

Superpixelwise Collaborative-Representation Graph Embedding for Unsupervised Dimension Reduction in Hyperspectral Imagery

Huan Liu, Wei Li [✉], Senior Member, IEEE, Xiang-Gen Xia, Fellow, IEEE, Mengmeng Zhang [✉], and Ran Tao [✉], Senior Member, IEEE

Abstract—Recently, graph-embedding framework has been developed for dimensionality reduction (DR) and classification of hyperspectral images (HSI). However, it suffers from intraclass difference and interclass similarity in complex scenarios. In this article, an unsupervised DR method called superpixelwise collaborative-representation graph embedding (SPCRGE) is proposed for the HSI classification. In SPCRGE, homogeneous regions called superpixels are generated by grouping spectral-similar and spatially adjacent pixels. Pixels in one homogeneous region come from one class with high probability. Then, Laplacian regularized superpixelwise collaborative representation (SPCR) of a query pixel, i.e., using all pixels in its superpixel to represent the pixel, is obtained by solving a generalized Sylvester equation to extract commonality and maintain individuality of the pixel to some extent. Finally, a global projection matrix to a low-dimensional space is calculated by reducing the discrepancy between SPCRs and the original spectral features, and reducing the differences between pixels from one superpixel and increasing the differences between pixels from different superpixels simultaneously. Superior classification performances on several HSI datasets demonstrate the effectiveness of the proposed SPCRGE.

Index Terms—Collaborative representation, graph embedding, hyperspectral image, Laplacian matrix, spectral-spatial dimensionality reduction.

I. INTRODUCTION

SPECTRAL information is important for discriminating different materials, since different materials possess different spectral properties. Compared with traditional and multispectral images, hyperspectral imagery (HSI) with hundreds of contiguous narrow bands, possesses more abundant spectral information, which has been extensively exploited in image processing

field, such as HSI classification [1]–[5], clustering [6], unmixing [7], [8], target detection [9], [10], anomaly detection [11], [12], and change detection [13]. However, in HSI classification, contiguous bands of hyperspectral image are often highly correlated, which may degrade its classification performance. Furthermore, high dimensionality of hyperspectral image may cause Hughes phenomenon [14], which further deteriorates the classification performance.

To alleviate the difficulty of HSI classification from redundant spectral information and the curse of high dimensionality, dimensionality reduction (DR) has been widely used. In DR, a mapping that transforms the original high-dimensional spectral features into new low-dimensional features is needed. The mapping could be nonlinear or linear, unsupervised, supervised, or semisupervised [15], [16]. Some traditional DR methods, such as principal component analysis (PCA) [17], linear discriminative analysis (LDA) [18], and nonparametric weighted feature extraction (NWFE) [19], construct the mappings from the statistical perspective. Manifold learning as one big family of DR methods tries to find the mapping by preserving local geometric structure. The relatively early methods of manifold learning include isometric mapping (Isomap) [20], locally linear embedding (LLE) [21], Laplacian eigenmaps (LE) [22], and the linear extensions of LLE and LE, i.e., neighborhood preserving embedding (NPE) [23], and locality preserving projection [24], respectively.

A general framework called graph embedding has been proposed [25], which unifies many DR methods and can be used as a general platform for developing new DR algorithms. In this framework, an intrinsic undirected weighted graph and a penalty undirected weighted graph are constructed to embed the statistic or geometrical information and low-dimensional features can be obtained by solving an optimization problem. After that, many new DR methods have been proposed. Unsupervised noise-robust sparsity preserving graph embedding (SPGE) [26] and its supervised version, i.e., sparsity graph-based discriminative analysis (SGDA) [27] learn the low-dimensional features by preserving sparse representation (SR) relationships. In [28], the performance of collaborative representation (CR) has been proved as good as that of SR while CR is superior to SR in terms of computational complexity. Subsequently, collaborative graph-based discriminative analysis (CGDA) and block collaborative graph-based discriminative analysis (BCGDA) have been

Manuscript received December 8, 2020; revised February 14, 2021; accepted April 30, 2021. Date of publication May 4, 2021; date of current version May 24, 2021. This work was supported in the Beijing Natural Science Foundation under Grant JQ20021 and Grant L191004, and in part by the National Natural Science Foundation of China under Grant 61922013 and Grant U1833203. (Corresponding author: Wei Li.)

Huan Liu, Wei Li, Mengmeng Zhang, and Ran Tao are with the School of Information, and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: huanliu233@gmail.com; leewei36@gmail.com; 752020002@bit.edu.cn; rantao@bit.edu.cn).

Xiang-Gen Xia is with the School of Information, and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with the Department of Electrical, and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: xianggen@udel.edu).

Digital Object Identifier 10.1109/JSTARS.2021.3077460

proposed [29] and experiments have shown that the classification performances of CGDA and BCGDA with lower computational complexity are even better than SGDA. Most recently, an unsupervised collaboration-competition graph preserving embedding (CCPGE) was proposed by combining global and local graphs [30].

