Spectral-Spatial and Superpixelwise PCA for Unsupervised Feature Extraction of Hyperspectral Imagery

Xin Zhang, Xinwei Jiang, Junjun Jiang, Yongshan Zhang, Xiaobo Liu and Zhihua Cai

Abstract—As the most classical unsupervised dimension reduction algorithm, principal component analysis (PCA) has been widely used in hyperspectral images (HSIs) preprocessing and analysis tasks. Recently proposed superpixelwise PCA (SuperPCA) has shown promising accuracy where superpixels segmentation technique was first used to segment a HSI to various homogeneous regions and then PCA was adopted in each superpixel block to extract the local features. However, the local features could be ineffective due to the neglect of global information especially in some small homogeneous regions and/or in some large homogeneous regions with mixed groundtruth objects. In this paper, a novel spectral-spatial and superpixelwise PCA (S^3-PCA) is proposed to learn the effective and low-dimensional features of HSIs. Inspired by SuperPCA we further adopt superpixels based local reconstruction to filter the HSIs and use the PCA based global features as the supplement of local features. It turns out that the global-local and spectral-spatial features can be well exploited. Specifically, each pixel of a HSI is reconstructed by the nearest neighbors pixels in the same superpixel block, which could eliminate the noise and enhance the spatial information adaptively. After the local reconstruction based data preprocessing, PCA is performed on each region and the entire HSI to obtain local and global features, respectively. Then we simply concatenate them to get the global-local and spectral-spatial features for HSIs classification. The experimental results on two HSIs datasets demonstrate the superiority of the proposed method over the state-of-the-art methods. The source code the proposed model is available at **https://github.com/XinweiJiang/S3-PCA**.

Index Terms—Hyperspectral image, superpixel segmentation, principal component analysis (PCA), dimensionality reduction, spectral-spatial feature.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) contain hundreds of spectral bands that provide rich information to identify

X. Zhang, X. Jiang, Y. Zhang and Z. Cai are with the School of Computer Science, and also with Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China (e-mail: xinz2801@gmail.com, ysjxw@hotmail.com, yszhang.cug@gmail.com, zhcai@cug.edu.cn).

J. Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail:jiangjunjun@hit.edu.cn).

X. Liu is with the School of Automation, and also with the Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, China University of Geosciences, Wuhan 430074, China (e-mail:xbliu@cug.edu.cn).

and classify ground objects [1], [2]. Although the high-dimensional spectral features could provide possibility to perform accurate HSIs classification, the curse of dimensionality issue [3], [4] could arise especially in the case of high-dimensional HSIs and insufficient labeled samples. Besides, noise in HSIs could also decrease the accuracy of the HSIs classification models [5]. Therefore, it is necessary to conduct dimensionality reduction (DR) to find low-dimensional features from high-dimensional HSIs so as to reduce training time and avoid overfitting for effective HSIs classification.

Various DR techniques had been introduced into HSIs data preprocessing, which are typically divided into feature selection [6] and feature extraction [3], [7], [8] approaches. The former tends to select a representative subset from the original high-dimensional features, while the latter usually tries to project the original data from the high-dimensional observation space to a low-dimensional subspace. In this paper, we focus on the feature extraction based DR models.

According to the availability of label information, feature extraction algorithms can be roughly divided into unsupervised and supervised models. Principal component analysis (PCA) [9] could be the most widely used unsupervised DR algorithm. Without using label information, PCA aims to find the optimal linear projections to transform high-dimensional data into low-dimensional subspace by maximizing the variance in each projected dimension. To handle the nonlinear structures, the manifold learning models for HSIs feature extraction had attracted researchers' attention, where the inherent manifold structures embedded in high-dimensional observations are kept in low-dimensional subspace, e.g., neighborhood preserving embedding (NPE) [10], locality preserving projection (LPP) [11], local pixel NPE (LPNPE) [12] and local neighborhood structure preserving embedding (LNSPE) [13]. By contrast, supervised DR models use extra label information to extract discriminative low-dimensional features. The representative supervised DR methods include linear discriminant analysis (LDA) [14], local fisher dicriminant analysis (LFDA) [15], nonparametric weighted feature extraction (NWFE) [16], etc..

Although the aforementioned methods had been successfully used for HSIs features extraction, the accuracy of the models could be compromised due to noise and lack of spatial information. It is inevitable that there are noise in HSIs when the images were captured, which could bring difficulties for the feature extraction and classification. To denoise HSIs, various image filtering models have been introduced, such as mean, gabor and propagated filters [17], which could efficiently

reduce noise and smooth the homogeneous regions in HSIs. However, since the filters usually use various windows to filter images, they could fail when processing complex boundary regions. Alternatively, there are some reconstruction based methods for HSIs denoising and data augmentation. For example, low-rank representation with neighborhood preserving regularization (LRR_NP) [18] have been proposed where LRR with the neighbor preserving regularization were used to preprocess HSIs. In LNSPE [13] the dual structure preserving model were developed by considering the nearest neighbor reconstruction points into the neighborhood scatter extraction model, which can be regarded as a new filter. Local geometric structure Fisher analysis [19] was introduced where locally reconstructed samples based on NPE along with original data are used to construct the intrinsic graph and penalty graph which can be viewed as a type of data augmentation. But spatial information in HSIs are not considered in these models which could decrease the accuracy of the models.

In addition, it is acknowledged that neighbor pixels are likely to belong to the same class, so the spatial information of HSIs is beneficial to HSIs feature extraction and classification [1]. Among these methods the image segmentation technology based spatial models have shown promising performance in terms of accuracy and complexity, because it can effectively utilize spatial and spectral information of HSIs based on superpixels [20]–[25]. For example, Gu et. al. [24] proposed a superpixel based tensor model for HSIs classification, where a multiattribute superpixel tensor model was constructed on the top of multiattribute superpixel maps based on the concept of extended morphological profiles to fully exploit the third-order spectral–spatial information. Chen et. al. [25] proposed a superpixel-based bilateral filtering algorithm combined bilateral filtering and superpixel segmentation to effectively extract the spectral-spatial features based on superpixel. However, the two models highly depend on the filters. By contrast, Jiang et. al. [20] proposed superpixelwise PCA (SuperPCA) without adopting filters based on the assumption that distinct superpixel blocks should have different PCA based linear projections to extract low-dimensional and local HSIs features. The new model has shown high accuracy compared to typical DR methods based on spectral information and other spectral-spatial feature extraction techniques [21], [22]. Similar to SuperPCA, superpixelwise LPP [21], superpixelwise Kernel PCA [22] and superpixel-based LDA [23] were also developed to extract the nonlinear features in superpixel blocks by employing LPP, KPCA and LDA to replace PCA in Super-PCA, respectively. However, *SuperPCA and its variants only focused on the local spatial information and ignored the global structures of HSIs, which compromises the accuracy of the feature extraction methods. Especially in the cases of limited data in some small homogeneous regions with noise and/or mixed groundtruth objects in some large superpixel blocks, results from SuperPCA based models could be unsatisfactory.*

In this paper, we propose a novel spectral-spatial feature extraction method, termed spectral-spatial and superpixelwise PCA (S^3-PCA), which can efficiently make use of spatial information to eliminate noise through superpixels based local reconstruction and to extract the global-local and spectral-

spatial features for HSIs classification. Specifically, the contributions of the paper are listed as follows: 1)By performing superpixels based local reconstruction, the spatial information such as edges in HSIs could be enhanced, and the noise in superpixel blocks could be eliminated effectively. 2)By adopting PCA on the global HSI data and each superpixel block, the global and local features can be extracted, respectively. Then by simply concatenating them the global-local and spectral-spatial features can be acquired. 3)The experimental results on two HSIs data sets demonstrate the proposed method outperforms typical and state-of-the-art DR models.

