

# SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery

Junjun Jiang, *Member, IEEE*, Jiayi Ma, *Member, IEEE*, Chen Chen, *Member, IEEE*, Zhongyuan Wang, *Member, IEEE*, Zhihua, Cai, and Lizhe Wang, *Senior Member, IEEE*

**Abstract**—As an unsupervised dimensionality reduction method, principal component analysis (PCA) has been widely considered as an efficient and effective preprocessing step for hyperspectral image (HSI) processing and analysis tasks. It takes each band as a whole and globally extracts the most representative bands. However, different homogeneous regions correspond to different objects, whose spectral features are diverse. It is obviously inappropriate to carry out dimensionality reduction through a unified projection for an entire HSI. In this paper, a simple but very effective superpixelwise PCA approach, called SuperPCA, is proposed to learn the intrinsic low-dimensional features of HSIs. In contrast to classical PCA models, SuperPCA has four main properties. (1) Unlike the traditional PCA method based on a whole image, SuperPCA takes into account the diversity in different homogeneous regions, that is, different regions should have different projections. (2) Most of the conventional feature extraction models cannot directly use the spatial information of HSIs, while SuperPCA is able to incorporate the spatial context information into the unsupervised dimensionality reduction by superpixel segmentation. (3) Since the regions obtained by superpixel segmentation have homogeneity, SuperPCA can extract potential low-dimensional features even under noise. (4) Although SuperPCA is an unsupervised method, it can achieve competitive performance when compared with supervised approaches. The resulting features are discriminative, compact, and noise resistant, leading to improved HSI classification performance. Experiments on three public datasets demonstrate that the SuperPCA model significantly outperforms the conventional PCA based dimensionality reduction baselines for HSI classification, and some state-of-the-art feature extraction approaches. The Matlab source code is available at <https://github.com/junjun-jiang/SuperPCA>.

**Index Terms**—Hyperspectral image classification, unsupervised dimensionality reduction, feature extraction, principal component analysis (PCA), superpixel segmentation.

## I. INTRODUCTION

J. Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 140001, China. E-mail: junjun0595@163.com.

J. Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China. E-mail: jyama2010@gmail.com.

C. Chen is with the Center for Research in Computer Vision, University of Central Florida (UCF), Orlando, FL 32816-2365 USA. E-Mail: chenchen870713@gmail.com.

Z. Wang is with the School of Computer, Wuhan University, Wuhan 430072, China. E-mail: wzy\_hope@163.com.

Z. Cai and L. Wang are with the School of Computer Science, China University of Geosciences, Wuhan 430074, China. E-mail: zhcai@cug.edu.cn; lizhe.wang@gmail.com.

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

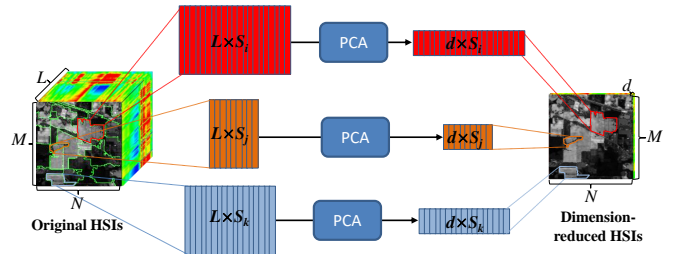


Fig. 1. Schematic of SuperPCA based dimensionality reduction for HSIs.

**H**YPERSPECTRAL image (HSI) acquired by spaceborne or airborne sensors, such as AVIRIS, HyMap, HYDICE, and Hyperion, typically record material's hundreds of thousands of spectral wavelengths for each pixel in the image, which has opened new perspectives in many applications in remote sensing [1], [2], [3]. Since the subtle differences in ground covers can be captured by different spectral signatures, hyperspectral imagery is a well-suited technology for discriminating materials of interest. Although the rich spectral signatures can provide useful information for data analysis, the high dimensionality of HSI data presents some new challenges: (i) increasing the burden of data transmission and storage; (ii) leading to the curse of dimensionality problem [4] which will reduce the generalization capability of classifiers and deteriorate the classification performance, especially when the available labeled samples are limited. Owing to (i) the dense sampling of spectral wavelengths and (ii) the spectral reflectance of most materials changes only gradually over certain spectral bands, many contiguous bands are highly correlated and not all features (or spectral bands) are expected to contribute useful information for the data classification/analysis task at hand. As one of the typical method to alleviate this problem, dimensionality reduction is widely used as a preprocessing step to remove the highly correlated and redundant measurements in the original high-dimensional HSI spectral space and preserve essential information in a low-dimensional subspace. It has attracted increasing attentions in recent years.

Generally speaking, dimensionality reduction of HSI data can be divided into two categories: feature selection [5], [6], [7], [8] and feature extraction [9], [10], [11]. The former tends to select a small subset of the most representative bands from the original bands, whereas the latter aims to find an

optimal transformation matrix to project the original high-dimensional spectral features into a low-dimensional subspace. Feature selection can only select existing bands from HSIs, whereas feature extraction can use entire bands to generate more discriminative features. In [12], a joint feature extraction and feature extraction method for HSI representation and classification has been developed. In this paper we mainly focus on employing feature extraction to reduce the feature dimensions of HSIs. Based on whether or not the label information is used, the feature extraction can be classified into unsupervised approaches and supervised approaches.

One of the most widely applied unsupervised dimensionality reduction techniques in HSI analysis is the principal component analysis (PCA) [13] and its variants [14], [15], [16], [17], [10]. Without any label information, PCA tends to find orthogonal transformations to maximize the total variance of the projected data. Different from preserving the largest data variance as in PCA, independent component analysis (ICA) [18] tries to find the independent components by maximizing the statistical independence of the estimated components. Recently, some nonlinear methods based on manifold learning [19], [20] have been used to compute the essential embedded low-dimensional space of observed high dimensional data [21], *e.g.*, locally linear embedding (LLE) [22], [23], neighborhood preserving embedding (NPE) [24], locality preserving projection (LPP) [25], and most recently proposed local pixel neighborhood preserving embedding (LPNPE) [26]. Other efficient unsupervised feature extraction and learning methods also include intrinsic representation [27], sub-feature learning [28], and latent subclass learning [29]. Supervised dimensionality reduction algorithms leverage the supervised information, *i.e.*, the labels, to learn the dimensionality reduced feature space. The most representative works include Fisher's linear discriminant analysis (LDA) [14] and Local Fisher discriminant analysis (LFDA) [30].

Most of the above feature extraction methods use only spectral signature of each pixel and the dimensionality reduction models cannot directly use spatial information of HSIs, which has been proven to be very effective to improve the HSI representation and classification accuracy [1], [31], [32], [33]. In [26], the spatial information is applied to spatial filtering (as a preprocessing) as well as modeling the spatial neighboring pixel correlations. Wen *et al.* proposed to incorporate the spatial information, *e.g.*, texture or morphological features, into the framework of orthogonal nonnegative matrix factorization [34]. The approach of [35] presents a novel spectral-spatial feature based similarity measurement which can be incorporated into existing dimensionality reduction methods including linear or nonlinear techniques. In [36], [37], spatial information is used to regularize the spectral representation.

#### A. Motivation and Contributions

Conventional methods usually learn a unified projection for HSI feature extraction [30], [38], [39], [40]. However, different regions in an HSI may correspond to different objects, whose spectral features are diverse. Therefore, a reasonable way is to learn different projection matrices for different regions. Image

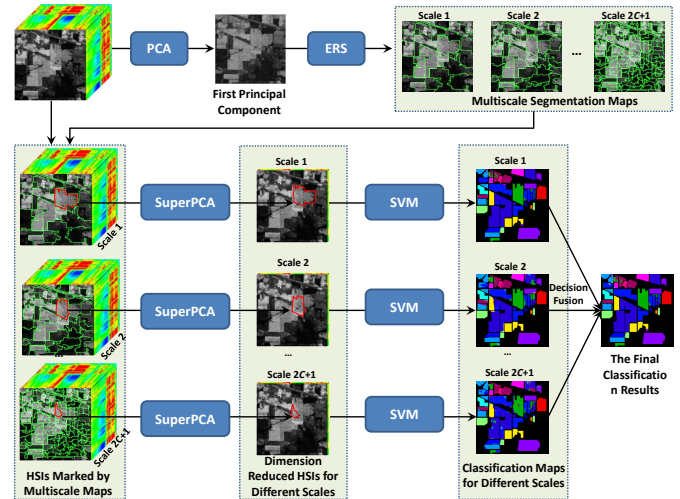


Fig. 2. Outline of the proposed multiscale SuperPCA based HSI classification framework.

segmentation can be seen as an exhaustive partitioning of the observed image into many different regions, and each of which is considered to be homogeneous [41]. These regions form a segmentation map that can be used as spatial structures for the spectral-spatial classification.

In this paper, we advocate a simple yet very effective unsupervised feature extraction method based on superpixelwise PCA, which is denoted as SuperPCA. It can learn the intrinsic low-dimensional features of different regions of the HSI data by performing PCA on each homogeneous region obtained by superpixel segmentation, as shown in Fig. 1. An HSI is firstly divided into many homogeneous regions via superpixel segmentation, which are denoted by matrices whose columns are the spectral vectors of pixels. PCA is applied to these high-dimension matrices to obtain the dimensionality reduced ones. Finally, we rearrange and combine all these low dimensional matrices to form the dimensionality reduced HSIs.