All the abovementioned DR methods utilize only spectral information. The classification performance is generally not satisfying because of intraclass difference and interclass similarity in complex scenarios. To obtain better performance, abundant spatial information should be utilized as well. Neighboring relations in spatial space can provide extra information since neighboring pixels have great chances to be from one class. To embed spatial information in DR, many spectral-spatial DR methods have been proposed. One strategy is to incorporate spatial information into distance calculation or selecting neighboring pixels. Three methods called spatial coherence distance [31], image patches distance [32], and spatial-spectral (SSCD) [33], design distance formulas between different patches, incorporate the spatial information into traditional DR methods and perform well as compared to traditional DR methods. Another strategy is to perform DR algorithm on a tensor patch of specified size. Tensor sparse and low-rank graph-based discriminant analysis (TSLGDA) [34] treats every tensor patch as features of center pixel and performs the DR algorithm on tensors. Other strategy like orthogonal total variation component analysis (OTVCA) [35] learns low-dimensional features with total variation regularization based on that the learned features should be smooth in spatial space.

However, these methods consider only specified neighborhood around a query pixel and cannot fully reflect local manifold in spectral-spatial feature space. Therefore, an adaptive neighborhood system is required. This system can be realized by generating nonoverlapping homogeneous regions. One of the most powerful tools to create homogeneous regions is superpixel segmentation. Based on superpixel segmentation, many classification methods have been proposed, such as Hidden Markov Random Fields-SVM (HMRF-SVM) [36], set-to-set distance-based spectral-spatial classification (SD-SSC) [37], superpixel-based extended random walker (SPERW) [38], and superpixel contracted graph-based learning (SGL) [39]. All these methods assume that pixels within a superpixel are from one class and have achieved good classification performances. There are some spectral-spatial DR methods based on superpixel segmentation. One method is called superpixel-based linear discriminative analysis (SPLDA) [40], where superpixelwise geometric structure is preserved. In [41], SuperPCA performs superpixelwise PCA to extract local principal components. However, there are some problems. First, local principal components lack global perspective. Since different superpixels may have different principal projections, projection values in different superpixels have different meanings. Second, after the number of principal components is specified, the number of pixels in a superpixel may be less than the number of principal components. Subsequently, the number of principal components of this superpixel has to be less than the specified number, which may deteriorate the

classification performance. This situation may exist in a complex urban scenario.

Based on the above observations, superpixelwise collaborative-representation graph embedding (SPCRGE) is proposed in this article. SPCRGE first performs PCA on the original HSI and utilizes the first three principal components to generate superpixels via a superpixel segmentation algorithm. Then, Laplacian regularized superpixelwise collaborative representation (SPCR) of a query pixel, i.e., using all pixels in its superpixel to represent the pixel, is obtained by solving a generalized Sylvester equation to extract commonality and maintain individuality of the pixel to some extent. Finally, a global projection matrix to a low-dimensional space is obtained under two principles. One principle is to reduce the discrepancy between SPCRs and the original spectral features. The other is to reduce the differences between SPCRs of pixels from one superpixel and increase the differences between SPCRs of pixels from different superpixels simultaneously. In this way, redundant spectral information is reduced and abundant spatial information is utilized as well.

The main contributions of this article are listed as follows.

1) From our best knowledge, this is the first time that intrinsic graph of graph-embedding DR method is based on the newly defined within-superpixel representation error matrix by combining superpixel segmentation and CR, saying, SPCR. SPCR helps to preserve the main spectral properties so as to alleviate environment-induced difference to some extent. As is known, spectral features can be influenced by local environment and two pixels from one class may have different spectral signatures in far-away space, but their main properties are believed to be similar. As we will see later, there are principal directions in a superpixel, which correspond to the main properties of the superpixel, and the directions of SPCRs transformed from the original features are closer to these principal directions. This means that the main properties of the pixels can be embodied with their SPCRs.

2) Between-superpixel scatter matrix of SPCR is defined in the first time for penalty graph in analogy with between-class scatter matrix. This helps to reduce the differences between SPCRs of pixels from one superpixel and increase the differences between SPCRs of pixels from different superpixels simultaneously. Someone may concern that it may enlarge the distance between the pixels that belong to one class but belong to different superpixels. However, if most of pixels in two superpixels belong to one class, SPCRs in two superpixels will be similar to each other, and thus the scatter matrix between them is small and has little contribution to the calculation of global low-dimensional projections.

3) The proposed two stages of feature extraction are novel. The first stage is to replace the original features with their SPCRs. SPCRs can keep the main properties of spectral features and discard unimportant information. This effect is similar to weighted mean filter, as it is used as a preprocessing tool in some DR methods [42]. In the second stage, a global linear projection matrix is obtained by reducing the discrepancy between the original features and their SPCRs and the distance between

pixels from one superpixel, and enlarging the distance between pixels from different superpixels that are not close in terms of SPCRs. In this way, the problem of spectral redundancy and Hughes phenomenon, and the problem of intraclass difference and interclass similarity can be alleviated simultaneously.

The remainder of this article is organized as follows. Section II briefly reviews some related works including superpixel segmentation and BCGDA. Section III explicitly explains the proposed SPCRG method. Experimental results are displayed in Section IV to demonstrate the superior performance of SPCRG. In Section V, we conclude the article.