The rest of this paper is organized as follows. In Section II we first review and introduce the related works including PCA and SuperPCA.The proposed method is described in Section III, and the experimental results are provided in Section IV, followed by the conclusion of our work.

II. RELATED WORKS

In this section, we briefly introduce the classic PCA and recently proposed SuperPCA.

A. PCA

Given the high-dimensional data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_P\} \in \mathbb{R}^{B \times P}$ with dimension B and the number of samples P, PCA tries to find a linear transformation from the original B-dimensional space to a low d-dimensional space $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_P\} \in \mathbb{R}^{d \times P}$ where the variances are maximized. Denote the transformation matrix as $\mathbf{W} \in \mathbb{R}^{B \times d}$, the linear projection between $\mathbf{X}$ and $\mathbf{Y}$ can be expressed by $\mathbf{Y} = \mathbf{W}^T \mathbf{X}$. Mathematically, the objective function of PCA is

$$\mathbf{W}^* = \underset{\mathbf{W}^T\mathbf{W}=\mathbf{I}}{\arg\max} \operatorname{Tr}(\mathbf{W}^T \operatorname{Cov}(\mathbf{X})\mathbf{W}), \qquad (1)$$

where $\operatorname{Cov}(\mathbf{X})$ is the covariance matrix of $\mathbf{X}$. In PCA, we select the eigenvectors $\mathbf{w}$ corresponding to the first d eigenvalues λ to form the projection matrix $\mathbf{W}$. Due to its effectiveness and efficiency, PCA has been widely used to extract low-dimensional features of HSIs data. In reality, the distribution of land objects presents the characteristics of regional homogeneity which means the closer the land objects, the more likely they are to belong to the same class and the reshaping from 3D cube to 2D matrix may ignore the differences of homogeneous regions which leads to poor results from PCA.

B. SuperPCA

SuperPCA [20] firstly adopted superpixel segmentation techniques to segment a HSI to various homogeneous regions, and then PCA is used in each superpixel block to extract low-dimensional and local features for HSIs classification. There are many superpixel segmentation methods. Roughly, the existing superpixel segmentation algorithms can be divided into the gradient-based methods [26] and graph-based methods [27]. Simple linear iterative clustering (SLIC) [26] is one of the most widely used gradient-based segmentation method due to its simplicity and effectiveness, which adopts a k-means clustering approach to generate superpixels blocks. A representative graph-based segmentation technique is the

Entropy Rate Segmentation (ERS) [27] which generates a 2D superpixel map. Due to its promising performance and high efficiency it is used for HSIs superpixel segmentation in SuperPCA.

Given a graph ($\mathcal{G} = (\mathbf{V}, \mathbf{E})$) for a HSI, the vertices set ($\mathbf{V}$) corresponds to the pixels in the image to be segmented and the edges set ($\mathbf{E}$) records the pairwise similarities. ERS tends to select a subset of edges $\mathbf{A} \subseteq \mathbf{E}$ such that the resulting graph, $\mathcal{G}^* = (\mathbf{V}, \mathbf{A})$, contains exactly K connected subgraphs. To generate the most suitable superpixel segmentation, the objective function of ERS is denoted by

$$\mathbf{A}^* = \arg \max_{\mathbf{A}} \text{Tr}\{H(\mathbf{A}) + \alpha B(\mathbf{A})\}, \text{s.t.} \mathbf{A} \subseteq \mathbf{E}. \quad (2)$$

In Eq. 2, $H(\mathbf{A})$ is the entropy rate term for generating homogeneous and compact clusters, while $B(\mathbf{A})$ is used to encourage the cluster with similar sizes. α is to balance the contributions of the entropy rate term $H(\mathbf{A})$ and the regularization term $B(\mathbf{A})$. To solve Eq. 2, a greedy heuristic algorithm is adopted. ERS has been proven to be a powerful superpixel segmentation method for HSIs spatial information extraction [20]. In SuperPCA, once ERS based superpixel segmentation is done, PCA is then used in each superpixel block to extract local spectral-spatial features for HSIs classification.

III. PROPOSED METHOD

In this section, we introduce the proposed spectral-spatial and superpixelwise PCA (S^3-PCA) model, which can efficiently eliminate the noise and extract the global-local and spectral-spatial features for HSIs classification. As can be seen from Fig. 1, a HSI is firstly divided into multiple homogeneous regions by ERS based superpixel segmentation technique, and then the pixels in each homogeneous region are spatially reconstructed by their neighbors belonging to the same superpixel block for HSI denoising. Secondly, SuperPCA based feature extraction is employed to obtain low-dimensional and local spectral-spatial features. Thirdly, in order to take into account the global structures we simply perform classic PCA on the reconstructed HSI globally to obtain the global features with low dimension, and then we concatenate and combine all these local and global features to form updated dimension-reduced HSI features. Finally, we make use of PCA again to reduce the dimension of the concatenated global-local features for classification. Compared to SuperPCA and other superpixel-based DR methods, the proposed method could efficiently extract the global-local and spectral-spatial features especially when dealing with pixels in some small homogeneous regions with noise and/or in some large superpixel blocks with mixed groundtruth objects.

Let's denote a HSI data by $\mathbf{X}_3 \in \mathbb{R}^{M \times N \times B}$, where M, N and B are the numbers of width, height, and spectral bands, respectively. Typically, we reshape it from 3D cube to 2D matrix for the subsequent feature extraction, leading to $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_P] \in \mathbb{R}^{B \times P}(P = MN)$ with each column vector $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{iB}]^T$ corresponding to one pixel of the HSI.

A. Superpixel based Local Reconstruction

Traditional spectral-spatial feature extraction methods try to consider the spatial information via a small spatial window [28], [29], but the spatial structures could not be effectively extracted because the local windows do not provide sufficient structural information and could cause spectral distortion [30]. To overcome the issues, we use the superpixel segmentation technique to extract the homogeneous objects, and then the spatial neighbor reconstruction model is developed to extract effective spatial information for each homogeneous region.

Superpixel segmentation can effectively extract spatial information of HSIs according to intrinsic features such as texture. In this paper, ERS is adopted to generate homogeneous regions from a HSI due to its high efficiency and performance. Before adopting ERS, we have to firstly perform PCA to obtain the first PC denoted by $I_f \in \mathbb{R}^{M \times N}$, to capture the major information of a HSI and reduce the computational burden in the process of superpixel segmentation. Then, ERS can be conducted on I_f to generate homogeneous superpixel blocks

$$I_f = \bigcup_{k=1}^{S} \mathcal{H}_k \quad \text{s.t.} \quad \mathcal{H}_k \cap \mathcal{H}_g = \emptyset, (k \neq g) \quad (3)$$

where S is the number of superpixels, and $\mathcal{H}_k$ represents the $k-$th superpixel.

To further reduce noise and obtain effective spectral-spatial features in superpixel blocks, inspired by LNSPE [13] the superpixel based local reconstruction method is developed for HSIs denoising. Different from LNSPE where local structure is based on k nearest neighbors without considering spatial distribution of HSIs, in this paper local reconstruction based on superpixels is performed for each pixel in every superpixel block as shown in the top right corner of Fig. 1. We aim to reconstruct each sample based on the rest data in the same superpixel block, that is, the spatial nearest neighbor pixels.