In an attempt to make full use of the spatial information contained in the HSI cube, we further develop a multiscale segmentation based SuperPCA model, namely MSuperPCA, which can effectively integrate multiscale spatial information to obtain the optimal classification result by decision fusion. Fig. 2 demonstrates the schematic of our proposed multiscale SuperPCA method. We first apply entropy rate superpixel (ESR) to obtain multiscale superpixel segmentations (by setting different superpixel numbers) based on the first principal component of the input HSIs. Then, for each scale, the proposed SuperPCA based unsupervised dimensionality reduction method is used to obtain the dimensionality reduced HSIs. Based on the predictions of different scales through support vector machine (SVM) classifier, we generate the final classification result via the majority voting decision fusion strategy.

To the best of our knowledge, this is the first time that a superpixelwise model is adopted for unsupervised dimensionality reduction and classification in hyperspectral imagery. Extensive experimental results demonstrate that, our method is not only simple and intuitive, but also achieves the most

competitive HSI classification results as compared with the state-of-the-art dimensionality reduction based methods, including some recently proposed supervised feature extraction techniques. When the label information is limited (a small number of labeled training samples, e.g. 5 samples per class), our proposed SuperPCA and MSuperPCA methods obtain even better classification accuracies than the state-of-the-art supervised feature extraction techniques.

### B. Organization of This Paper

The remainder of the paper is organized as follows. Section II firstly reviews and introduces the ESR superpixel segmentation algorithm. Section III introduces notations and then explains the details of the proposed HSI classification approach based on SuperPCA and the multiscale extension of SuperPCA model. And then, we also give some analysis of the proposed SuperPCA algorithm. Section IV presents the experimental results and analysis. Finally, the concluding remarks are stated in Section V.

## II. ENTROPY RATE SUPER-PIXEL SEGMENTATION (ERS)

For a superpixel segmentation algorithm, it should have the following characteristics. Firstly, superpixels should adhere well to the object boundaries. Secondly, as a preprocessing process, superpixel segmentation should be of low computational complexity itself.

Recently, graph structure based segmentation approaches are widely used in superpixel segmentation [42] and applications [43]. A typical superpixel segmentation technique is the eigen-based solution to the normalized cuts (NCuts) [44]. However, it needs to construct a very large graph ( $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ ) whose vertices ( $\mathbf{V}$ ) are the pixels in the image to be segmented, the edge set ( $\mathbf{E}$ ) consists of the pairwise similarities by the weight function  $s : \mathbf{E} \rightarrow \mathbb{R}^+ \cup \{0\}$ . Therefore, performing eigenvalue decomposition on such a large similarity matrix is very time consuming, which will take several minutes for segmenting an image of moderate size, e.g., around  $500 \times 300$  pixels. TurboPixel [45] is an efficient alternative to achieve a similar regularity. However, it sacrifices fine image details and results in a low boundary recall. In [46], an ERS segmentation approach is proposed, and the graph is partitioned into a connected subgraph by choosing a subset of edges  $\mathbf{A} \subseteq \mathbf{E}$  such that the resulting graph  $\mathcal{G}' = (\mathbf{V}, \mathbf{A})$  consists of smaller connected components/subgraphs. In the objective function of ERS, it incorporates an entropy rate term  $H(\mathbf{A})$  and a balancing term  $B(\mathbf{A})$  to optimize the superpixel segmentation:

$$\mathbf{A}^* = \underset{\mathbf{A}}{\operatorname{argmax}} \operatorname{Tr} \{H(\mathbf{A}) + \alpha B(\mathbf{A})\}, \quad s.t. \mathbf{A} \subseteq \mathbf{E}. \quad (1)$$

Here, the first term favors the formation of homogeneous and compact clusters, while the second term can be used to encourage the cluster with similar sizes.  $\alpha$  is used to balance the contributions of the entropy rate term  $H(\mathbf{A})$  and the balancing term  $B(\mathbf{A})$ . As described in [47], a greedy algorithm effectively solves the optimization problem in (1). This method is highly efficient, which only takes about 2.5 seconds to segment an image of size  $500 \times 300$  pixels.

## III. SUPERPIXELWISE PRINCIPAL COMPONENT ANALYSIS (SUPERPCA)

An HSI cube  $\mathbf{X} \in \mathbb{R}^{M \times N \times L}$  is made up with hundreds of nearly contiguous spectral bands, with high (5-10 nm) spectral resolution, from the visible to infrared spectrum for each image pixel. Here,  $M$ ,  $N$  and  $L$  are the number of image rows, columns and sampled wavelengths, respectively. We can reshape the 3D cube to a 2D matrix,  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_P] \in \mathbb{R}^{L \times P}$  ( $P = MN$ ), in which each column represents one pixel vector that reflects the energy spectrum of the materials within the spatial area covered by the pixel.

Denote  $\mathbf{x}_i \in \mathbb{R}^L$  ( $1 \leq i \leq P$ ) the  $i$ -th pixel vector of the observed HSI cube  $\mathbf{X} \in \mathbb{R}^{L \times P}$ ,

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iL}]^T. \quad (2)$$

PCA performs the dimensionality reduction by computing the low-dimensional representation that maximizes data variance in the dimensionality reduced space. Specifically, it finds a linear mapping from the original  $L$ -dimensional space  $\mathbf{X} \in \mathbb{R}^{L \times P}$  to a low  $d$ -dimensional space  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_P] \in \mathbb{R}^{d \times P}$ ,  $d < L$ . Without loss of generality, we denote the transformation matrix by  $\mathbf{W}$ . That is,  $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$ . Mathematically, it aims at finding the linear transformation matrix by solving the following objective function,

$$\mathbf{W}^* = \underset{\mathbf{W}^T \mathbf{W} = \mathbf{I}}{\operatorname{argmax}} \operatorname{Tr} (\mathbf{W}^T \operatorname{Cov}(\mathbf{X}) \mathbf{W}), \quad (3)$$

where  $\operatorname{Cov}(\mathbf{X})$  stands for the covariance matrix of the data set  $\mathbf{X}$ , and  $\operatorname{Tr}(\mathbf{X})$  denotes the trace of an  $n$ -by- $n$  square matrix  $\mathbf{X}$ .

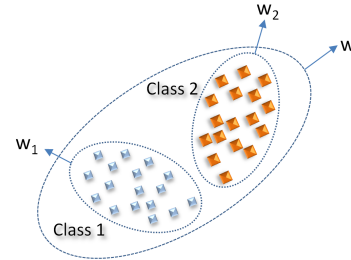


Fig. 3. The principal projection directions of global PCA and class specific PCA.

Owing to its simplicity, effectiveness, and robustness to noise, PCA has been widely used as a preprocessing step of many HSI based applications. However, in an HSI, there are many homogeneous regions. Within each region, pixels are more likely to be the same class [48], [49], [50], [51]. The global PCA approach considers the entire data space (composed of all the pixel vectors of the HSI cube), and tries to find the best transformation vector for this space. It may ignore the differences of homogeneous regions. As illustrated by a toy example (Fig. 3), we suppose that the data space is formed by class 1 (marked with blue squares) and class 2 (marked with orange squares), which could possibly represent distributions of samples from two different homogeneous regions of HSIs. We can obviously see that the transformation vectors  $\mathbf{w}_1$  and  $\mathbf{w}_2$  for class 1 and class 2 are significantly different, and they are also different from the transformation vector  $\mathbf{w}$  generated

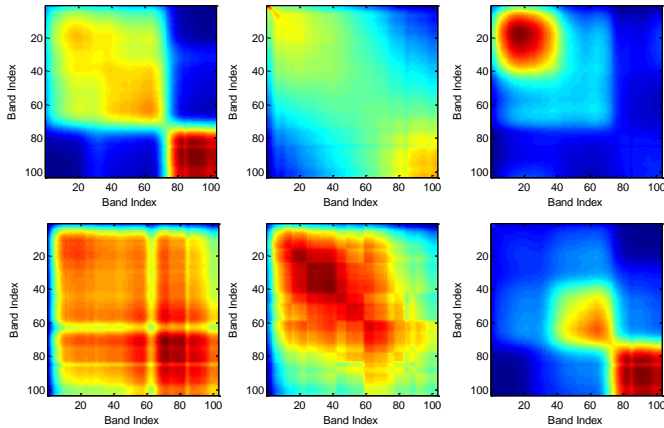


Fig. 4. Visualization of the correlation matrices of spectral bands of the entire University of Pavia image (the top left subfigure) and different homogeneous regions (the rest subfigures).

for the entire data space. As shown in Fig. 4, we plot the correlation matrices of spectral bands of the entire University of Pavia image as well as some typical homogeneous regions. From this figure, we can learn that the correlation matrices are variant. Therefore, different regions will have varying transformation vectors (see Eq. (3)).

#### A. Generation of Homogeneous Regions

Inspired by the above observation, in this paper we propose a divide-and-conquer strategy to perform unsupervised feature extraction based on PCA for each homogeneous region. By extracting the same number of principal components (PCs) for each homogeneous region, we can combine them to form the dimensionality reduced HSIs (Fig. 2). In the following, we will introduce the construction of homogeneous regions using superpixel segmentation, which can exhaustively partition the image into many homogeneous regions.