II. RELATED WORKS

A. Superpixel Segmentation

Superpixel segmentation divides an image as perceptually disconnected homogeneous regions, each one of which is called a superpixel within which pixels share similar properties. Taking account of the merit of superpixel segmentation, it has been used widely in the preprocessing stage of HSI classification. Two popular superpixel segmentation algorithms are simple linear iterative clustering (SLIC) [39], [40], [43] and entropy rate superpixel (ERS) [37], [41], [44], [45], because of their low computational complexity and satisfactory performance. In this article, since ERS is empirically better than SLIC in terms of classification performance, we use ERS as superpixel segmentation algorithm, which is briefly reviewed here. In fact, the superpixel segmentation algorithm in the first step of our proposed DR method is not limited to ERS, and other alternative superpixel segmentation algorithms can be also considered [46].

ERS segmentation is one of graph-based superpixel segmentation algorithms. First, a graph is constructed by building edges between a pixel and its four or eight spatially neighboring pixels and calculating the pairwise similarities. Then, ERS segmentation is to choose a subset of the edges so that many disconnected regions are generated. To generate K nonoverlapping homogeneous regions, the objective function of ERS segmentation combines an entropy rate term $\mathcal{H}(E_s)$ with a balancing term $\mathcal{B}(E_s)$ with respect to the subset E_s of edge set E

$$\begin{aligned} & \max_{E_s} \mathcal{H}(E_s) + \gamma \mathcal{B}(E_s) \\ & \text{s.t. } N_{E_s} \geq K \text{ and } E_s \subseteq E \end{aligned} \quad (1)$$

where γ is the weight that balances the entropy rate term and the balancing term, and N_{E_s} is the number of nonoverlapping homogeneous regions. The entropy rate term is to obtain compact and homogeneous regions. The balancing term favors fewer and similar-size homogeneous regions. To solve the optimization problem (1), a greedy algorithm is adopted. In each iteration of the algorithm, one edge is added, yielding the largest gain of the objective function. The iterations are stopped when the number of homogeneous regions is equal to K . As described in [44], this method has achieved good performance and is also highly efficient as it only takes about 2.5 s on average to segment an image of size 481×321 .

B. BCGDA for HSI Dimensionality Reduction

HSI is a 3-D tensor $\mathbf{H} \in \mathbb{R}^{m \times n \times d}$, where m and n are the height and width, respectively, and d is the number of spectral bands. For pixelwise hyperspectral classification, we transform the 3-D image into pixelwise 2-D matrix $\mathbf{X} = [\mathbf{x}_i] \in \mathbb{R}^{d \times mn}$ with $\mathbf{x}_i \in \mathbb{R}^{d \times 1}$. In the framework of graph-embedding DR, two graphs are constructed: an intrinsic graph $\mathbf{G} = \{\mathbf{X}, \mathbf{W}\}$ and a penalty graph $\mathbf{G}^p = \{\mathbf{X}, \mathbf{W}^p\}$, where \mathbf{W} and \mathbf{W}^p are the corresponding weight matrices. To obtain a linear projection matrix $\mathbf{P} \in \mathbb{R}^{d \times k}$ ($k \ll d$) that transforms the original high-dimensional features \mathbf{X} into a low-dimensional subspace $\mathbf{P}^T \mathbf{X}$, an objective function is mathematically formed as

$$\mathbf{P} = \arg \min_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{P})}{\text{Tr}(\mathbf{P}^T \mathbf{X} \mathbf{L}^p \mathbf{X}^T \mathbf{P})} \quad (2)$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ and $\mathbf{L}^p = \mathbf{D}^p - \mathbf{W}^p$ are the Laplacian matrices of graphs \mathbf{G} and \mathbf{G}^p , respectively, with \mathbf{D} and \mathbf{D}^p being diagonal matrices subject to $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ and $\mathbf{D}_{ii}^p = \sum_j \mathbf{W}_{ij}^p$, and $\text{Tr}(\cdot)$ is the trace operator. This optimization problem can be solved as a generalized eigenvalue decomposition

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{p}_i = \lambda_i \mathbf{X} \mathbf{L}^p \mathbf{X}^T \mathbf{p}_i \quad (3)$$

where λ_i is the i th smallest nonzero eigenvalue and \mathbf{p}_i is the corresponding eigenvector. The $d \times k$ projection matrix \mathbf{P} is obtained just by grouping $[\mathbf{p}_i] (1 \leq i \leq k)$.

BCGDA is established under the above graph-embedding framework. The main work of BCGDA is to construct its intrinsic graph. Let C be the number of classes, the training number be N_L , and the number of training samples in the i th class be n_i satisfying $\sum_{i=1}^C n_i = N_L$. First, training samples are grouped as matrices \mathbf{X}_i where every column vector \mathbf{x}_i in \mathbf{X}_i belongs to the i th class. Then, CR, i.e., using all training samples in the i th class to represent a query training sample \mathbf{x}_i is obtained by solving the optimization problem

$$\arg \min_{\alpha_i} \|\mathbf{x}_i - \mathbf{X}_i \alpha_i\|_2^2 + \lambda \|\alpha_i\|_2^2 \quad (4)$$

where $\mathbf{X}_i \alpha_i$ is the CR of \mathbf{x}_i , α_i is the corresponding CR coefficient, and λ is the weight balancing two terms. It is noted that \mathbf{x}_i is excluded in \mathbf{X}_i in (4). The solution to (4) is

$$\alpha_i = (\mathbf{X}_i^T \mathbf{X}_i + \lambda \mathbf{I}_i)^{-1} \mathbf{X}_i^T \mathbf{x}_i. \quad (5)$$

