.

[23] J. Zabalza, R. Jinchang, W. Zheng, S. Marshall, and W. Jun, "Singular Spectrum Analysis for Effective Feature Extraction in Hyperspectral Imaging," *IEEE Geoscience and Remote Sensing Letters,* vol. 11, no. 11, pp. 1886-1890, 2014.

[24] J. Zabalza, J. Ren, Z. Wang, H. Zhao, J. Wang, and S. Marshall, "Fast Implementation of Singular Spectrum Analysis for Effective Feature Extraction in Hyperspectral Imaging," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 8, no. 6, pp. 2845-2853, 2015.

[25] N. Golyandina, and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*, 2013.

[26] L. Rodríguez-Aragón, and A. Zhigljavsky, "Singular spectrum analysis for image processing," *Statistics and Its Interface,* vol. 3, 01/01, 2010.

[27] N. Golyandina, and K. Usevich, "2D-Extension of Singular Spectrum Analysis: Algorithm and Elements of Theory," *Matrix Methods: Theory, Algorithms and Applications*, 10/26, 2010.

[28] J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li, and S. Marshall, "Novel Two-Dimensional Singular Spectrum Analysis for Effective Feature Extraction and Data Classification in Hyperspectral Imaging," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 53, no. 8, pp. 4418-4433, 2015.

[29] 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," *Journal of the Franklin Institute,* vol. 355, no. 4, pp. 1733-1751, 2018/03/01/, 2018.

[30] H. Fu, G. Sun, J. Zabalza, A. Zhang, J. Ren, and X. Jia, "A Novel Spectral-Spatial Singular Spectrum Analysis Technique for Near Real-Time In Situ Feature Extraction in Hyperspectral Imaging," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 13, pp. 2214-2225, 2020.

[31] T. Liu, Y. Gu, J. Chanussot, and M. Dalla Mura, "Multimorphological Superpixel Model for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 12, pp. 6950-6963, 2017.

[32] S. Jia, K. Wu, J. Zhu, and X. Jia, "Spectral–Spatial Gabor Surface Feature Fusion Approach for Hyperspectral Imagery Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 57, no. 2, pp. 1142-1154, 2019.

[33] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation." pp. 2097-2104.

[34] M. Van den Bergh, X. Boix, G. Roig, and L. Van Gool, "SEEDS: Superpixels Extracted Via Energy-Driven Sampling," *International Journal of Computer Vision,* vol. 111, no. 3, pp. 298-314, 2015/02/01, 2015.

[35] C. Yang, L. Bruzzone, H. Zhao, Y. Tan, and R. Guan, "Superpixel-Based Unsupervised Band Selection for Classification of Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, no. 12, pp. 7230-7245, 2018.

[36] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–Spatial Classification of Hyperspectral Images With a Superpixel-Based

Discriminative Sparse Model," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 53, no. 8, pp. 4186-4201, 2015.

[37] F. Melgani, and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 42, no. 8, pp. 1778-1790, 2004.

[38] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification," *IEEE Transactions on Image Processing,* vol. 17, no. 4, pp. 622-629, 2008.

[39] R. Archibald, and G. Fann, "Feature Selection and Classification of Hyperspectral Images With Support Vector Machines," *IEEE Geoscience and Remote Sensing Letters,* vol. 4, no. 4, pp. 674-677, 2007.

[40] B. Demir, and S. Ertürk, "Empirical Mode Decomposition of Hyperspectral Images for Support Vector Machine Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 48, no. 11, pp. 4071-4084, 2010.

[41] R. Green, M. Eastwood, C. Sarture, T. Chrien, M. Aronsson, B. Chippendale, J. Faust, B. Pavri, C. Chovit, M. Solis, M. Olah, and O. Williams, "Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sensing of Environment,* vol. 65, pp. 227-248, 08/31, 1998.

[42] C. Chih-Chung, and L. Chih-Jen, "Libsvm: a library for support vector machines," *ACM Trans. Intell. Syst. Technol., vol. 2(3),* pp. 1-27, 01/01, 2011.

[43] Y. Li, H. Zhang, and Q. Shen, "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network," *Remote Sensing,* vol. 9, no. 1, 2017.

[44] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, no. 2, pp. 847-858, 2018.

[45] W. Ma, Q. Yang, Y. Wu, W. Zhao, and X. Zhang, "Double-Branch Multi-Attention Mechanism Network for Hyperspectral Image Classification," *Remote Sensing,* vol. 11, no. 11, 2019.

[46] J. Peng, X. Jiang, N. Chen, and H. Fu, "Local adaptive joint sparse representation for hyperspectral image classification," *Neurocomputing,* vol. 334, pp. 239-248, 2019/03/21/, 2019.

[47] 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 Transactions on Geoscience and Remote Sensing,* vol. 56, no. 8, pp. 4581-4593, 2018.

[48] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, E. Tsougenis, Q. Huang, M. Cai, and P. A. Heng, "Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis," *IEEE Transactions on Cybernetics,* vol. 50, no. 9, pp. 3950-3962, 2020.

[49] Y. F. Li, L. Z. Guo, and Z. H. Zhou, "Towards Safe Weakly Supervised Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 43, no. 1, pp. 334-346, 2021.