Specifically, for each sample $\mathbf{x}_i \in \mathbb{R}^{B \times 1}$ in a superpixel block, its k nearest spatial neighbors belonging to the same superpixel can be denoted by $\mathbf{Z}_i = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k\} \in \mathbb{R}^{B \times k}$. Let's further define the similarity weight between $\mathbf{x}$ and $\mathbf{z}_j(j = 1, ..., k)$ by

$$w_j = \frac{\exp(-||x_i - z_j||_2^2/2t^2)}{h} \quad (4)$$

where $t = \frac{1}{k} \sum_{i=1}^{k} ||x_i - z_j||_2^2, h = \sum_{i=1}^{k} \exp(-||x - z_j||_2^2/2t^2)$. Then we perform the calculation in Eq. 4 based on all elements in $\mathbf{Z}$ to obtain the reconstruct pixel

$$\mathbf{x}_i^* = \sum_{j=1}^{k} w_j \times z_j. \quad (5)$$

Finally, we replace the original pixel $\mathbf{x}_i$ with the reconstructed pixel $\mathbf{x}_i^*$ and reconstruct all pixels in each homogeneous region to obtain the denoised HSI data. To demonstrate the effectiveness of the superpixels based local reconstruction, three false colour composites (FCC) are provided in Fig. 2 corresponding to the raw, 5×5 mean filtering and superpixels based locally reconstructed Indian Pines data, respectively. The number of superpixels S is 75, the number of spatial neighbors k is 15, and the selected bands are 26,14,8. It is clear that the noisy points in raw image corresponding to some white and black points are eliminated by the proposed superpixels based local reconstruction, and the edges are enhanced as well.
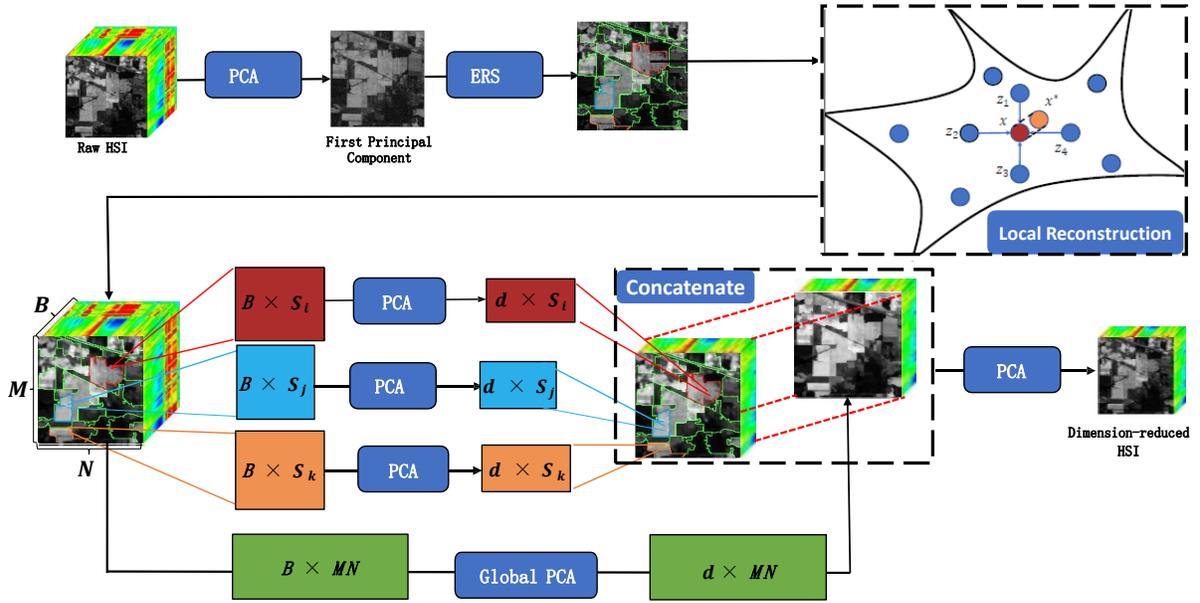
Fig. 1: Outline of the proposed S³-PCA based HSI classification framework.

(a) Raw   (b) Mean filter   (c) Local reconstruction
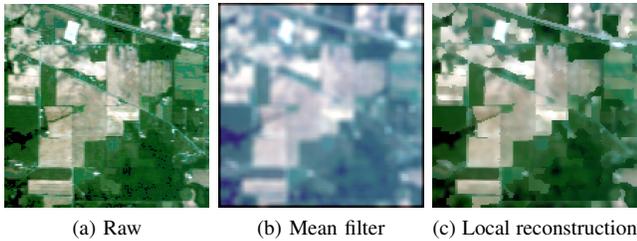
Fig. 2: The false colour composites (FCC) corresponding to the raw, 5×5 mean filtering and superpixels based locally reconstructed Indian Pines data, respectively.

B. Global-Local and Spectral-Spatial Feature Extraction

To further improve the accuracy of SuperPCA where only local spectral-spatial features are extracted and the global structures of a HSI are ignored, the global-local and spectral-spatial feature extraction model is developed by simply combining global PCA and local SuperPCA based features. Although solely using global PCA for HSIs feature extraction could be unsatisfactory, the global information of a HSI could be valuable to a certain extent, because it could provide global structures of the HSI for effective feature extraction. To take advantage of the merits of local SuperPCA and global PCA, we try to fuse them in a simple way by concatenating the features based on local SuperPCA and global PCA to form a composite global-local and spectral-spatial feature.

Mathematically, with global PCA and local SuperPCA we can get two features, which are the global PCA based global feature $H_g \in \mathbb{R}^{d \times P}$ and the SuperPCA based local feature $H_l \in \mathbb{R}^{d \times P}$, respectively. Then we simply concatenate them along the spectral dimension, leading to the novel global-local and spectral-spatial feature H

$$H = [H_g, H_l] \in \mathbb{R}^{2d \times P}. \qquad (6)$$

Since both the features are the linear transformation based on PCA, there is no need to consider that the concatenated multiple feature sets may be incompatible, resulting in the problem of highly nonlinearity in the induced feature space [31]. After concatenating the global and local features, the dimension of the new feature H could become large, thus it is necessary to perform PCA again on the new global-local and spectral-spatial feature to reduce its feature dimension. We will compare the three features, global PCA based feature, local SuperPCA based feature as well as the global-local and spectral-spatial feature in the next section.