After α_i is determined for all the training samples, the CR coefficient matrix \mathbf{A} can be constructed by zero-padding α_i in the corresponding positions and then grouping them together

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{A}_C \end{bmatrix} \quad (6)$$

where \mathbf{A}_i is the CR coefficient matrix of size n_i in the i th class. Laplacian matrix of the intrinsic graph is $\mathbf{L} = (\mathbf{I} - \mathbf{A})(\mathbf{I} - \mathbf{A})^T$; thus, within-class representation error matrix is

$$\mathbf{S}_w = \mathbf{X} \mathbf{L} \mathbf{X}^T = \sum_i^C \mathbf{X}_i (\mathbf{I}_i - \mathbf{A}_i) (\mathbf{I}_i - \mathbf{A}_i)^T \mathbf{X}_i^T \quad (7)$$

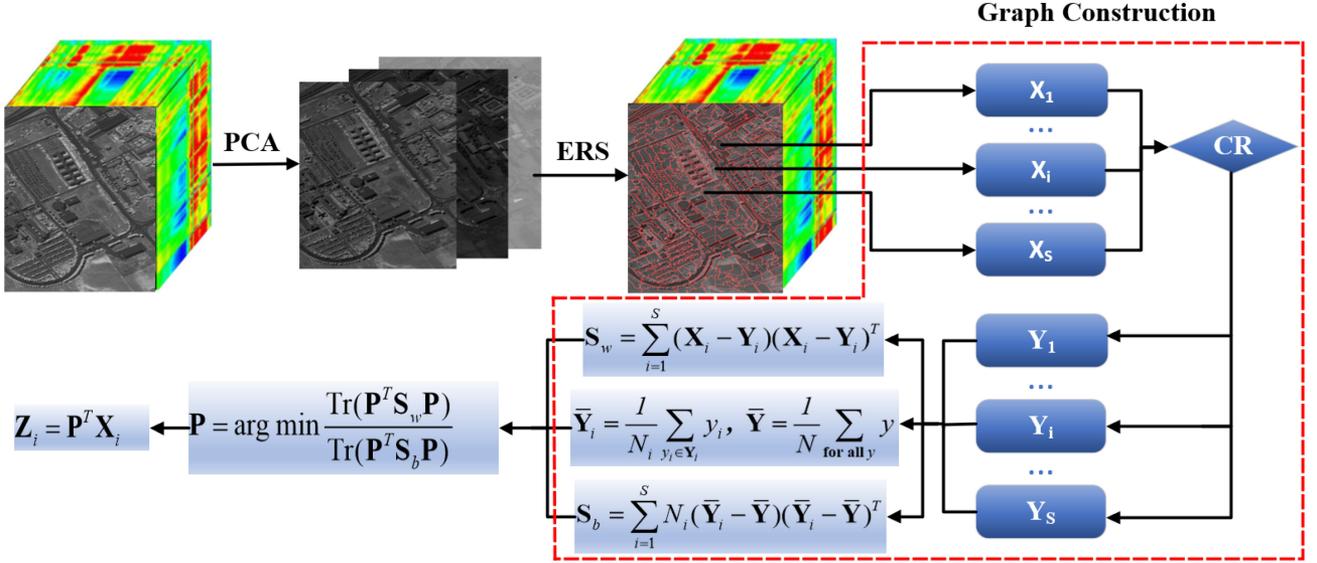


Fig. 1. Flowchart of the proposed SPCRGE for unsupervised dimensionality reduction on HSI.

where \mathbf{I}_i is an identity matrix of size n_i . The Laplacian matrix of penalty graph \mathbf{L}_p is set as \mathbf{I} , which is an identity matrix of size n . Then, projection matrix \mathbf{P} can be obtained by solving the eigenvalue decomposition problem (3). As described in [29], BCGDA has achieved better HSI classification performance than most of state-of-the-art methods.

III. PROPOSED UNSUPERVISED SPCRGE FRAMEWORK

In this section, the flowchart of SPCRGE is shown in Fig. 1, including superpixel segmentation and superpixelwise CR graph embedding.

A. Superpixel Segmentation for HSI

HSI classification is based on discriminative spectral information. However, spatial information that neighboring pixels bear great chances to be from one class should be utilized as well. Many works incorporate spatial information by a specified neighborhood system. This may not be appropriate since pixels within one class distribute around spatial space in a more adaptive way and constitute homogeneous regions. Therefore, it is better to form an adaptive neighborhood system via generating nonoverlapping homogeneous regions. ERS segmentation is one excellent method for doing so.

Before performing ERS segmentation algorithm, PCA is executed on HSI to extract the main spatial structure information. ERS segmentation algorithm is then executed on the first three principal components to generate nonoverlapping homogeneous regions—superpixels. Let the number of superpixels be S , the total number of pixels be N , and the number of pixels in the i th superpixel be N_i satisfying $\sum_{i=1}^S N_i = N$. The pixels in the i th superpixel are grouped as \mathbf{X}_i .