As in many superpixel segmentation based hyperspectral image classification and restoration methods [48], [49], [52], [53], we adopt ERS due to its promising performance in both efficiency and efficacy. Other state-of-the-art methods such as simple linear iterative clustering (SLIC) [54] can also be used to replace the ERS. Specially, we first obtain the first principal component of HSIs,  $I_f$ , capturing the major information of HSIs. This further reduces the computational cost for superpixel segmentation. And then, we perform ESR on  $I_f$  to obtain the superpixel segmentation,

$$I_f = \bigcup_k^S \mathcal{X}_k, \text{ s.t. } \mathcal{X}_k \cap \mathcal{X}_g = \emptyset, (k \neq g), \quad (4)$$

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

#### B. Multiscale Extension of SuperPCA

By segmenting the HSIs to superpixels, it will be beneficial to exploit rich spatial information about the land surface [52], [32]. However, how to select an optimal value for the number

of superpixels is a very challenging problem in actual applications [46]. When the superpixels are too large (by setting a small superpixel number), the resultant under-segmentation can lead to ambiguity-labeled boundary superpixels that require further segmentation. When superpixels are too small (by setting a large superpixel number), the features computed from the over-segmented regions may become less distinctive, making it more difficult to infer correct labels. In addition, as reported in [55], there is no single region size that would adequately characterize the spatial information of HSIs. Inspired by the classifier and decision fusion techniques [56], [57], [58], in this paper we propose the multiscale segmentation strategy to enhance the performance of single scale SuperPCA based method, thus alleviating above-mentioned problem. More specifically, the principal component image  $I_f$  (the first principal component of HSIs) is segmented into  $2C+1$  scales. The number of superpixel of the  $c$ -th scale is  $S_c$ ,

$$S_c = (\sqrt{2})^c S_f, \quad c = 0, \pm 1, \pm 2, \dots, \pm C, \quad (5)$$

where  $S_f$  is the fundamental superpixel number and is set empirically. Since the value of  $S_c$  may not be an integer number in  $\{1, 2, \dots, P\}$ , we reset it as  $S_c = \min(\max(1; \text{round}(S_c)); P)$ . Here,  $P$  is the number of total pixels in the HSIs.

By taking advantage of the multiscale superpixels, the decision fusion strategy can boost the classification accuracy, especially in conflicting situations. Specifically, we fuse the label information of each test pixel predicted by different multiscale superpixels. That is, given that the fundamental image  $I_f$  is segmented to  $2C+1$  scales, and there will be  $2C+1$  different classification results for an HSI. Then, we can aggregate the results through an effective decision fusion strategy. In this paper, we leverage the majority voting (MV) based decision fusion strategy due to its insensitivity to inaccurate estimates of posterior probabilities:

$$l = \arg \max_{i \in \{1, 2, \dots, G\}} N(i), \quad N(i) = \sum_{j=1}^{2C+1} \alpha_j I(l_j = i), \quad (6)$$

where  $l$  is the class label from one of the  $G$  possible classes for the test pixel,  $j$  is the classifier index,  $N(i)$  represents the number of times that class  $i$  is predicted in the bank of classifiers, and  $I$  denotes the indicator function. In Eq. (6),  $\alpha_j$  denotes the voting strength of the  $j$ -th classifier. One possible way of performing this adaptive voting mechanism is to weigh a classifier's vote based on its confidence score, which can be learned from training data. In this paper, we directly use the equal voting strength,  $\alpha_j = \frac{1}{2C+1}$  ( $j = 1, 2, \dots, 2C+1$ ).

Fig. 2 shows the framework of the proposed multiscale SuperPCA method for HSI classification. We firstly obtain the first principal component of the input HSIs. Then, it is segmented to multiple scales based on the ESR algorithm [46] with different superpixel numbers. For each scale, we perform PCA dimensionality reduction on each homogeneous region and combine all regions to form the dimension-reduced HSIs. Lastly, we apply SVM classification to each dimension-reduced HSIs and fuse the classification results by majority voting to predict the final labels for testing samples.

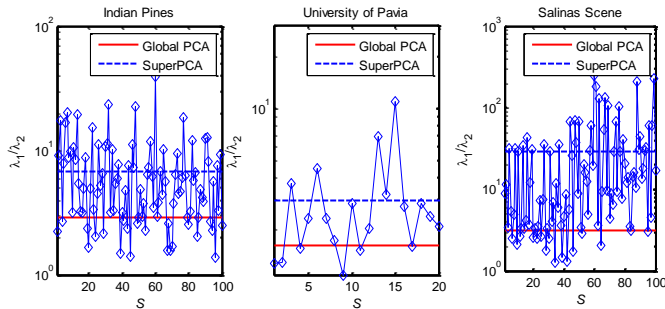


Fig. 5. The ratio between the first and second eigenvalues ( $\lambda_1/\lambda_2$ ). The red lines are the ratios of the global PCA method, while the blue plots are the ratios of all the homogeneous regions based on the proposed SuperPCA method when the number of superpixel is set to the optimal value,  $S = 100$ ,  $S = 20$ , and  $S = 100$ , for Indian Pines, University of Pavia, and Salinas Scene, respectively. The blue horizontal line represents the average ratio of all the homogeneous regions. For the convenience of observation, we use a logarithmic scale for the values of ratios.

### C. Analysis of the Proposed SuperPCA

*Remark 1.* Through superpixel segmentation, we can obtain different homogeneous regions, in which pixels are more likely to fall in the same class [48], [49], [50]. By dividing the global HSIs to some small regions, it becomes easier to find the intrinsic projection directions. Fig. 5 shows the ratios between the first and second eigenvalues of PCA (global based) and the proposed SuperPCA on Indian Pines, University of Pavia, and Salinas Scene HSI datasets (for more detailed information about the datasets and the parameter setting of the number of superpixel  $S$ , please refer to the experimental section). Obviously, the larger the ratio, the more representative and discriminant the primary projected features are. By segmenting the HSIs to different homogeneous regions, SuperPCA gains larger ratio than conventional global PCA method (see the blue and red horizontal lines). It is worth noting that larger  $S$  results in smaller homogeneous regions, and each of which has a better consistency. However, it does not necessarily lead to better classification performance. This is because, when the homogeneous region (superpixel) is too small, there will be few data samples in each superpixel, which may cause instability for PCA. From the experimental analysis, it is clear that the divide-and-conquer strategy of unsupervised feature extraction based on SuperPCA can significantly increase the eccentricity in the direction of the first eigenvector. This further corroborates our claim that a homogeneous region based PCA will be more effective in preserving the essential data information in a low dimensional space.

*Remark 2.* There are currently a number of region-based PCA methods for feature extraction or other related applications. For example, in region-based PCA face recognition [59], [60], [61], they divide the whole face image into small patches, and then use PCA to extract the local features that cannot be captured by traditional global face based PCA algorithm; in region-based PCA image denoising [62], they first divide the whole face image into small patches, and then stack similar noisy patches and apply PCA to exploit these consistency structure among similar patches (thus removing the noise). However, when we directly apply the regular

TABLE I  
CLASSIFICATION RESULTS (IN TERMS OF OA) OF THREE  
DIVIDE-AND-CONQUER STRATEGIES ON THE INDIAN PINES DATASET  
USING SVM AND NN CLASSIFIERS.

Noise	T.N.s/C	Global PCA	ClusterPCA	SquarePCA	SuperPCA
$\sigma = 0$	5	46.37, 46.94	46.37, 46.94	67.32, 65.64	77.34, 75.85
	10	55.72, 52.06	55.72, 51.83	77.59, 76.89	85.76, 83.79
	20	62.97, 56.88	62.97, 56.65	84.32, 83.97	92.87, 91.94
	30	67.27, 59.50	67.27, 59.37	87.36, 87.02	94.62, 93.78
$\sigma = 10$	5	35.25, 36.17	37.08, 36.09	64.68, 63.49	74.26, 74.20
	10	38.63, 37.68	39.76, 39.08	75.95, 75.14	82.52, 82.18
	20	44.40, 39.65	44.40, 41.19	81.71, 81.33	90.42, 89.05
	30	45.51, 41.13	45.51, 42.04	84.00, 83.86	93.36, 90.84

patch based PCA algorithm to hyperspectral images, it cannot fully exploit the rich spatial information contained in HSIs. To this end, we propose a novel region-based PCA through superpixel segmentation strategy. Table I shows the average overall classification accuracies of three divide-and-conquer strategies<sup>1</sup>, Clustering dependent PCA (ClusterPCA for short), Square patch dependent PCA (SquarePCA for short), and the proposed SuperPCA, with different training sample numbers on the Indian Pines dataset. In addition, the Global PCA method is used as a baseline for comparison. Without loss of generality, we only conduct experiments on this dataset and similar conclusions can be found on the other two datasets.

It should be noted that we use two different classifiers to conduct the classification, i.e., SVM and nearest neighbor (NN). For each result in the bracket, the left is based on the SVM classifier while the right is based on the NN classifier, respectively. To evaluate the performance, we randomly choose  $N = 5, 10, 20, 30$  samples from each class to form the training set<sup>2</sup>, and the rest of the samples for testing. Due to space limitation, we use “T.N.s/C” to denote training numbers in each class in the table. In comparison to ClusterPCA and SquarePCA, the proposed SuperPCA method is more efficient. Global PCA and ClusterPCA have the similar results, which indicates that the preprocessing of clustering is invalid. This is because ClusterPCA does not use the spatial information, and considers each pixel as an isolated data sample. In contrast, SquarePCA and SuperPCA leverage the spatial information inside a square patch or a superpixel region, thus leading to better performance. Such advantage becomes more obvious in the case of noise presence. To demonstrate this, we add additive white Gaussian noise (AWGN) with the variance of  $\sigma = 10$  to the original HSIs. Please refer to the third block in Table I. The performance of ClusterPCA drops drastically when adding noise, while SquarePCA and SuperPCA methods are less affected by the noise. For example, when the noise level is  $\sigma = 10$  and the number of training samples per class is 30, the classification accuracy of ClusterPCA is less than 50%, while SquarePCA and SuperPCA can go beyond 80%. Our SuperPCA method even reaches 93.36% (for

<sup>1</sup>The differences of these three strategies lie in their dividing strategies. In ClusterPCA, all the pixels are clustered by  $K$ -means, and then PCA is applied to each cluster to obtain the dimensionality reduced features. SquarePCA directly performs PCA dimensionality reduction on the squared patches of HSIs.