It is also worth emphasizing that the proposed S^3-PCA is related to SuperPCA, the key ideas behind the proposed S^3-PCA consisting of the superpixels based local reconstruction for HSIs denoising and the global-local and spectral-spatial feature extraction can be integrated into SuperPCA separately, leading to the Reconstructed SuperPCA (RSuperPCA) and Concatenated SuperPCA (CSuperPCA), respectively, and then the proposed S^3-PCA can be viewed as the combination of RSuperPCA and CSuperPCA. We will analyze and compare these models in the experiments.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, two HSIs data are used to verify the effectiveness of the proposed model. *The Indian Pines data* was gathered by AVIRIS sensor in 1992 which consists of 145×145 pixels and 224 spectral reflectance bands in the wavelength range 0.4-2.5 µm. 20 bands were discarded due to the effects of noise and water absorption. The total number of labeled samples was 10249 from 16 classes. *The University of Pavia data* was acquired by ROSIS sensor in 2002 which contains 103 spectral bands with the size of 610×610 after

TABLE I: The number of samples for each class in Indian Pines and University of Pavia

Indian Pines		University of Pavia	
Class Names	Numbers	Class Names	Numbers
Alfalfa	46	Asphalt	6631
Corn-notill	1428	Bare soil	18649
Corn-mintill	830	Bitumen	2099
Corn	237	Bricks	3064
Grass-pasture	483	Gravel	1345
Grass-trees	730	Meadows	5029
Grass-pasture-mowed	28	Metal sheets	1330
Hay-windrowed	478	Shadows	3682
Oats	20	Trees	947
Soybean-notill	972		
Soybean-mintill	2455		
Soybean-clean	593		
Wheat	205		
Woods	1265		
Buildings-Grass-Trees-Drives	386		
Stone-Steel-Towers	93		
Total Number	10249	Total Number	42776

removing the noisy and water absorption bands. The geometric resolution is 1.3m with the spectral coverage from 0.43-0.86 μm. The data contain 42776 labeled pixels with 9 classes.

To verify the proposed method, eight DR methods are compared with the proposed model as well as the baseline approaches with the raw spectral features. The comparative models includes six unsupervised feature extraction methods (PCA [9], LPP [11], NPE [10],LPNPE [12],LNSPE [13], SuperPCA [20]), and two supervised feature extraction methods (NWFE [16], LFDA [15]). Similar to the previous work in [20], we use three measurements, including the overall accuracy (OA), average accuracy (AA) and Kappa to evaluate the performance of different dimension reduction algorithms for HSIs classification. To boost the accuracy of the aforementioned models, 5×5 weighted mean filter is initially adopted to preprocess the HSIs data for all the feature extraction methods except SuperPCA and the proposed S^3-PCA, then support vector machine (SVM) [4] classifier is used for HSIs classification based on the extracted features with RBF kernel and parameters grid search. All methods are tested on MATLAB R2018a and Windows 10 64bit platforms equipped with Intel Core i7-7700 CPU (3.60GHz) and 8GB memory.

In our experiments, the training and testing samples are randomly selected from the available groundtruth maps for the two data sets. The sample sizes regarding each groundtruth class in the two data sets are shown in Table I. For each land cover class in Indian Pines and University of Pavia data sets, we randomly choose $T = 5, 10, 20, 30, 40, 50, 60$ samples to be the training data, with the remaining samples to be the testing data. For some classes with a small sample size, such as grass-pasture-mowed and oats in the Indian Pines data, we only selected half of the total number of samples as the training set. To make fair comparison, all the experiments are repeated 10 times and the average results are reported.

A. Parameters Settings

There are a few parameters in the proposed S^3-PCA method to be manually preset, including the number of nearest neighbor k in local reconstruction and the number of superpixels

blocks S in ERS based superpixels segmentation. In order to analyze the influence of the parameters, we conduct the parameters sensitivity experiments for the proposed S^3-PCA. We also provide extra experiments in supplementary documentation to testify the influence of the dimensions settings for the three PCA models used in S^3-PCA. Based on the results we simply set the reduced dimensions corresponding to the three PCA models in S^3-PCA to be 30 in the following experiments.

Firstly, by referring to the experiments settings in SuperPCA, we similarly set the number of superpixels blocks S =100, 20 for Indian Pines and University of PaviaU, respectively. Then the optimal number of nearest neighbor k on the two data is investigated. Fig. 3(a) and (b) shows the OAs corresponding to S^3-PCA when the number of nearest neighbors k ranges in $\{5, 7, 9, 11, 13, 15, 17\}$ and $T = 30$. If the number of pixels in some superpixel blocks is smaller than the number of nearest neighbors k, all pixels in the superpixel block except the target pixel are used for local reconstruction. From the results in Fig. 3(a) and (b), we can conclude that with the increase of the number of nearest neighbors, the OAs in the two data tend to rise first and then decrease. If the number of nearest neighbors k is too small, it is difficult to obtain enough spatial information through reconstruction. By contrast, if the value of k is too large, it is difficult to ensure that the neighbors participating in the reconstruction are all from the same groundtruth class, which leads to a poor reconstruction results. By setting a proper value for the number of nearest neighbors, S^3-PCA and RSuperPCA always perform better than SuperPCA and LNSPE. Obviously, the proposed S^3-PCA eliminate the noise by adopting the superpixels to locally reconstruct each pixel, which is more effective than LNSPE without using spatial information. Based on the above results, we can determine that the optimal parameters $k = 15, 13$ for Indian Pines and University of Pavia data, respectively.

Furthermore, with the optimal parameter k we also investigate the influence of the number of superpixel blocks. Fig. 3(c) and (d) shows the influence of the number of superpixels blocks in terms of OAs based on SuperPCA, Concatenated-SuperPCA (CSuperPCA), Reconstructed-SuperPCA (RSuperPCA) and S^3-PCA in the two data, where the number of superpixel S is chosen from $\{1, 3, 5, 10, 20, 30, 40, 50, 75, 100, 150, 200, 300\}$ and we can get the best number of superpixels $S = 75, 30$ in Indian Pines and the University of Pavia for S^3-PCA, respectively. As can be seen from Fig. 3(c) and (d) that with the number of superpixels S increases, the OAs based on SuperPCA, CSuerpPCA, RSuperPCA and S^3-PCA show the trend of rising first and then falling. Moreover, by setting a suitable number of superpixels, the classification accuracy based on the proposed CSuperPCA, RSuperPCA and S^3-PCA are usually higher than SuperPCA. Specifically, in the Indian Pines, the accuracy improvement by superpixels based local reconstruction is more obvious than the improvement by the global-local feature concatenation. This is possibly because in the data set the samples are more clustered and show the effect of regionalization so that enough spatial information can be obtained through spatial reconstruction via superpixels. The results in the University of Pavia data display that the accuracy

(a) Indian Pines   (b) University of Pavia

(c) Indian Pines   (d) University of Pavia
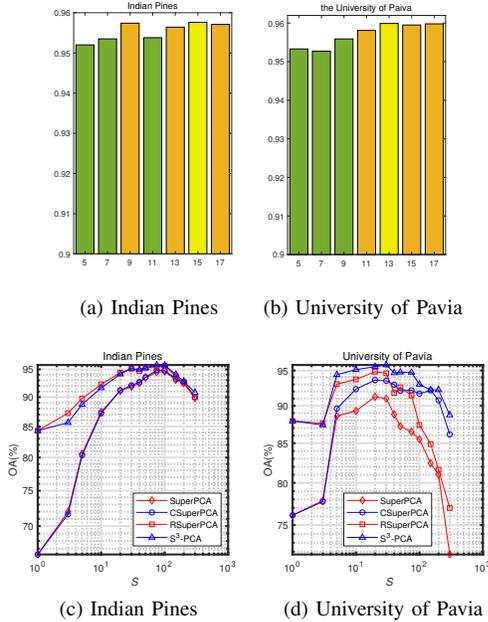
Fig. 3: OAs from the proposed method in figures (a,b) correspond to different numbers of neighbors k for the two data when the numbers of superpixels S are set to be 100,20, respectively, and OAs in figures (c,d) correspond to different settings of the numbers of superpixels S when numbers of neighbors k are optimally set to be 15,13, respectively.

improvement by the global-local feature concatenation is more obvious than another data set. This is possibly because in the University of Pavia data samples are more scattered than the Indian Pines data set. Consequently, the results of the proposed S^3-PCA illustrate that combination of the superpixels based local reconstruction for HSIs denoising plus the global-local and spectral-spatial feature extraction can be effective.