B. Superpixelwise CR Graph Embedding for HSI Dimensionality Reduction

After superpixels are generated via ERS segmentation algorithm, SPCR coefficient matrix \mathbf{A}_i in the i th superpixel are obtained by solving the optimization problem:

$$\arg \min_{\mathbf{A}_i} \|\mathbf{X}_i - \mathbf{X}_i \mathbf{A}_i\|_F^2 + \lambda \|\mathbf{A}_i\|_F^2 + \beta \text{Tr}(\mathbf{X}_i \mathbf{A}_i \mathbf{M}_i \mathbf{A}_i^T \mathbf{X}_i^T) \quad (8)$$

where \mathbf{M}_i is the symmetrical Laplacian matrix of graph matrix \mathbf{W}_i in the i th superpixel whose m th element is $\mathbf{W}_{i(m,n)} = \exp(-\|\mathbf{x}_m - \mathbf{x}_n\|_2^2 / \gamma_m \gamma_n)$, where $\gamma_m = \|\mathbf{x}_m - \mathbf{x}_m^{NN}\|_2$ and \mathbf{x}_m^{NN} is the nearest neighbor of \mathbf{x}_m in the i th superpixel. Similar to LapCGDA [47], Laplacian regularization, i.e., the last term in (8) is to preserve the local manifold. However, LapCGDA preserves the local manifold by conforming CR coefficients of each pixel to the local manifold while our method by conforming SPCR per se to the local manifold.

From the minimization problem (8), we can obtain the corresponding analytical solution by solving a generalized Sylvester equation. By setting the derivative of objective function (8) with respect to \mathbf{X}_i to zero, we can obtain

$$(\mathbf{X}_i^T \mathbf{X}_i + \lambda \mathbf{I}_i) \mathbf{A}_i + \beta \mathbf{X}_i^T \mathbf{X}_i \mathbf{A}_i \mathbf{M}_i = \mathbf{X}_i^T \mathbf{X}_i. \quad (9)$$

Substituting $\mathbf{X}_i^T \mathbf{X}_i$ with its eigenvalue decomposition $\mathbf{Q}_1 \mathbf{\Lambda}_1 \mathbf{Q}_1^T$ and \mathbf{M}_i with its eigenvalue decomposition $\mathbf{Q}_2 \mathbf{\Lambda}_2 \mathbf{Q}_2^T$, (9) becomes

$$(\mathbf{\Lambda}_1 + \lambda \mathbf{I}_i) \mathbf{Q}_1^T \mathbf{A}_i \mathbf{Q}_2 + \beta \mathbf{\Lambda}_1 \mathbf{Q}_1^T \mathbf{A}_i \mathbf{Q}_2 \mathbf{\Lambda}_2 = \mathbf{\Lambda}_1 \mathbf{Q}_1^T \mathbf{Q}_2 \quad (10)$$

which is a generalized Sylvester equation. From (10), we can obtain

$$\text{vec}(\mathbf{Q}_1^T \mathbf{A}_i \mathbf{Q}_2) = \text{vec}(\mathbf{\Lambda}_2 \mathbf{Q}_1^T \mathbf{Q}_2) / [\mathbf{1} \otimes \text{diag}(\mathbf{\Lambda}_1 + \lambda \mathbf{I}_i) + \beta \text{diag}(\mathbf{\Lambda}_2) \otimes \text{diag}(\mathbf{\Lambda}_1)] \quad (11)$$

where $\text{vec}(\cdot)$, $\text{diag}(\cdot)$ are the vectorization operator and the operator of vectorizing the diagonal elements of a matrix, respectively, \otimes is Kronecker product operator, and $\mathbf{1}$ is a vector of all ones with the same dimension as $\text{diag}(\Lambda_1)$. Then, \mathbf{A}_i can be obtained after $\mathbf{Q}_1^T \mathbf{A}_i \mathbf{Q}_2$ is retrieved from its vector form.

After all \mathbf{A}_i are obtained, like (6), SPCR coefficient matrix \mathbf{A} can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{A}_S \end{bmatrix} \quad (12)$$

where \mathbf{A}_i is the SPCR coefficient matrix of size N_i in the i th superpixel. Then, within-superpixel representation error matrix of SPCRGE is defined as

$$\begin{aligned} \mathbf{S}_w &= \sum_{i=1}^S \mathbf{X}_i (\mathbf{I}_i - \mathbf{A}_i) (\mathbf{I}_i - \mathbf{A}_i)^T \mathbf{X}_i^T \\ &= \sum_{i=1}^S (\mathbf{X}_i - \mathbf{Y}_i) (\mathbf{X}_i - \mathbf{Y}_i)^T \end{aligned} \quad (13)$$

where \mathbf{I}_i is an identity matrix of size N_i , and $\mathbf{Y}_i = \mathbf{X}_i \mathbf{A}_i$. Seemingly, representation error matrix of SPCRGE is similar to that of BCGDA, and except for Laplacian regularization, the only difference is that the former is superpixel-dependent while the latter is class-dependent. However, to obtain low-dimensional features, a projection matrix in BCGDA is ultimately performed on the original spectral features \mathbf{X} while that in SPCRGE is ultimately performed on SPCR $\mathbf{Y} = \mathbf{X}\mathbf{A}$. Therefore, physical meanings of representation error matrices of SPCRGE and BCGDA are different. In BCGDA, a projection matrix is found to preserve CR relationships. However, in SPCRGE, a projection matrix is found to reduce the discrepancy between SPCR and the original spectral features. In analogy with between-class scatter matrix, between-superpixel scatter matrix of SPCRGE is defined as