<sup>2</sup>At a maximum half of the total samples in Grass-pasture-mowed and Oats classes, which have relatively small sample sizes, are chosen.

TABLE II  
NUMBER OF SAMPLES IN THE INDIAN PINES, UNIVERSITY OF PAVIA, AND SALINAS SCENE IMAGES

Indian Pines		University of Pavia		Salinas Scene	
Class Names	Numbers	Class Names	Numbers	Class Names	Numbers
Alfalfa	46	Asphalt	6631	Brocoli_green_weeds_1	2009
Corn-notill	1428	Bare soil	18649	Brocoli_green_weeds_2	3726
Corn-mintill	830	Bitumen	2099	Fallow	1976
Corn	237	Bricks	3064	Fallow_rough_plow	1394
Grass-pasture	483	Gravel	1345	Fallow_smooth	2678
Grass-trees	730	Meadows	5029	Stubble	3959
Grass-pasture-mowed	28	Metal sheets	1330	Celery	3579
Hay-windrowed	478	Shadows	3682	Grapes_untrained	11271
Oats	20	Trees	947	Soil_vinyard_develop	6203
Soybean-notill	972			Corn_senesced_green_weeds	3278
Soybean-mintill	2455			Lettuce_roumaine_4wk	1068
Soybean-clean	593			Lettuce_roumaine_5wk	1927
Wheat	205			Lettuce_roumaine_6wk	916
Woods	1265			Lettuce_roumaine_7wk	1070
Buildings-Grass-Trees-Drives	386			Vinyard_untrained	7268
Stone-Steel-Towers	93			Vinyard_vertical_trellis	1807
Total Number	10249	Total Number	42776	Total Number	54129

SVM classifier) and 90.48% (for NN classifier). In all cases, our method achieves the best performance. In comparison to the SquarePCA method, which also takes into account the spatial information, our proposed SuperPCA method also yields significant performance gains, with an average of 8% increase no matter what kind of classifier is used. We attribute this superiority of SuperPCA over SquarePCA to that pixels in a superpixel are much more like to be the same class than those in a regular patch, and our method can exploit the spatial information more effectively. In summary, clearly demonstrates the robustness of SuperPCA to noise in HSIs for image classification.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we first introduce the three HSI datasets used in our experiments. Then, we assess the impact of the number of superpixels and the reduced dimension on the classification performance using SuperPCA. The comparison results with the state-of-the-art dimensionality reduction approaches are presented.

##### A. Datasets and Experimental Procedure

In order to evaluate the proposed SuperPCA method, we use three publicly available HSI datasets<sup>3</sup>.

- 1) The first HSI dataset is the *Indian Pine*, which is acquired by the AVIRIS sensor in June 1992. The scene is with  $145 \times 145$  pixels and 220 bands in the 0.4-2.45 m region covering the agricultural fields with regular geometry. In this paper, 20 low SNR bands are removed and a total of 200 bands are used for classification. It contains 16 different land-covers, and approximately 10249 labeled pixels are from the ground-truth map.
- 2) The second HSI dataset is the *University of Pavia*, which contains a spatial coverage of  $610 \times 340$  pixels and is collected by the ROSIS sensor under the HySens project managed by DLR (the German Aerospace Agency). It generates 115 spectral bands, of which 12 noisy and

water-bands are removed. It has a spectral coverage from 0.43-0.86  $\mu\text{m}$  and a spatial resolution of 1.3 m. Approximately 42776 labeled pixels with nine classes are from the ground truth map.

- 3) The third HSI dataset is the *Salinas Scene*, collected by the 224-band AVIRIS sensor over Salinas Valley, California, capturing an area over Salinas Valley, CA, USA. It generates  $512 \times 217$  pixels and 204 bands over 0.4-2.5  $\mu\text{m}$  with spatial resolution of 3.7 m, of which 20 water absorption bands are removed before classification. In this image, there are approximately 54129 labeled pixels with 16 classes sampled from the ground truth map.

For the three datasets, the training and testing samples are randomly selected from the available ground truth maps. The class-specific numbers of labeled samples are shown in Table II. To evaluate the performance of our proposed SuperPCA algorithm, we randomly choose  $T = 5, 10, 20, 30$  samples from each class to build the training set, leaving the rest samples to form the testing set. For some classes, e.g., Grass-pasture-mowed and Oats in the Indian Pines image, which have a few labeled samples, we only select a maximum half of the total samples in them. To avoid any bias, all the experiments are repeated 10 times, and we report the average classification accuracy.

We compare the proposed methods with two baseline methods (raw spectral features based and PCA method), as well as the state-of-the-art dimensionality reduction approaches, including five unsupervised feature extraction methods (PCA [13], ICA [18], LPP [25], NPE [24] and LPNPE [26]), and two supervised feature extraction methods (LDA [14] and LFDA [30]). Similar to many previous representative works [31], [63], [64], three measurements, overall accuracy (OA), average accuracy (AA) and Kappa, are used to evaluate the performance of different dimensionality reduction algorithms for HSI classification. Similar to [25], all above-mentioned feature extraction methods are performed on the filtered data using a *5times5* weighted mean filter, and then SVM classifier is applied to filtered data. It is worth noting that for all the comparison methods, they all go through these two procedures. Firstly, they extract the features of the input

<sup>3</sup>[http://www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes)

HSIs with an unsupervised or supervised manner, and then the supervised SVM classifier is adopted to test their classification performances.

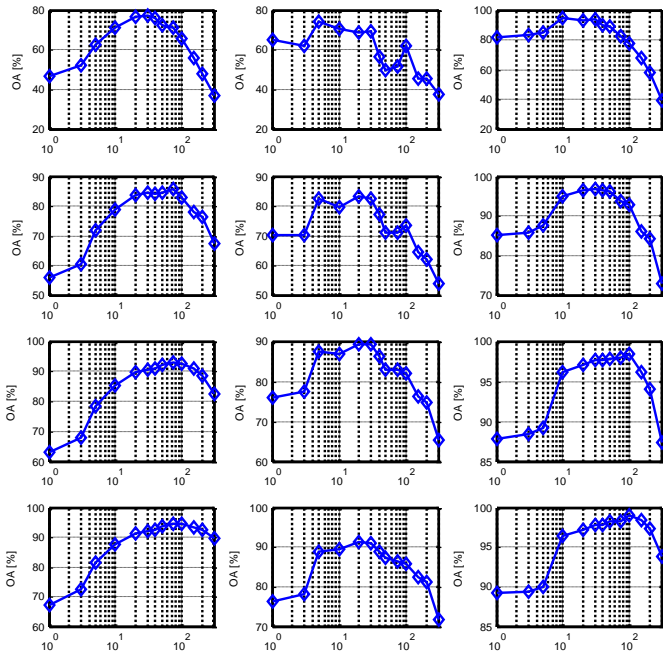


Fig. 6. The influences of the number of superpixels on the overall classification accuracy (%) of the proposed SuperPCA method for Indian Pines (first column), University of Pavia (second column), and Salinas Scene (third column). Different rows represent results of using different training sizes. Specifically, the first to fourth row are the performance when the training size is 5, 10, 20 and 30 samples per class, respectively.

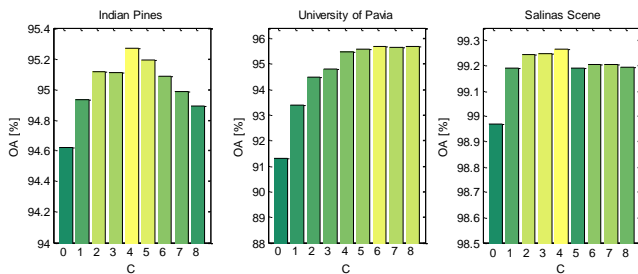


Fig. 7. The OA results of the proposed approaches according to the scale number of  $C$  for Indian Pines, University of Pavia, and Salinas Scene. The best performance is achieved when  $C$  is set to 4, 6, and 4, for these three datasets respectively.

## B. Parameter Tuning

In this subsection, we investigate the influences of (i) the number of superpixels in the SuperPCA approach, (ii) the number of scales of the proposed Multiscale SuperPCA, *i.e.*, the value of the power exponent in Eq. (5) on the performance of the proposed SuperPCA method. Fig. 6 illustrates the OA of SuperPCA as a function of the number of superpixels,  $S_f$ , whose value is chosen from  $\{1, 3, 5, 10, 20, 30, 40, 50, 75, 100, 150, 200, 300\}$ . From the parameter tuning results, we can at least draw the following two conclusions:

- With the increase of the number of superpixels, it shows that the overall performance will first ascend and then

descend. Too large or too small number of superpixel will lead to reduced performance of the proposed SuperPCA method. This is mainly because that too large number of superpixel will result in over-segmented regions and cannot make full use of all the samples belong to the homogeneous area, while a too small number of superpixel will result in under-segmentation and introduce some samples from different homogeneous areas. In addition, when the number of superpixel is too large, each region will have a limited number of pixels, it will not guarantee the stability and reliability of the PCA results, *i.e.*, limited number of samples are not enough to ensure that the real projection.