B. SuperPCA based Models Comparison

The proposed S^3-PCA is related to SuperPCA with the differences of the local reconstruction based denoising plus global-local and spectral-spatial feature extraction. Generally speaking, the two ideas can be separately integrated into SuperPCA, leading to the Reconstructed SuperPCA (RSuperPCA) and Concatenated SuperPCA (CSuperPCA), respectively. To further demonstrate the effectiveness of the two ideas behind S^3-PCA, firstly we compare the four SuperPCA based methods on the Indian Pines and University of Pavia sets when the number of superpixles are $S = 100, 20$ and the numbers of local reconstruction neighbors are $k = 15, 13$, respectively. Fig. 5 shows the ratios between the first and second eigenvalues of SuperPCA and RSuperPCA. It is acknowledged that the larger the ratio, the more representative and discriminative the features are as illustrated in SuperPCA. Since local reconstruction is carried out by using the spatial neighbors belonging to the same superpixel blocks, RSuperPCA could gain larger ratio than SuperPCA as can be seen from the blue and red horizontal lines in Fig. 5 which

TABLE II: OAs from the four methods on Indian Pines and University of Pavia data with SVM Classifier

Datasets	T.N.s/C	SuperPCA	CSuperPCA	RSuperPCA	S^3-PCA
Indian Pines	5	77.14%	77.77%	78.74%	**80.36%**
	10	85.75%	86.04%	86.57%	**87.44%**
	20	92.80%	93.28%	93.79%	**93.98%**
	30	94.61%	94.81%	95.69%	**95.81%**
University of Pavia	5	74.36%	78.92%	79.17%	**83.70%**
	10	83.39%	89.02%	89.06%	**90.88%**
	20	89.34%	92.22%	93.50%	**95.63%**
	30	91.26%	93.61%	94.82%	**96.24%**

corresponds to the average ratio of RSuperPCA and SuperPCA in all the homogeneous regions, respectively. It also means that compared to SuperPCA, RSuperPCA can obtain more discriminative features for HSIs classification, which proves our claim that superpixel based local reconstruction is capable of effectively eliminating noise in HSIs. Moreover, we conduct the comparative experiments on the two HSIs data to compare SuperPCA and CSuperPCA where the features from global PCA and local SuperPCA are simply concatenated without the aforementioned local reconstruction based HSIs denoising. For the Indian Pines and University of Pavia data sets, we randomly select samples with $T = 5$, 10, 20, and 30 from each class to be the training set, and the remaining samples as the test set. We use "T.N.s/C" to denote the numbers of training data from each class in Table II. As can be seen from Table II that CSuperPCA is more effective than SuperPCA which proves our argument that the global information from global PCA can enhance the local features from SuperPCA to a certain extent. Also, by comparing SuperPCA to RSuperPCA, we can obtain the similar conclusion in Fig. 5 because the superpixel based local reconstruction can efficiently remove the noise and improve the features in each homogeneous region. Based on the results in Table II, we can see that the proposed S^3-PCA outperforms SuperPCA, CSuperPCA and RSuperPCA, which can efficiently extract the global-local and spectral-spatial features to improve the classification accuracy. Extra experiments in supplementary documentation are also provided to compare SuperPCA and S^3-PCA by data visualization in 2D and 3D feature space, which subjectively show the superiority of the proposed S^3-PCA over SuperPCA.

C. Comparisons with State-of-the-Arts Models

In this subsection, we conduct extensive experiments to compare the proposed S^3-PCA to classical and state-of-the-art DR models. We firstly show the classification results when the number of samples for each class in the training set is 30 with the numbers of superpixles $S = 75, 30$, the numbers of spatial neighbors $k = 15, 13$ in Indian Pines and the University of Pavia respectively, and the reduced dimension $d = 30$.

The classification maps of Indian Pines and the University of Pavia data are provided in Figs. 4 and 6. From the classification results we can see that the traditional unsupervised spectrum-based DR methods, such as PCA [9], LPP [11], NPE [10] perform poorly. By contrast, better results can be obtained based on some state-of-the-art unsupervised DR algorithms such as LPNPE [12] where spatial information is extracted by
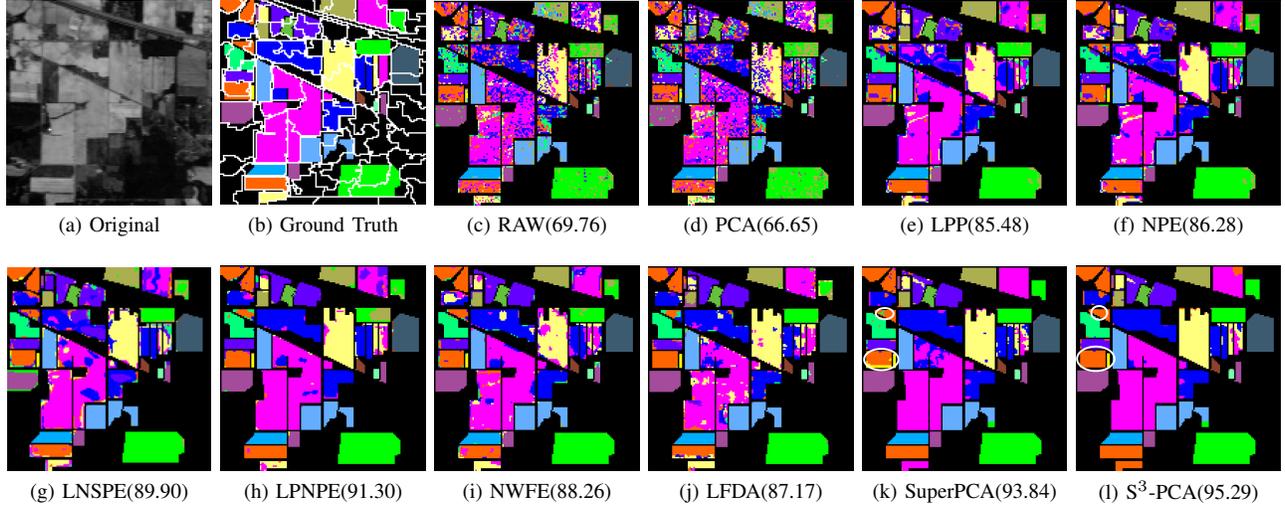
Fig. 4: Classification maps of the Indian Pines data from different models. (a) first PCA of original HSI. (b) Ground Truth. (c) Raw pixel. (d) PCA. (e) LPP. (f) NPE. (g) LNSPE. (h) LPNPE. (i) NWFE. (j) LFDA. (k) SuperPCA. (l) S^3-PCA.