$$\mathbf{S}_b = \sum_{i=1}^S N_i (\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}}) (\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})^T \quad (14)$$

where $\bar{\mathbf{Y}}$ and $\bar{\mathbf{Y}}_i$ are the mean of all SPCRs and SPCRs in the i th superpixel, respectively, and S is the number of superpixels. Then, projection matrix is obtained by solving the optimization problem:

$$\mathbf{P} = \arg \min_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T \mathbf{S}_w \mathbf{P})}{\text{Tr}(\mathbf{P}^T \mathbf{S}_b \mathbf{P})} \quad (15)$$

and its corresponding generalization eigenvalue problem is

$$\mathbf{S}_w \mathbf{p}_i = \lambda_i \mathbf{S}_b \mathbf{p}_i. \quad (16)$$

Similar to (3), \mathbf{P} is obtained by grouping $[\mathbf{p}_i] (1 \leq i \leq k)$ corresponding to the k smallest nonzero eigenvalues. The overall description of the proposed SPCRGE is given as Algorithm 1.

C. Analysis of Proposed SPCRGE

In this section, some physical explanations are given to illustrate the rationality behind the proposed SPCRGE.

Algorithm 1: Proposed SPCRGE Algorithm.

Input: HSI data $\mathbf{X} = \{\mathbf{x}_j\}_{j=1}^N \in \mathbb{R}^d$, number of superpixels S , and the regularization parameters λ and β
 Obtain first three principal components by PCA on \mathbf{X}
 and segment on the three principal components via ERS segmentation algorithm to obtain superpixels;
for $i = 1$ to S
 Grouping pixels of the i th superpixel as
 $\mathbf{X}_i = [\mathbf{x}_i] \in \mathbb{R}^{d \times N_i}$;
 Obtain its Laplacian Matrix \mathbf{M}_i via pairwise distance calculation;
 Eigenvalue decompositions: $\mathbf{X}_i^T \mathbf{X}_i = \mathbf{Q}_{i1} \Lambda_{i1} \mathbf{Q}_{i1}^T$,
 $\mathbf{M}_i = \mathbf{Q}_{i2} \Lambda_{i2} \mathbf{Q}_{i2}^T$;
 Obtain $\text{vec}(\mathbf{Q}_{i1}^T \mathbf{A}_i \mathbf{Q}_{i2})$, i.e., the vector form of $\mathbf{Q}_{i1}^T \mathbf{A}_i \mathbf{Q}_{i2}$ via solving (11) in a closed form;
 Retrieve SPCR coefficient matrix \mathbf{A}_i from $\mathbf{Q}_{i1}^T \mathbf{A}_i \mathbf{Q}_{i2}$, i.e., the matrix form of $\text{vec}(\mathbf{Q}_{i1}^T \mathbf{A}_i \mathbf{Q}_{i2})$ and obtain $\mathbf{Y}_i = \mathbf{X}_i \mathbf{A}_i$;
endfor;
 Obtain within-superpixel representation error matrix \mathbf{S}_w via (13) and between-superpixel scatter matrix \mathbf{S}_b via (14);
 Obtain projection matrix \mathbf{P} via solving (16) and group all \mathbf{Y}_i as \mathbf{Y} ;
Output: A projection matrix \mathbf{P} and low-dimensional features $\mathbf{P}^T \mathbf{Y}$.

1) *Remark 1:* Performing singular value decomposition (SVD) on $\mathbf{X}_i = \mathbf{U} \Sigma \mathbf{V}^T$, where Σ is a diagonal matrix whose diagonal elements are the singular values $\sigma_i > 0$ in descent order. Then, according to (5), expression of SPCR, \mathbf{y}_i of \mathbf{x}_i in the i th superpixel can be written as [48]

$$\mathbf{y}_i = \mathbf{X}_i \alpha_i = \mathbf{U} \hat{\Sigma} \mathbf{U}^T \mathbf{x}_i \quad (17)$$

where $\hat{\Sigma}$ is a diagonal matrix whose diagonal elements are $\hat{\sigma}_j = \sigma_j^2 / (\sigma_j^2 + \lambda)$. From (17), some points with respect to \mathbf{y}_i can be obtained. a) The original spectral feature \mathbf{x}_i is first projected into the space \mathbf{U} spanned by \mathbf{X}_i with the projection value reduced by the ratio of $1 - \hat{\sigma}_j$ in the j th projection direction, and then \mathbf{y}_i is obtained by multiplying \mathbf{U} with a vector whose elements are the corresponding reduced projection values. b) The reduction ratio in every direction is proportional to λ and the projection values in the perpendicular direction to \mathbf{U} are lost. The reduction ratio is zero when $\lambda = 0$ and $1 - \sigma_i^2 / \lambda$ when λ is large. c) The reduction ratio decreases with the increase of the singular value in that projection direction, so the projection values in the principal directions corresponding to the larger singular values are relatively larger than that in other directions.