- By setting a proper value of the number of superpixels, the performance is always better than when the number of superpixels is set to 1 (which reduces to the case of traditional global PCA method). It is evident that the proposed SuperPCA, which takes the spatial homogeneity of HSIs into account, is much more effective than the traditional PCA for capturing the intrinsic data structure.

Based on the above experiments, we can obtain the optimal fundamental superpixel number  $S_f$  for Indian Pines, University of Pavia, and Salinas Scene, which is set to 100, 20, and 100, respectively.

In order to verify the necessity of multiscale fusion, we report the classification results of the proposed SuperPCA approach with different segmentation scales, *i.e.*,  $c$  is set from -5 to 5. When  $c$  is set to zero, it means that the input HSI is segmented with the fundamental superpixel number. Tables III, IV, and V tabulate the OA, AA, and Kappa coefficient when the training number is 30, under different segmentation scales for Indian Pines, University of Pavia, and Salinas Scene images, respectively. The best performance for each class is highlighted in bold typeface. From these tables, we learn that even though SuperPCA can obtain the best overall performance when the power exponent  $c$  is set to zero, *i.e.*, under the fundamental superpixel number  $S_f$ , it cannot achieve the best performance in every class (for the sake of convenience, we highlight the best performance for each class (row) in bold). For example, as shown in Table III, when  $c = -4$ , the OA is obviously inferior to the best scale, *i.e.* 92.15% to 94.62%. In this case, however, the sixth and eighth classes get the best classification accuracy. Similar results can be also observed in Table IV and Table V (please refer to the case when  $c = -5$ ). The OA is minimum, but it achieves the best classification performances in the fifth and ninth classes. All these results demonstrate that the superpixel segmentation based on a single scale is not able to fully model the complexity and diversity of HSIs. Therefore, it is an effective and reliable choice to perform multiscale segmentation based decision fusion for HSI classification.

To further verify the usefulness of the multiscale segmentation strategy as well as to assess the influence of the different values of  $C$ , as shown in Fig. 7, we show the OA results of the proposed multiscale SuperPCA method according to scale number  $C$  for the images of Indian Pines, University of Pavia, and Salinas Scene. By fusing multiscale segmentation based classification results, we can expect better results than single

TABLE III

PERFORMANCE OF THE PROPOSED SUPERPCA APPROACH ON THE INDIAN PINES DATASET WITH DIFFERENT SEGMENTATION SCALES (-5 TO 5).

Class Names	c = -5	c = -4	c = -3	c = -2	c = -1	c = 0	c = 1	c = 2	c = 3	c = 4	c = 5
Alfalfa	97.83	96.96	96.52	96.96	<b>100</b>	<b>100</b>	<b>100</b>	99.13	99.13	99.13	99.13
Corn-notill	83.13	84.48	87.83	87.03	<b>91.12</b>	89.67	88.04	82.85	77.52	72.38	60.67
Corn-mintill	77.95	82.23	87.80	86.23	86.18	<b>92.44</b>	88.88	87.63	85.29	81.90	78.30
Corn	91.64	94.93	92.56	93.43	94.01	95.51	96.28	<b>97.63</b>	96.23	91.84	88.41
Grass-pasture	94.92	96.36	95.92	96.42	<b>97.00</b>	96.56	95.74	95.65	94.02	90.77	89.89
Grass-trees	98.47	<b>99.61</b>	98.49	98.57	98.47	97.93	96.99	94.93	92.51	82.19	78.50
Grass-pasture-mowed	95.71	94.29	89.29	90.00	<b>97.14</b>	97.14	97.14	97.14	97.14	97.14	97.14
Hay-windrowed	99.98	<b>100</b>	99.80	99.80	<b>100</b>	99.64	99.64	99.64	99.64	96.38	93.62
Oats	94.00	99.00	99.00	98.00	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
Soybean-notill	84.71	85.12	91.09	<b>91.40</b>	91.19	90.69	90.67	90.10	85.22	77.14	69.72
Soybean-mintill	90.76	93.30	91.18	93.97	94.72	94.48	94.08	96.02	<b>96.51</b>	89.61	88.34
Soybean-clean	84.74	90.50	91.08	89.40	90.48	92.97	92.13	<b>94.65</b>	93.36	89.17	79.98
Wheat	99.20	99.31	<b>99.43</b>	<b>99.43</b>	<b>99.43</b>	<b>99.43</b>	<b>99.43</b>	<b>99.43</b>	98.46	98.29	96.46
Woods	98.85	98.85	98.84	<b>98.94</b>	98.76	98.89	95.33	91.21	83.65	83.31	72.69
Buildings-Grass-Trees-Drives	97.61	98.06	98.43	98.62	97.33	<b>98.65</b>	98.60	98.26	97.36	94.58	90.73
Stone-Steel-Towers	97.14	97.30	96.98	97.14	98.41	98.41	98.89	99.52	<b>99.52</b>	98.57	98.57
OA[%]	90.37	92.15	93.01	93.45	94.26	<b>94.62</b>	93.41	92.49	89.84	84.83	79.30
AA[%]	92.92	94.39	94.64	94.71	95.89	<b>96.40</b>	95.74	95.24	93.47	90.15	86.38
Kappa	0.8899	0.9101	0.9200	0.9251	0.9342	<b>0.9383</b>	0.9243	0.9133	0.8821	0.8241	0.7579

TABLE IV

PERFORMANCE OF THE PROPOSED APPROACH ON THE UNIVERSITY OF PAVIA DATASET WITH DIFFERENT SEGMENTATION SCALES (-5 TO 5).

Class Names	c = -5	c = -4	c = -3	c = -2	c = -1	c = 0	c = 1	c = 2	c = 3	c = 4	c = 5
Asphalt	71.68	71.99	78.32	79.10	78.06	81.40	<b>86.80</b>	83.46	79.01	82.16	77.30
Bare soil	74.67	93.97	86.32	92.12	91.22	<b>94.41</b>	92.63	88.63	84.32	83.76	77.13
Bitumen	81.58	89.98	88.89	89.96	94.40	<b>97.09</b>	96.02	95.22	89.93	93.33	92.35
Bricks	90.89	92.75	<b>92.90</b>	86.34	89.85	86.21	82.88	78.60	78.70	75.80	70.59
Gravel	<b>97.89</b>	97.59	97.53	97.00	97.47	96.65	97.05	96.05	91.58	91.59	90.31
Meadows	69.40	89.11	86.95	90.86	91.02	<b>92.23</b>	89.96	90.13	84.78	83.20	82.11
Metal sheets	90.78	89.50	94.69	90.76	94.20	94.55	95.66	95.68	95.98	<b>97.45</b>	96.94
Shadows	73.50	81.07	86.08	88.11	87.58	88.16	92.44	<b>95.52</b>	91.62	91.64	86.66
Trees	<b>99.97</b>	99.44	99.64	99.65	97.56	98.53	98.85	97.47	97.72	97.26	97.04
OA[%]	76.74	88.69	86.61	89.36	89.32	<b>91.30</b>	91.23	88.83	84.92	84.97	80.28
AA[%]	83.37	89.49	90.15	90.43	91.26	92.14	<b>92.48</b>	91.20	88.18	88.47	85.60
Kappa	0.7030	0.8516	0.8265	0.8608	0.8604	<b>0.8856</b>	0.8851	0.8547	0.8057	0.8064	0.7485

TABLE V

PERFORMANCE OF THE PROPOSED APPROACH ON THE SALINAS SCENE DATASET WITH DIFFERENT SEGMENTATION SCALES (-5 TO 5).

Class Names	c = -5	c = -4	c = -3	c = -2	c = -1	c = 0	c = 1	c = 2	c = 3	c = 4	c = 5
Brocoli_green_weeds_1	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	98.74	93.75
Brocoli_green_weeds_2	<b>99.99</b>	99.95	99.8	99.96	99.85	99.78	99.73	97.9	96.96	91.57	82.39
Fallow	<b>100</b>	<b>100</b>	99.97	99.23	98.21	99.67	98.32	99.22	99.54	96.78	96.03
Fallow_rough_plow	99.07	99.02	99.05	99.12	98.91	99.16	<b>99.32</b>	99.27	96.74	95.50	93.20
Fallow_smooth	99.45	<b>99.46</b>	99.44	99.38	98.69	99.37	98.62	98.70	96.72	95.32	89.13
Stubble	<b>99.88</b>	99.87	99.76	99.82	99.79	98.37	98.33	98.68	94.83	87.45	80.62
Celery	<b>99.91</b>	99.55	98.06	98.19	98.11	97.78	98.00	97.81	96.73	94.65	81.93
Grapes_untrained	96.63	93.66	94.87	96.62	96.52	99.39	<b>99.48</b>	98.13	98.93	94.74	97.02
Soil_vinyard_develop	99.25	99.39	99.54	<b>99.58</b>	99.51	99.02	99.57	95.11	89.46	81.65	68.6
Corn_senesced_green_weeds	92.55	96.09	<b>97.25</b>	97.08	96.59	97.16	94.35	94.59	89.89	83.56	79.72
Lettuce_roumaine_4wk	98.01	98.30	98.05	98.14	98.61	98.38	98.42	98.72	<b>98.78</b>	95.31	92.36
Lettuce_roumaine_5wk	99.28	99.35	99.27	99.67	98.34	<b>99.80</b>	99.51	99.10	97.53	96.85	91.75
Lettuce_roumaine_6wk	98.21	98.21	98.21	98.19	98.23	<b>98.28</b>	98.09	97.99	97.79	97.81	97.14
Lettuce_roumaine_7wk	95.41	97.83	97.74	98.06	<b>98.25</b>	97.95	98.01	96.29	94.92	93.65	92.57
Vinyard_untrained	91.17	97.46	97.11	95.96	96.12	<b>99.05</b>	97.43	94.58	83.91	71.9	60.08
Vinyard_vertical_trellis	99.18	<b>99.33</b>	98.98	99.18	98.92	98.99	98.66	97.97	91.54	87.30	85.17
OA[%]	97.29	97.78	97.93	98.16	97.99	<b>98.97</b>	98.57	97.26	94.19	88.85	83.12
AA[%]	98.00	98.59	98.57	98.64	98.42	<b>98.89</b>	98.49	97.75	95.27	91.42	86.34
Kappa	0.9698	0.9753	0.9770	0.9795	0.9776	<b>0.9886</b>	0.9841	0.9694	0.9349	0.8746	0.8088



TABLE VI  
CLASSIFICATION RESULTS (IN TERMS OF OA) OF THE PROPOSED APPROACHES AND EIGHT COMPARISON ALGORITHMS ON THREE HSI DATASETS WITH DIFFERENT TRAINING NUMBERS.