TABLE III: OAs from PCA, LPP, NPE, LNSPE, LPNPE, NWFE, LFDA, SuperPCA and S^3-PCA on Indian Pines data

Indian Pines										
Class	Training Samples	PCA	LPP	NPE	LNSPE	LPNPE	NWFE	LFDA	SuperPCA	S^3-PCA
Alfalfa	23	95.65%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Corn-notill	30	50.07%	74.75%	80.47%	82.12%	92.13%	79.26%	82.98%	92.85%	90.70%
Corn-mintill	30	69.13%	89.38%	88.63%	95.75%	95.63%	90.38%	93.13%	92.50%	99.00%
Corn	30	79.23%	93.72%	93.72%	91.79%	91.79%	93.24%	94.69%	95.65%	97.58%
Grass-pasture	30	91.17%	93.16%	94.70%	92.72%	97.57%	95.58%	94.26%	99.34%	99.56%
Grass-trees	30	86.14%	97.00%	98.86%	97.86%	99.57%	99.00%	99.00%	97.14%	97.14%
Grass-pasture-mowed	14	85.71%	100.00%	100.00%	100.00%	100.00%	92.86%	100.00%	92.86%	85.71%
Hay-windrowed	30	86.61%	96.43%	98.44%	99.55%	97.77%	98.88%	99.78%	99.55%	100.00%
Oats	10	60.00%	100.00%	90.00%	90.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Soybean-notill	30	71.02%	78.56%	77.39%	83.86%	88.43%	82.80%	81.53%	83.55%	86.20%
Soybean-mintill	30	46.60%	79.05%	78.43%	84.95%	80.70%	81.03%	75.67%	91.59%	94.89%
Soybean-clean	30	68.21%	87.39%	88.28%	94.49%	97.51%	92.90%	90.05%	92.72%	95.38%
Wheat	30	97.71%	97.71%	100.00%	99.43%	100.00%	99.43%	99.43%	99.43%	99.43%
Woods	30	85.99%	92.23%	90.85%	92.96%	94.74%	95.14%	93.20%	99.84%	99.84%
Buildings-Grass-Trees-Drives	30	56.46%	93.54%	96.07%	99.16%	96.63%	96.07%	95.79%	98.60%	96.91%
Stone-Steel-Towers	30	98.41%	96.83%	100.00%	100.00%	100.00%	98.41%	98.41%	98.41%	98.41%
OA		66.65%	85.48%	86.28%	89.90%	91.30%	88.26%	87.17%	93.84%	95.29%
AA		76.76%	91.86%	92.24%	94.04%	95.78%	93.44%	93.62%	95.88%	96.30%
Kappa		62.68%	83.49%	84.43%	88.50%	90.10%	86.64%	85.42%	92.94%	94.61%
Time (s)		0.072	3.251	9.326	26.242	3.881	2.482	1.674	0.304	2.228

TABLE IV: OAs from PCA, LPP, NPE, LNSPE, LPNPE, NWFE, LFDA, SuperPCA and S^3-PCA on University of Pavia data

University of Pavia										
Class	Training Samples	PCA	LPP	NPE	LNSPE	LPNPE	NWFE	LFDA	SuperPCA	S^3-PCA
Asphalt	30	78.91%	85.43%	82.84%	84.87%	91.76%	87.11%	75.44%	79.64%	93.21%
Bare soil	30	77.24%	93.69%	90.50%	91.04%	91.67%	92.75%	87.93%	93.08%	94.65%
Bitume	30	80.04%	88.88%	88.84%	84.20%	89.51%	88.06%	72.16%	97.54%	99.81%
Bricks	30	95.02%	93.05%	91.83%	93.24%	93.28%	93.05%	95.22%	85.33%	93.47%
Gravel	30	99.92%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	96.88%	100.00%
Meadow	30	64.69%	88.86%	83.40%	89.98%	92.98%	87.00%	85.10%	94.80%	99.68%
Metal sheet	30	83.31%	93.85%	90.77%	91.92%	90.38%	92.85%	92.23%	92.38%	98.31%
Shadows	30	53.48%	84.26%	76.75%	86.34%	83.57%	82.97%	77.33%	92.55%	99.48%
Trees	30	99.45%	94.44%	95.09%	95.20%	98.26%	98.26%	91.49%	95.53%	98.15%
OA		76.75%	90.96%	87.71%	89.77%	91.51%	90.50%	85.08%	90.96%	95.95%
AA		81.34%	91.38%	88.89%	90.75%	92.38%	91.34%	86.32%	91.97%	97.42%
Kappa		70.11%	88.11%	83.89%	86.62%	88.86%	87.48%	80.55%	88.16%	94.68%
Time (s)		0.362	171.038	564.237	1679.264	194.471	2.269	1.414	0.504	154.029

TABLE V: OAs from the proposed method and nine comparative algorithms on two HSIs data with different training data

Datasets	T.N.s/C	Raw	PCA	LPP	NPE	LNSPE	LPNPE	NWFE	LFDA	SuperPCA	S^3-PCA
Indian Pines	5	44.87%±6.50%	45.47%±5.33%	51.56%±6.89%	53.26%±6.42%	55.01%±4.48%	71.73%±3.66%	60.08%±2.68%	59.12%±3.75%	77.14%±5.16%	**80.36%±4.21%**
	10	55.77%±3.27%	55.07%±2.87%	71.09%±2.60%	71.44%±3.39%	73.73%±2.94%	81.60%±3.87%	70.03%±2.52%	64.06%±2.95%	85.75%±3.06%	**87.44%±1.37%**
	20	63.81%±3.37%	62.16%±2.64%	81.88%±1.49%	82.84%±1.37%	85.28%±1.42%	89.41%±1.73%	85.08%±1.67%	82.48%±2.02%	92.80%±1.46%	**94.16%±1.61%**
	30	68.77%±1.27%	66.24%±0.58%	85.45%±2.16%	86.45%±1.91%	88.78%±1.91%	92.74%±1.56%	88.86%±1.61%	87.58%±1.14%	94.61%±0.81%	**95.81%±0.77%**
	40	71.64%±1.05%	68.90%±1.06%	87.51%±1.68%	88.54%±1.32%	90.60%±1.19%	93.37%±1.56%	90.37%±1.48%	89.68%±0.60%	95.33%±0.97%	**96.24%±0.84%**
	50	74.18%±1.20%	70.70%±1.01%	89.15%±1.20%	89.95%±1.26%	91.96%±1.15%	94.57%±1.26%	91.78%±1.32%	90.97%±0.75%	95.42%±1.02%	**96.67%±0.85%**
	60	75.24%±1.12%	71.79%±1.33%	90.16%±0.92%	90.97%±0.72%	93.01%±0.85%	95.25%±0.89%	92.52%±0.81%	91.71%±0.91%	95.72%±0.56%	**97.05%±0.85%**
University of Pavia	5	64.59%±5.07%	65.26%±5.14%	70.62%±5.79%	68.66%±6.05%	68.26%±8.56%	79.63%±3.17%	72.18%±5.46%	75.27%±2.83%	74.36%±3.65%	**83.70%±3.41%**
	10	70.22%±3.05%	67.00%±3.06%	79.12%±3.82%	77.76%±4.24%	78.71%±6.26%	86.09%±2.72%	81.31%±2.27%	78.90%±2.84%	83.39%±3.33%	**90.88%±2.58%**
	20	75.85%±2.37%	75.80%±2.31%	86.92%±1.60%	84.70%±1.91%	87.71%±2.02%	90.62%±1.89%	89.14%±1.94%	81.65%±2.14%	89.34%±1.54%	**95.63%±1.10%**
	30	76.45%±1.50%	76.13%±1.85%	89.59%±1.64%	87.33%±1.69%	90.61%±1.41%	91.70%±1.12%	90.64%±1.60%	85.56%±1.62%	91.26%±0.99%	**96.24%±1.33%**
	40	77.76%±1.26%	77.41%±1.49%	91.12%±1.45%	89.13%±1.17%	92.50%±1.08%	93.35%±0.81%	91.77%±1.19%	88.27%±1.33%	92.19%±0.81%	**96.72%±1.14%**
	50	79.12%±1.31%	78.77%±1.31%	92.38%±1.23%	91.01%±1.46%	93.62%±0.79%	94.09%±0.68%	92.66%±0.65%	89.75%±1.58%	93.25%±0.90%	**97.34%±0.77%**
	60	80.52%±1.32%	79.87%±1.16%	93.52%±0.89%	92.05%±1.07%	94.59%±0.62%	94.74%±0.80%	93.62%±0.92%	91.40%±1.13%	93.99%±1.08%	**97.84%±0.49%**