From the above three points, we can conclude that the direction of SPCR \mathbf{y}_i in one superpixel transformed from the original feature \mathbf{x}_i is closer to the principal directions of the space spanned by all pixels in the same superpixel irrespective of shorter length of \mathbf{y}_i compared to \mathbf{x}_i . In addition, the larger the regularization parameter λ , the closer the direction of \mathbf{y}_i is to the principal directions. The principal directions corresponding to

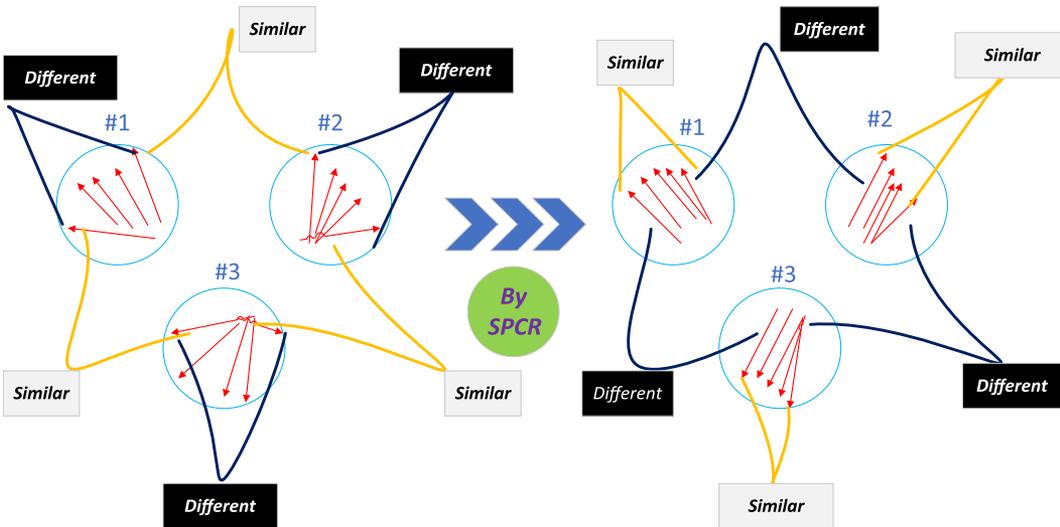


Fig. 2. Illustration of the effect of SPCR for reducing intraclass difference and suppressing interclass similarity.

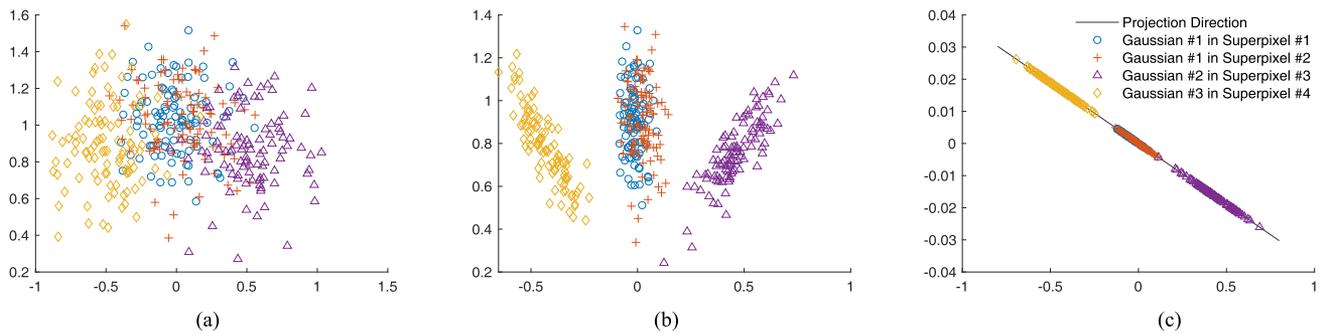


Fig. 3. Effect of SPCR of enhancing intraclass similarity and suppressing interclass similarity on the three simulated Gaussian distributions. (a) Original points. (b) SPCRs of the original points. (c) The points in the low-dimensional space.

the larger singular values can be seen as commonality of pixels in a superpixel, and the other directions corresponding to the smaller singular values as individuality. By SPCR, commonality of pixels in the superpixel is enhanced. This is similar to PCA as it also keeps the main properties of the original features, but the difference lies in that individuality is also kept in SPCR, though with less ratio.

This property can be utilized to alleviate the problem of intraclass difference and interclass similarity in HSI classification, which can be illustrated in Fig. 2. The superpixels #1, #2, and #3 represented by circles are obtained by grouping spectral-similar and spatially adjacent pixels represented by red arrows, based on the assumption that spatial-spectral similar pixels are highly likely to be one class. By the original features, large difference exists between pixels from one superpixel and large similarity also exists between pixels from different superpixels, which correspond to intraclass difference and interclass similarity, respectively. By SPCR, commonality of pixels from one superpixel is enhanced and their directions are more closer to the principal directions. As a result, similarity of pixels from one superpixel are enhanced while similarity of pixels from different superpixels are suppressed.

However, discrepancy exists between the original spectral features and SPCRs. In (8), Laplacian regularization is added to conform SPCRs to the local manifold, so discrepancy is alleviated to some extent. Furthermore, the purpose in the intrinsic graph of SPCRGE is also to find a projection space where discrepancy can be highly reduced.

2) *Remark 2:* For penalty graph, between-superpixel scatter matrix is utilized. This is enlightened by between-class scatter matrix. As demonstrated in [25], benefits of between-class scatter matrix are twofolded: enlarging the distances of pixels between different classes and reducing the distances of pixels from one class. Similarly, utilization of between-superpixel scatter matrix is to further enhance the similarity between pixels from one superpixel and suppress the similarity between pixels from different superpixels.