Datasets	T.N.s/C	Raw	PCA	ICA	LPP	NPE	LPNPE	LDA	LFDA	SuperPCA	MSuperPCA
Indian Pines	5	44.88	46.37	45.21	53.58	53.68	67.25	59.95	59.62	77.34	<b>78.68</b>
	10	55.77	55.72	57.12	70.41	70.49	76.45	69.30	64.91	85.76	<b>87.12</b>
	20	63.81	62.97	64.41	80.26	79.87	83.51	76.56	74.01	93.90	<b>95.69</b>
	30	68.77	67.27	68.92	84.43	83.98	90.10	89.51	90.19	94.62	<b>96.78</b>
	200	84.01	84.40	82.86	94.31	94.16	97.80	98.55	<b>99.15</b>	97.13	98.25
University of Pavia	5	64.59	65.26	66.58	70.86	68.35	76.12	72.43	74.67	74.39	<b>78.49</b>
	10	70.22	70.15	71.39	81.29	80.63	82.55	81.24	78.95	83.42	<b>91.67</b>
	20	75.85	75.91	76.65	86.00	85.69	88.56	85.00	86.98	89.38	<b>95.37</b>
	30	76.45	76.31	76.87	86.90	87.19	90.56	87.91	90.19	91.30	<b>95.68</b>
	200	85.71	85.70	85.79	94.08	93.69	97.50	95.72	98.72	96.99	<b>98.84</b>
Salinas Scene	5	81.79	81.87	81.75	85.23	84.86	92.09	89.03	88.83	94.42	<b>95.00</b>
	10	85.24	85.28	85.74	88.60	88.99	94.52	91.46	82.77	96.78	<b>98.15</b>
	20	87.85	87.79	88.08	90.61	90.69	95.89	93.72	93.56	98.37	<b>99.04</b>
	30	88.93	89.24	89.28	91.73	91.69	96.66	95.87	95.89	98.97	<b>99.27</b>
	200	91.48	91.94	91.74	96.18	95.88	99.09	99.87	99.57	99.63	<b>99.70</b>

scale SuperPCA method, *i.e.*, setting  $C$  to 0. When setting the value of  $C$  to 4, 6, and 4 for Indian Pines, University of Pavia, and Salinas Scene images, the improvements of single scale SuperPCA method over multiscale SuperPCA method are 0.65%, 4.38%, and 0.30%, respectively. From these results, we observe that the improvement on the University of Pavia image is more obvious than the other two images. This is mainly due to the following two reasons: on the one hand, the single scale SuperPCA performs not very well and has a relatively large space for improvement. On the other hand, the University of Pavia image has richer and more complex texture information, and it is much more difficult for the single scale based segmentation method to capture these useful spatial knowledge. At the same time, we also observe another phenomenon: in order to achieve high classification accuracy, the relatively complex HSIs may require a larger scale number to exploit its spatial information, *e.g.*, the optimal scale number of the University of Pavia image is 6, which is larger than that of the other two HSI datasets.

### C. Comparison Results with State-of-the-arts

The classification maps obtained with above-mentioned three public HSI datasets for the proposed SuperPCA and MSuperPCA approaches and the comparison methods are given in Fig. 8, Fig. 9, and Fig. 10. Here, we only show the results when the number of training samples is set to 30. From these maps, we can learn that raw spectral features based method, PCA [13], ICA [18], LPP [25], and NPE [24] exhibit higher classification errors than other methods. Among the comparison unsupervised methods, LPNPE [26] achieves the best performance due to its local spatialCpectral scatter based effective spatial information extraction. By taking advantage of the discrimination information of the labeled samples, these supervised methods (LDA [14] and LFDA [30]) can produce very good results. As for the datasets of Indian Pines and University of Pavia, the proposed SuperPCA and MSuperPCA are clearly better than previous arts. When comparing the classification maps of our methods with LDA [14] and LFDA [30] (please refer to the University of Pavia dataset), it can be observed that our methods can achieve much better results for these large regions (*i.e.*, Bare soil and Meadows), which

can be attributed to the efficient segmentation. Through the fusion of multiscale segmentation based classification results, MSuperPCA can improve the result of single scale segmentation based SuperPCA and obtain accurate classification maps (please refer to the edges and the holes of regions).

According to the experimental settings of LPNPE method [26], which can be seen as the best unsupervised feature extraction method for HSI classification to the best of our knowledge, we further randomly choose  $T = 5, 10, 20, 30$  samples from each class to form the training set to test the comparison results (in terms of OA), respectively. Table VI tabulates the OA performance of different approaches. From the results of each individual method, the OA performance of the proposed SuperPCA is better than others in most instances, and this advantage is particularly evident when the number of training samples is small. With the increase of training number, the performance of supervised methods becomes much better. This can be explained as follows: with the increase of labeled training samples, these supervised methods are able to use more discriminant information from the training samples.

To further utilize the spatial information of HSIs, MSuperPCA is advocated to fuse the decisions of SuperPCA with different segmentation scales. By comparing the last two columns, we can clearly see that the performance of MSuperPCA is better than that of SuperPCA in all cases (regardless of different datasets or different training sample numbers). In particular, the improvement of MSuperPCA over SuperPCA is much more impressive on the University of Pavia dataset, that is over 4% higher accuracy. It is because of the rich texture information contained in that dataset.

The above results show that our proposed method can achieve good performance when the number of training samples are small. At the fourth row of each block, we additionally provide the results when the number of training samples is relatively large, *e.g.*,  $T = 200$ . In this situation, these supervised methods (LDA [14] and LFDA [30]) can learn more discriminative information from the labeled training data for classification. Therefore, all of them have considerable performance improvements as compared to the situation when the training samples are limited. Nevertheless, the results of SuperPCA, which do not use any label information, are

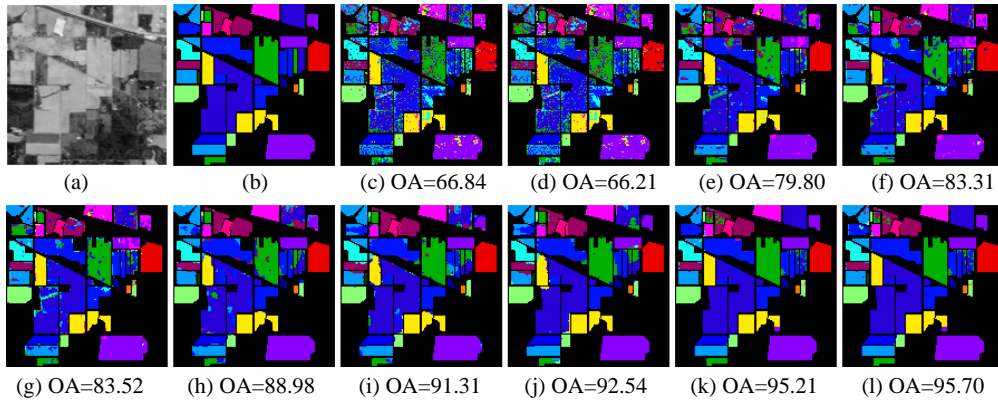


Fig. 8. Classification maps obtained with the Indian Pine dataset. (a) First principal component, (b) Ground truth, (c) Raw pixel, (d) PCA [13], (e) ICA [18], (f) LPP [25], (g) NPE [24], (h) LPNPE [26], (i) LDA [14], (j) LFDA [30], (k) SuperPCA, (l) MSuperPCA.