(a) Indian Pines   (b) University of Pavia
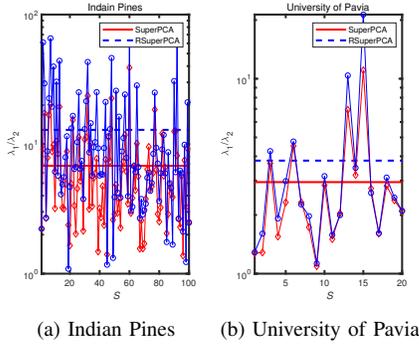
Fig. 5: Comparisons of ratio between the first and second eigenvalues (λ_1/λ_2). The red and blue plots are the ratios corresponding to all the homogeneous regions based on SuperPCA and RSuperPCA when the numbers of superpixels are set to the optimal values $S = 100, 20$ for Indian Pines and the University of Pavia, respectively. The red and blue horizontal line represents the average ratio of all the homogeneous regions based on SuperPCA and RSuperPCA. For convenience, the logarithmic function is used to scale the values of ratios.

local spatial-spectral scattering and LNSPE [13] where local reconstruction based on k nearest neighbors is adopted for feature extraction by combining the dual structure preserving model with neighborhood scatter extraction model. Also, the supervised DR methods such as LFDA [15] and NWFE [16], provides relatively high classification accuracy because they make use of label information to obtain more discriminative features. SuperPCA efficiently considers the spatial information of HSIs through superpixel segmentation, which performs better than the aforementioned algorithms, and the proposed S^3-PCA method achieves the best results.

To compare and analyze the superpixels based models, the superpixel segmentation results are also provided in Fig. 4(b) and Fig. 6(b). As can be seen from Fig. 4, compared to SuperPCA the proposed S^3-PCA can significantly improve the accuracy for the large regions in Indian Pines data (e.g., the Soybean-notill located in the middle yellow area and the Soybean-mintill located in the middle pink area) possibly because of the superpixels based local reconstruction which could eliminate noise. More interestingly, when the ERS based superpixel segmentation provides poor results where different groundtruth objects are clustered into the same superpixel

such as the area of small circle in Fig. 4(k,l) and the area of big circle in Fig. 6(k,l), the proposed S^3-PCA could still provide satisfactory classification results compared to SuperPCA, possibly because the supplementary features based on global PCA could be discriminative. This phenomenon is very prominent in the University of Pavia data where different groundtruth objects are typically mixed together, as the optimal number of superpixels is set to 30. Furthermore, S^3-PCA shows promising accuracy in the small regions with limited samples, such as the area of big circle in Fig. 4(k,l) and the area of small circle in Fig. 6(k,l), which demonstrates the superiority of the proposed model. Table III-IV show the detail of classification results based on different DR models and SVM when the training data of each class $T = 30$. Obviously, the proposed method S^3-PCA achieved the best performance in terms of OA, AA and Kappa in the two data sets. The comparison of training times are also provided in the tables. It's worth noting that all the data including the unlabeled samples are firstly processed by the unsupervised DR models followed by the train-test splitting, which correspond to 21025 and 207400 samples for the two HSIs data. Consequently, the proposed S^3-PCA is more time-consuming than SuperPCA, because the superpiexls based local construction takes time.

To further verify the proposed S^3-PCA, we randomly choose $T = 5, 10, 20, 30, 40, 50, 60$ samples from each class to be the training data with the remaining to be the testing data in Indian Pines and University of Pavia data to show the changes of OAs when the number of training data increases. Table V shows the results from different approaches, where S^3-PCA achieves the highest OA than other comparative models especially in the case of limited training data. For example, when $T = 5$, S^3-PCA can achieve 3.22%, 9.34% improvements compared to SuperPCA in the two HSIs data.

V. CONCLUSION

In this paper, we propose S^3-PCA which uses the super-pixel based local reconstruction for HSIs denoising plus the global-local and spectral-spatial feature extraction. It turns out that the new method can extract effective features for HSIs classification especially when dealing with pixels in some small homogeneous regions with noise and in some large homogeneous regions with mixed groundtruth objects. With the superpixel segmentation technique local spatial structures of HSIs can be extracted, then superpixel based spatial nearest neighbor reconstruction for each pixel in every homogeneous

(a) Original (b) Ground Truth (c) RAW(76.18) (d) PCA(76.75) (e) LPP(90.96) (f) NPE(87.71)

(g) LNSPE(89.77) (h) LPNPE(91.51) (i) NWFE(90.50) (j) LFDA(85.08) (k) SuperPCA(90.96) (l) S^3-PCA(95.95)
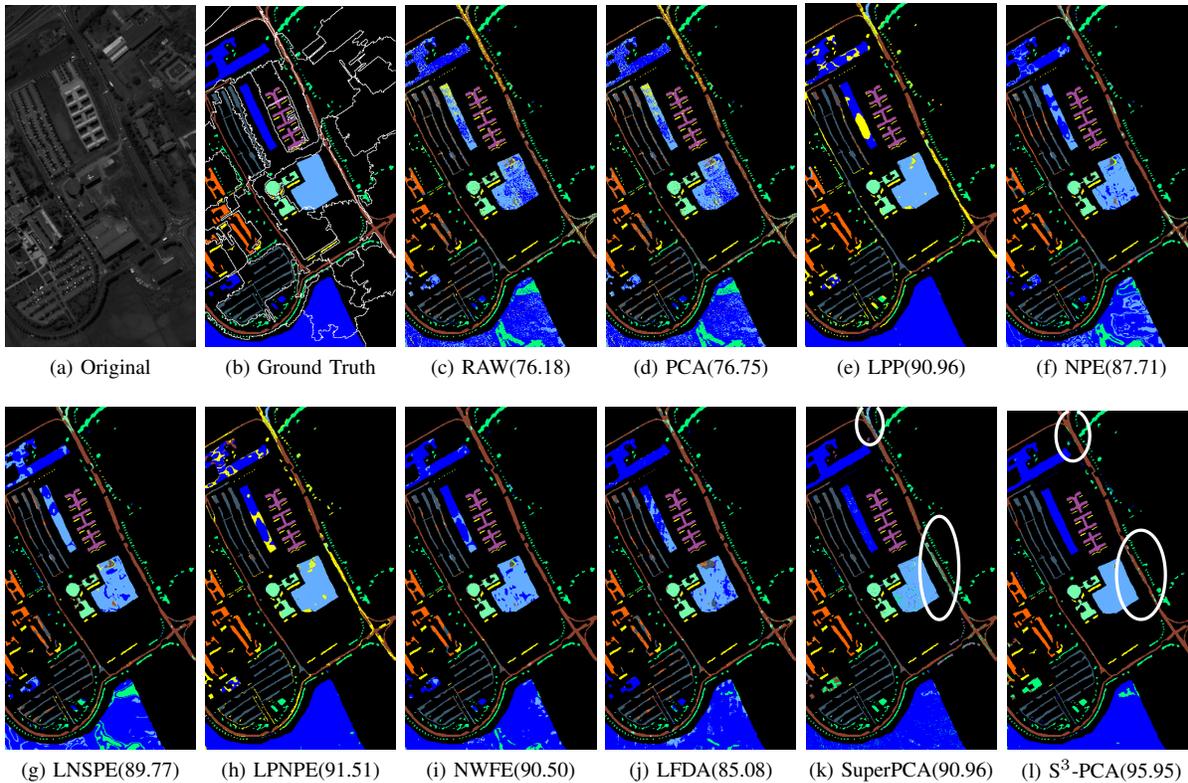
Fig. 6: Classification maps of the University of Pavia data from different models. (a) first PCA of original HSI. (b) Ground Truth. (c) Raw pixel. (d) PCA. (e) LPP. (f) NPE. (g) LNSPE. (h) LPNPE. (i) NWFE. (j) LFDA. (k) SuperPCA. (l) S^3-PCA.