[50] B. Li, W. Xiong, O. Wu, W. Hu, S. Maybank, and S. Yan, "Horror Image Recognition Based on Context-Aware Multi-Instance Learning," *IEEE Transactions on Image Processing,* vol. 24, no. 12, pp. 5193-5205, 2015.

[51] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Context-Aware Local Binary Feature Learning for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 5, pp. 1139-1153, 2018.

[52] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning Rotation-Invariant and Fisher Discriminative Convolutional Neural Networks for Object Detection," *IEEE Transactions on Image Processing,* vol. 28, no. 1, pp. 265-278, 2019.

[53] P. Xuan, L. Gao, N. Sheng, T. Zhang, and T. Nakaguchi, "Graph Convolutional Autoencoder and Fully-Connected Autoencoder with Attention Mechanism Based Method for Predicting Drug-Disease Associations," *IEEE Journal of Biomedical and Health Informatics,* vol. 25, no. 5, pp. 1793-1804, 2021.

[54] C. Ge, Q. Du, W. Sun, K. Wang, J. Li, and Y. Li, "Deep Residual Network-Based Fusion Framework for Hyperspectral and LiDAR Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 14, pp. 2458-2472, 2021.

[55] F. Guo, W. Wang, Z. Shen, J. Shen, L. Shao, and D. Tao, "Motion-Aware Rapid Video Saliency Detection," *IEEE Transactions on*

Circuits and Systems for Video Technology, vol. 30, no. 12, pp. 4887-4898, 2020.

[56]   C. Chang, S. Wang, K. Liu, M. Chang, and C. Lin, "Progressive Band Dimensionality Expansion and Reduction Via Band Prioritization for Hyperspectral Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 4, no. 3, pp. 591-614, 2011.

[57]   H. Sun, J. Ren, H. Zhao, G. Sun, W. Liao, Z. Fang, and J. Zabalza, "Adaptive Distance-Based Band Hierarchy (ADBH) for Effective Hyperspectral Band Selection," *IEEE Trans Cybern*, Mar 24, 2020.

Genyun Sun (M'16) received the BSc degree from Wuhan University, Wuhan, China, in 2003 and Ph.D. degree in Institute of Remote Sensing Applications, Chinese Academy of Sciences in 2008. He is currently a Professor with China University of Petroleum (East China), Qingdao, China.

His research interests cover remote sensing image processing including hyperspectral and high resolution remote sensing, and intelligent optimization..

Hang Fu received the BSc degrees in geomatics engineering from China University of Petroleum (East China), Qingdao, China, in 2019. He is currently pursuing the PhD degree in geomatics engineering at China University of Petroleum (East China), Qingdao, China.

His research interests include feature extraction and hyperspectral classification.

Jinchang Ren (Senior Member, IEEE) received BEng, MEng and D.Eng. degrees from the Northwestern Polytechnical University, Xian, China in 1992, 1997 and 2000, respectively. He was also awarded a Ph.D. degree in 2019 from the University of Bradford, Bradford, U.K.

He is currently a Professor with the National Subsea Centre, Robert Gordon University, Aberdeen, U.K. His research interests include image processing, computer vision, machine learning, and big data analytics. He acts as Associate Editor for several international journals, including IEEE Transactions on Geoscience and Remote Sensing and J. Franklin Institute et al.

Aizhu Zhang (M'19**)** received the BSc, MSc and Ph.D. degree egrees from China University of Petroleum (East China), Qingdao, China, in 2011, 2014, and 2017, respectively. She is currently a Lecturer with China University of Petroleum (East China), Qingdao, China.

Her research interests are in artificial intelligence, pattern recognition, city remote sensing, and wetland remote sensing.

Jaime Zabalza (M'14) received the M.Eng. degree in industrial engineering from the Universitat Jaume I (UJI), Castellón de la Plana, Spain, in 2006; and the M.Sc. degree (with distinction) and the Ph.D. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 2012 and 2015, respectively. He was awarded the IET Image & Vision Section prize for best Ph.D. thesis for his work in hyperspectral remote sensing under the supervision of Prof. Jinchang Ren.

He is currently with the Department of Electronic and Electrical Engineering, University of Strathclyde. His research interests include hyperspectral data analysis as well as signal and image processing in a wide range of applications..

Xiuping Jia (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982 and the Ph.D. degree in Electrical Engineering from The University of New South Wales, Australia, in 1996. Since 1988, she has been with the School of Engineering and Information Technology, The University of New South Wales at Canberra, Australia, where she is currently an Associate Professor. Her research interests include remote sensing, hyperspectral image processing and spatial data analysis. Dr. Jia has authored or coauthored more than 270 referred papers, including over 160 journal papers, addressing various topics including data correction, feature reduction and image classification using machine learning techniques. She has coauthored of the remote sensing textbook titled *Remote Sensing Digital Image Analysis* [Springer-Verlag, 3rd (1999) and 4th eds. (2006)]. She is a Subject Editor for the Journal of Soils and Sediments and an Associate Editor-in-Chief of the IEEE Transactions on Geoscience and Remote Sensing. She is an IEEE Fellow.

Huimin Zhao was born in Shaanxi, China, in 1966. He received the B.Sc. and the M.Sc. degree in signal processing in 1992 and 1997 from Northwestern Polytechnical University, Xian, China, respectively. He received the Ph.D. degree in electrical engineering from the Sun Yat-sen University in 2001.

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