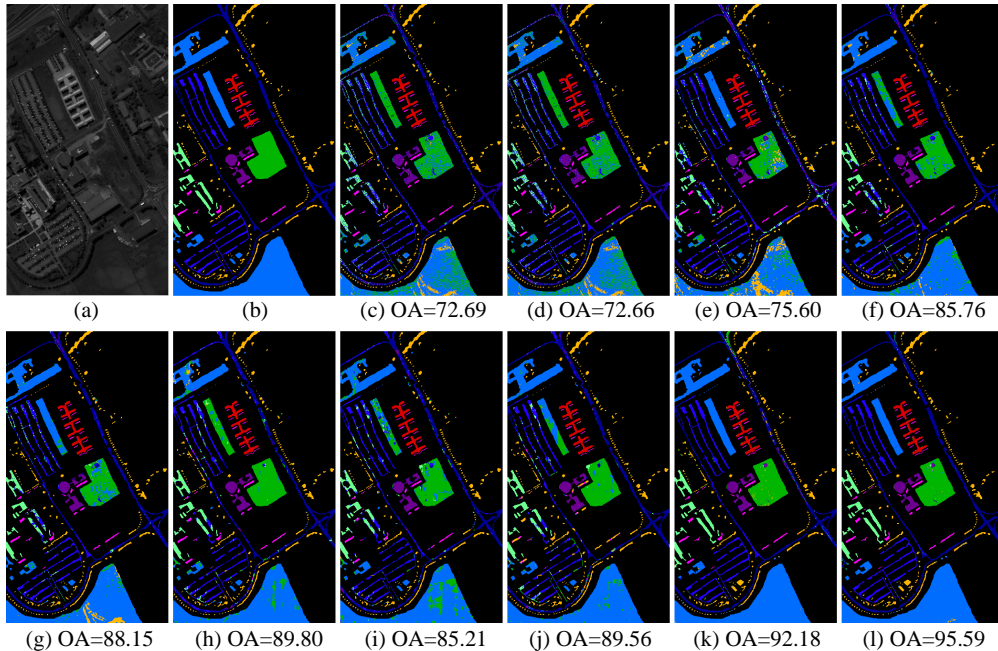


Fig. 9. Classification maps obtained with the University of Pavia dataset. (a) First principal component, (b) Ground truth, (c) Raw pixel, (d) PCA [13], (e) ICA [18], (f) LPP [25], (g) NPE [24], (h) LPNPE [26], (i) LDA [14], (j) LFDA [30], (k) SuperPCA, (l) MSuperPCA.

still very competitive in this situation. By fusing multiscale classification results, MSuperPCA can even surpass LDA [14] and LFDA [30] on the University of Pavia and Salinas datasets. This proves the effectiveness of the proposed method once again.

#### D. Running Times

In Table VII, we report the run times of extracting the dimensionality reduced features of different algorithms on the Indian Pine, University of Pavia, and Salinas Scene images with different training numbers ( $T = 5, 10, 20, 30$ ). As for the proposed method, we report the whole running time including the segmentation and dimensionality reduction of all superpixels. All methods were tested on MATLAB R2014a using an Intel Xeon CPU with 3.50 GHz and 16G memory PC with Windows platform. The testing time of all methods is measured using a single-threaded MATLAB process. It should be noted that for these unsupervised methods, PCA [13], ICA

[18], LPP [25], NPE [24], LPNPE [26], and our proposed SuperPCA, the running time will not change with the training number per class. PCA, LDA [14], and LFDA [30] show the fastest performance. While LPP [25] and NPE [24] need to construct the large similarity graph and decompose it via SVD, and thus the computational complexities of them are relatively high. With the increase of training numbers, the run times of these supervised methods (LDA [14] and LFDA [30]) will also increase. The timings reveal that although our method requires pre-segmentation and feature extraction for each region, the run time is still acceptable. Also, thanks to the independence of the dimensionality reduction of each superpixel, we can accelerate the algorithm simply by parallel computation.

#### E. Discussions

Since the key idea of the proposed methods is to oversegment the HSIs and perform PCA superpixelwisely, how to

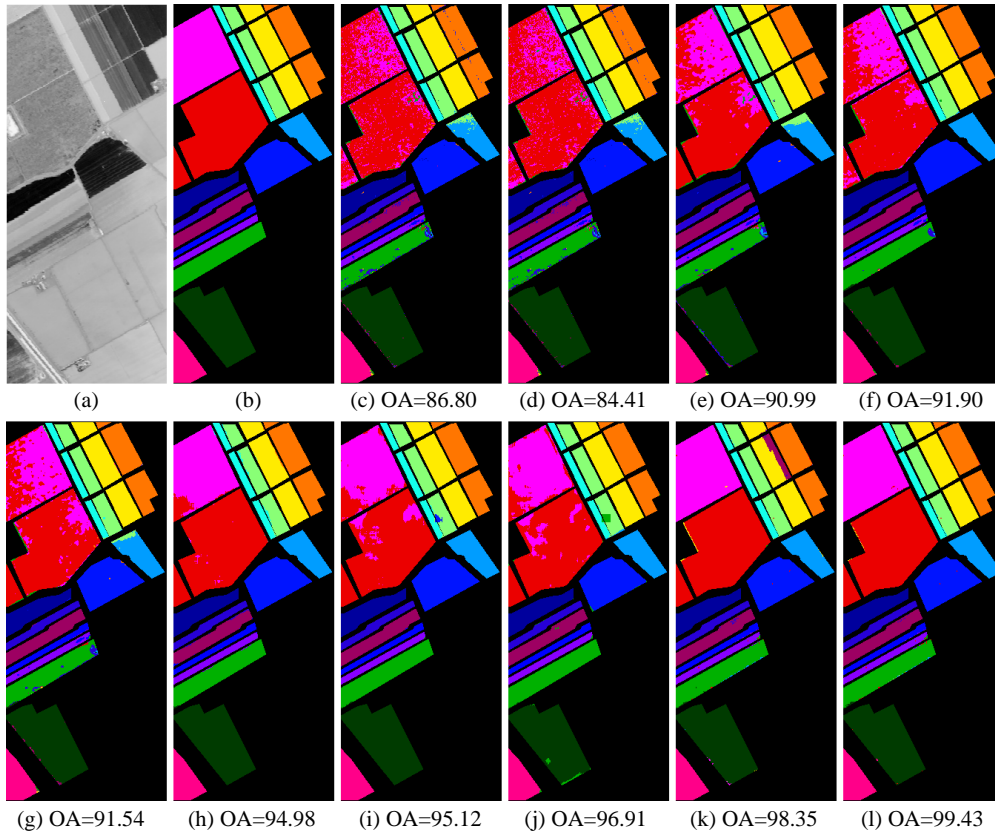


Fig. 10. Classification maps obtained with the Salinas Scene dataset. (a) First principal component, (b) Ground truth, (c) Raw pixel, (d) PCA [13], (e) ICA [18], (f) LPP [25], (g) NPE [24], (h) LPNPE [26], (i) LDA [14], (j) LFDA [30], (k) SuperPCA, (l) MSuperPCA.

TABLE VII

RUNNING TIMES OF THE FEATURE EXTRACTION PROCESS (IN SECONDS) OF THE PROPOSED APPROACH AND SOME COMPARISON ALGORITHMS ON THE THREE HSI DATASETS WITH DIFFERENT TRAINING NUMBERS.

Dataset	T.N.s/C	PCA	ICA	LPP	NPE	LPNPE	LDA	LFDA	SuperPCA
Indian Pines	5	0.0076	2.5417	0.2428	0.7018	0.3834	0.0027	0.0167	0.6879
	10	0.0076	2.5417	0.2428	0.7018	0.3834	0.0047	0.0249	0.6879
	20	0.0076	2.5417	0.2428	0.7018	0.3834	0.0054	0.0388	0.6879
	30	0.0076	2.5417	0.2428	0.2428	0.3834	0.0057	0.0356	0.6879
University of Pavia	5	0.4004	3.1445	2.9477	6.5322	1.2357	0.0017	0.0053	2.8867
	10	0.4004	3.1445	2.9477	6.5322	1.2357	0.0022	0.0110	2.8867
	20	0.4004	3.1445	2.9477	6.5322	1.2357	0.0024	0.0097	2.8867
	30	0.4004	3.1445	2.9477	6.5322	1.2357	0.0024	0.0095	2.8867
Salinas Scene	5	0.4145	5.8702	4.7492	9.8365	1.2003	0.0027	0.0174	2.7452
	10	0.4145	5.8702	4.7492	9.8365	1.2003	0.0046	0.0260	2.7452
	20	0.4145	5.8702	4.7492	9.8365	1.2003	0.0066	0.0377	2.7452
	30	0.4145	5.8702	4.7492	9.8365	1.2003	0.0067	0.0391	2.7452

determine the parameter of the superpixel segmentation model (i.e., the superpixel number in ERS) and the segmentation scales is a crucial and open problem. In this paper, we set them experimentally to achieve the best performance. In fact, the segmentation scales and superpixel number jointly determine the minimum and maximum homogeneous regions, which can be deduced from the Eq. (5). The searching of optimal segmentation scales and superpixel number can be converted to the problem of setting the size of minimum and maximum homogeneous regions of the given HSIs. Obviously, the size of homogeneous region is determined by the texture information. Therefore, the most direct approach is to detect the edges in a given images through some edge detectors, such

as Canny and Sobel. Therefore, we can obtain the texture ratio, which can be used to define the size of homogeneous region.

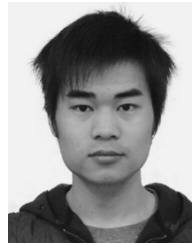
## V. CONCLUSIONS

In this paper, we propose a simple but very effective technique for unsupervised feature extraction of hyperspectral imagery based on superpixelwise principal component analysis (SuperPCA). By segmenting the entire hyperspectral image (HSI) to many different homogeneous regions, which have similar reflectance properties, it can facilitate the dimensionality reduction process of finding the essential low-dimensional feature space of HSIs. To take full advantage of the spatial information contained in the HSIs, which cannot be extracted using a single scale, we further advocate a decision fusion strategy through multiscale segmentation based on the SuperPCA model (MSuperPCA). Extensive experiments on three standard HSI datasets demonstrate that the proposed SuperPCA and MSuperPCA algorithms outperform the existing state-of-the-art feature extraction methods, including unsupervised feature extraction methods as well as supervised feature extraction methods, especially when the training samples are limited. When the number of the training samples is relatively large, the proposed algorithm can still obtain very competitive classification results when compared with these supervised feature extraction methods. Because of inheriting the merits of PCA technology, the proposed SuperPCA can be also used as a preprocessing for many hyperspectral image processing and analysis tasks.