region is introduced to filter HSIs with noise elimination. Furthermore, given the fact that the low-dimensional features based on typical PCA could also provide global information of HSIs, we fuse the global features from global PCA and local features from SuperPCA, leading to the global-local and spectral-spatial feature. Experiments on two HSIs data have demonstrated that the proposed S^3-PCA outperforms the classical and state-of-the-art feature extraction methods.

For the future work, how to automatically determine the optimal number of superpixels for superpixels segmentation technique should be addressed. A possible solution could be to initially set a large number of superpixels and then to adaptively combine the similar superpixel blocks. Also, the superpixels based local reconstruction has proven to be an effective filter to denoise HSIs, which can be integrated into other classification methods such as SVM and deep learning models as an effective data preprocessing approach.

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. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4435–4447, 2018.
[3] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.
[4] 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.
[5] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, "Noise reduction in hyperspectral imagery: Overview and application," *Remote Sensing*, vol. 10, no. 3, p. 482, 2018.
[6] Y. Cai, X. Liu, and Z. Cai, "Bs-nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 1969–1984, 2020.
[7] X. Jiang, X. Song, Y. Zhang, J. Jiang, J. Gao, and Z. Cai, "Laplacian regularized spatial-aware collaborative graph for discriminant analysis of hyperspectral imagery," *Remote Sensing*, vol. 11, no. 1, p. 29, 2018.
[8] Y. Cai, Z. Zhang, Z. Cai, X. Liu, X. Jiang, and Q. Yan, "Graph convolutional subspace clustering: A robust subspace clustering framework for hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–12, 2020.
[9] M. D. Farrell and R. M. Mersereau, "On the impact of pca dimension reduction for hyperspectral detection of difficult targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 2, pp. 192–195, 2005.
[10] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, vol. 2. IEEE, 2005, pp. 1208–1213.
[11] X. He and P. Niyogi, "Locality preserving projections," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, 2004, pp. 153–160.
[12] Y. Zhou, J. Peng, and C. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 1082–1095, 2014.
[13] G. Shi, H. Huang, and L. Wang, "Unsupervised dimensionality reduction for hyperspectral imagery via local geometric structure feature learning," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2019.
[14] 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.
[15] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image anal-

ysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1185–1198, 2011.

[16] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096–1105, 2004.

[17] J.-H. Rick Chang and Y.-C. Frank Wang, "Propagated image filtering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 10–18.

[18] M. Wang, J. Yu, L. Niu, and W. Sun, "Feature extraction for hyperspectral images using low-rank representation with neighborhood preserving regularization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 6, pp. 836–840, 2017.

[19] 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, no. 8, p. 790, 2017.

[20] 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.

[21] L. He, X. Chen, J. Li, and X. Xie, "Multiscale superpixelwise locality preserving projection for hyperspectral image classification," *Applied Sciences*, vol. 9, no. 10, p. 2161, 2019.

[22] L. Zhang, H. Su, and J. Shen, "Hyperspectral dimensionality reduction based on multiscale superpixelwise kernel principal component analysis," *Remote Sensing*, vol. 11, no. 10, p. 1219, 2019.

[23] H. Xu, H. Zhang, W. He, and L. Zhang, "Superpixel-based spatial-spectral dimension reduction for hyperspectral imagery classification," *Neurocomputing*, vol. 360, pp. 138–150, 2019.

[24] Y. Gu, T. Liu, and J. Li, "Superpixel tensor model for spatial-spectral classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2019.

[25] Z. Chen, J. Jiang, C. Zhou, S. Fu, and Z. Cai, "Superbf: Superpixel-based bilateral filtering algorithm and its application in feature extraction of hyperspectral images," *IEEE Access*, vol. 7, pp. 147 796–147 807, 2019.

[26] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[27] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 2097–2104.

[28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[29] C. Tao, H. Pan, Y. Li, and Z. Zou, "Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 12, pp. 2438–2442, 2015.

[30] C. Shi and C. Pun, "Multiscale superpixel-based hyperspectral image classification using recurrent neural networks with stacked autoencoders," *IEEE Transactions on Multimedia*, vol. 22, no. 2, pp. 487–501, 2020.

[31] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3681–3693, 2015.

Xinwei Jiang Xinwei Jiang received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2012. He is an Associate Professor with the China University of Geosciences, Wuhan. His research interests include nonparametric statistical models, dimensionality reduction and hyperspectral image processing.
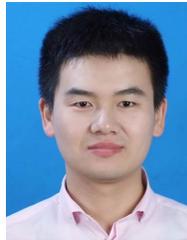
Junjun Jiang Junjun Jiang (Member, IEEE) received the B.S. degree from the Department of Mathematics, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014. From 2015 to 2018, he was an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan. From 2016 to 2018, he was a Project Researcher with the National Institute of Informatics, Tokyo, Japan. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His interests mainly focus on computer vision, image processing, and hyperspectral image analysis.

Yongshan Zhang Yongshan Zhang received her bachelor and Ph.D degree in Computer Science in 2014 and 2019 from China University of Geosciences, Wuhan, China, where she is an Associate Professor. She was a visiting student in the Department of Computer Science, University of Illinois at Chicago. Her research focuses on data mining and machine learning.

Xiaobo Liu Xiaobo Liu (Member, IEEE) received the M.S. degree in computer science and the Ph.d. degree in geosciences information engineering from the China University of Geosciences, Wuhan, China, in 2008 and 2012, respectively. He is an Associate Professor with the School of Automation, China University of Geosciences. His research interests include machine learning, evolutionary computation, and hyperspectral remote sensing image processing.

Xin Zhang Xin Zhang received the B.Eng. degree in Computer Science from the China University of Geosciences, Wuhan, China, in 2018, where he is pursuing the master degree with the School of Computer Science. His research interests include machine learning and hyperspectral image processing.

Zhihua Cai Zhihua Cai received the B.S. degree from Wuhan University, Wuhan, China, in 1986, the M.Sc. degree from the Beijing University of Technology, Beijing, China, in 1992, and the Ph.D. degree from the China University of Geosciences, Wuhan, in 2003. He is a Faculty Member with the School of Computer Science, China University of Geosciences. He has published over 50 research articles in journals and conferences. His main research areas include data mining, machine learning, evolutionary computation, and their applications.