## REFERENCES

- [1] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
- [2] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.
- [3] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, 2018, DOI: 10.1109/TGRS.2018.2820040.
- [4] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [5] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, 2008.
- [6] H. Yang, Q. Du, and G. Chen, "Unsupervised hyperspectral band selection using graphics processing units," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 660–668, 2011.
- [7] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, June 2016.
- [8] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, 2016.
- [9] L. M. Bruce, C. H. Koger, and J. Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2331–2338, 2002.
- [10] W. Zhao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, 2016.
- [11] W. Sun, G. Yang, B. Du, L. Zhang, and L. Zhang, "A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4032–4046, 2017.
- [12] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cyber.*, 2016.
- [13] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [14] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 625–629, 2008.
- [15] M. A. Hossain, M. Pickering, and X. Jia, "Unsupervised feature extraction based on a mutual information measure for hyperspectral image classification," in *IGARSS*. IEEE, 2011, pp. 1720–1723.
- [16] V. Laparra, J. Malo, and G. Camps-Valls, "Dimensionality reduction via regression in hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1026–1036, Sept 2015.
- [17] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, 2019.
- [18] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, 2006.
- [19] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [20] J. Ma, J. Jiang, C. Liu, and Y. Li, "Feature guided gaussian mixture model with semi-supervised em and local geometric constraint for retinal image registration," *Inf. Sci.*, vol. 417, pp. 128–142, 2017.
- [21] D. Lungu, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 55–66, 2014.
- [22] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, 2005.
- [23] L. Ma, M. M. Crawford, and J. Tian, "Anomaly detection for hyperspectral images based on robust locally linear embedding," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 31, no. 6, pp. 753–762, 2010.
- [24] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *ICCV*, vol. 2. IEEE, 2005, pp. 1208–1213.
- [25] X. He and P. Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*, 2004, pp. 153–160.
- [26] Y. Zhou, J. Peng, and C. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, 2015.
- [27] L. Xu, A. Wong, F. Li, and D. A. Clausi, "Intrinsic representation of hyperspectral imagery for unsupervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 1118–1130, 2016.
- [28] V. Slavkovic, S. Verstockt, W. D. Neve, S. V. Hoecke, and R. V. D. Walle, "Unsupervised spectral sub-feature learning for hyperspectral image classification," *International Journal of Remote Sensing*, vol. 37, no. 2, pp. 309–326, 2016.
- [29] W. Wei, Y. Zhang, and C. Tian, "Latent subclass learning-based unsupervised ensemble feature extraction method for hyperspectral image classification," *Remote Sensing Letters*, vol. 6, no. 4, pp. 257–266, 2015.
- [30] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, 2012.
- [31] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitioned clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, 2009.
- [32] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, 2014.
- [33] L. Wang, J. Zhang, P. Liu, K.-K. R. Choo, and F. Huang, "Spectral-spatial multi-feature-based deep learning for hyperspectral remote sensing image classification," *Soft Computing*, vol. 1, no. 21, pp. 213–221, 2016.
- [34] J. Wen, J. E. Fowler, M. He, Y. Q. Zhao, C. Deng, and V. Menon, "Orthogonal nonnegative matrix factorization combining multiple features for spectral-spatial dimensionality reduction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4272–4286, July 2016.
- [35] H. Pu, Z. Chen, B. Wang, and G. M. Jiang, "A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7008–7022, Nov 2014.
- [36] L. Ma, X. Zhang, X. Yu, and D. Luo, "Spatial regularized local manifold learning for classification of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 609–624, 2016.
- [37] J. Jiang, C. Chen, Y. Yu, X. Jiang, and J. Ma, "Spatial-aware collaborative representation for hyperspectral remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 404–408, 2017.
- [38] N. H. Ly, Q. Du, and J. E. Fowler, "Collaborative graph-based discriminant analysis for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2688–2696, 2014.
- [39] —, "Sparse graph-based discriminant analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3872–3884, 2014.
- [40] W. Li, J. Liu, and Q. Du, "Sparse and low-rank graph for discriminant analysis of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4094–4105, 2016.
- [41] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern recognition*, vol. 26, no. 9, pp. 1277–1294, 1993.
- [42] F. Verdoja and M. Grangetto, "Fast superpixel-based hierarchical approach to image segmentation," in *International Conference on Image Analysis and Processing*. Springer, 2015, pp. 364–374.
- [43] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *CVPR*, 2013, pp. 1155–1162.
- [44] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.
- [45] A. Levinstein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, 2009.
- [46] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *CVPR*. IEEE, 2011, pp. 2097–2104.
- [47] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions-I," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.

- [48] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, 2015.
- [49] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, 2015.
- [50] W. Li, S. Prasad, and J. E. Fowler, "Classification and reconstruction from random projections for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 833–843, 2013.
- [51] B. Zhang, J. Gu, C. Chen, J. Han, X. Su, X. Cao, and J. Liu, "One-to-one networks for compression artifacts reduction in remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018.
- [52] S. Zhang, S. Li, W. Fu, and L. Fang, "Multiscale superpixel-based sparse representation for hyperspectral image classification," *Remote Sensing*, vol. 9, no. 2, p. 139, 2017.
- [53] F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Information Sciences*, vol. 397, pp. 48–68, 2017.
- [54] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, p. 2274, 2012.
- [55] C. Coburn and A. C. Roberts, "A multiscale texture analysis procedure for improved forest stand classification," *International journal of remote sensing*, vol. 25, no. 20, pp. 4287–4308, 2004.
- [56] S. Prasad and L. M. Bruce, "Decision fusion with confidence-based weight assignment for hyperspectral target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1448–1456, 2008.
- [57] M. Ding, S. Antani, S. Jaeger, Z. Xue, S. Candemir, M. Kohli, and G. Thoma, "Local-global classifier fusion for screening chest radiographs," in *Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*, vol. 10138. International Society for Optics and Photonics, 2017, p. 101380A.
- [58] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, 2016.
- [59] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, and T.-H. Ma, "Patch-based principal component analysis for face recognition," *Computational intelligence and neuroscience*, vol. 2017, 2017.
- [60] Y. Zhao, X. Shen, N. D. Georganas, and E. M. Petriu, "Part-based pca for facial feature extraction and classification," in *HAVE*. IEEE, 2009, pp. 99–104.
- [61] Y. Gao, J. Ma, and A. L. Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2545–2560, 2017.
- [62] C.-A. Deledalle, J. Salmon, A. S. Dalalyan *et al.*, "Image denoising with patch based pca: local versus global," in *BMVC*, vol. 81, 2011, pp. 425–455.
- [63] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Svm-and mrf-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, 2010.
- [64] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, 2006.



**Jiayi Ma** received the B.S. degree from the Department of Mathematics, and the Ph.D. Degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. From 2012 to 2013, he was with the Department of Statistics, University of California at Los Angeles. He is now an Associate Professor with the Electronic Information School, Wuhan University. His current research interests include in the areas of computer vision, machine learning, and pattern recognition.



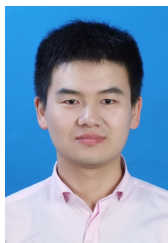
**Chen Chen** received the B.E. degree from Beijing Forestry University, Beijing, China, in 2009 and the MS degree from Mississippi State University, Starkville, in 2012 and the PhD degree from the University of Texas at Dallas, Richardson, TX in 2016. He is currently a Post-Doc in the Center for Research in Computer Vision at University of Central Florida (UCF). His research interests include compressed sensing, signal and image processing, pattern recognition and computer vision.



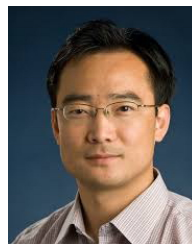
**Zhongyuan Wang** received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2008. Dr. Wang is now an associate professor with School of Computer, Wuhan University, Wuhan, China. He is currently directing two projects funded by the National Natural Science Foundation Program of China. His research interests include video compression, image processing, and multimedia communications, etc.



**Zhihua Cai** received the B.Sc. degree from Wuhan University, Wuhan, China, in 1986, the M.Sc. degree from Beijing University of Technology, Beijing, China, in 1992, and the Ph.D. degree from China University of Geosciences, Wuhan, in 2003. He is currently a faculty member with the School of Computer Science, China University of Geosciences. His main research areas include data mining, machine learning, evolutionary computation, and their applications.



**Junjun Jiang** received the B.S. degree from School of Mathematical Sciences, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from School of Computer, Wuhan University, Wuhan, China, in 2014. Since Oct. 2016, he has been a Project Researcher with the National Institute of Informatics, Tokyo, Japan. His research interests include image processing and computer vision.



**Lizhe Wang** received the B.Eng. (Hons.) and the M.Eng. degrees from Tsinghua University, Beijing, China, and the Doctor of Engineering degree in applied computer science (magna cum laude) from University Karlsruhe (currently, Karlsruhe Institute of Technology), Karlsruhe, Germany.

He is a "ChuTian" Chair Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. Prof. Wang is a fellow of Institution of Engineering and Technology and British Computer Society.