It is noted that the spectral features of two pixels from one class but from different superpixels are similar irrespective of intraclass difference. By SPCR, their features are more similar to each other as compared to their original features since SPCR extracts the principal properties of two superpixels and the principal properties of two pixels from one class should be the same. This is illustrated in Fig. 3, where the pixels of

TABLE I
CLASS LABELS AND TRAIN-TEST DISTRIBUTION OF SAMPLES FOR PAVIAU

No.	Class Name	Train	Test
1	Asphalt	30	6601
2	Meadows	30	18619
3	Gravel	30	2069
4	Trees	30	3034
5	Painted metal sheets	30	1315
6	Bare Soil	30	4999
7	Bitumen	30	1300
8	Self-Blocking bricks	30	3652
9	Shadows	30	917
Total		270	42506

TABLE II
CLASS LABELS AND TRAIN-TEST DISTRIBUTION OF SAMPLES FOR MUFFL

No.	Class Name	Train	Test
1	Trees	150	23096
2	Mostly grass	150	4120
3	Mixed ground surface	150	6732
4	Dirt and sand	150	1676
5	Road	150	6537
6	Water	150	316
7	Building shadow	150	2083
8	Building	150	6090
9	Sidewalk	150	1235
10	Yellow curb	150	33
11	Cloth panels	150	119
Total		1650	52037

TABLE III
CLASS LABELS AND TRAIN-TEST DISTRIBUTION OF SAMPLES FOR HOUSTON2013

No.	Class Name	Train	Test
1	Grass healthy	198	1053
2	Grass stressed	190	1064
3	Grass synthetics	192	505
4	Tree	188	1056
5	Soil	186	1056
6	Water	182	143
7	Residential	196	1072
8	Commercial	191	1053
9	Road	193	1059
10	Highway	191	1036
11	Railway	181	1054
12	Parking lot 1	192	1041
13	Parking lot 2	184	285
14	Tennis court	181	247
15	Running track	187	473
Total		2832	12197

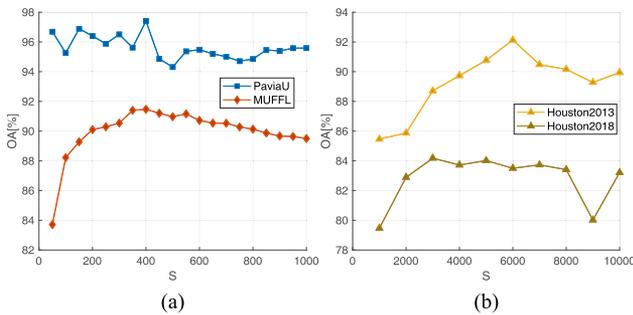


Fig. 4. OA [%] versus S , i.e., the number of superpixels, using the proposed SPCRGE on the HSI datasets. (a) PAVIAU and MUFFL. (b) Houston2013 and Houston2018.

TABLE IV
CLASS LABELS AND TRAIN-TEST DISTRIBUTION OF SAMPLES FOR HOUSTON2018

No.	Class Name	Train	Test
1	Healthy grass	300	9799
2	Stressed grass	300	32502
3	Artificial turf	300	684
4	Evergreen trees	300	13595
5	Deciduous trees	300	5021
6	Bare earth	300	4516
7	Water	133	266
8	Residential buildings	300	39772
9	Non-residential buildings	300	223752
10	Roads	300	45866
11	Sidewalks	300	34029
12	Crosswalks	300	1518
13	Major thoroughfares	300	46348
14	Highways	300	9865
15	Railways	300	6937
16	Paved parking lots	300	11500
17	Unpaved parking lots	68	146
18	Cars	300	6547
19	Trains	300	5369
20	Stadium seats	300	6824
Total		5601	504656

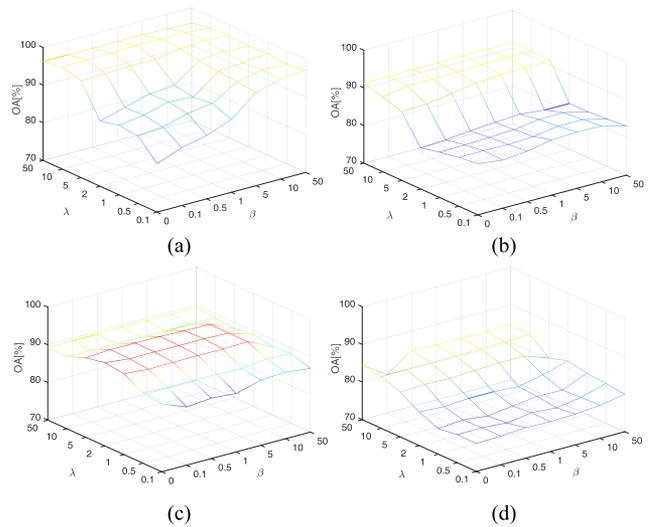


Fig. 5. OA [%] versus parameters λ and β by several DR methods on the HSI datasets. (a) PAVIAU. (b) MUFFL. (c) Houston2013. (d) Houston2018.

yellow diamond, blue circle, and purple triangle are generated by different Gaussian distributions and the pixels of blue circle and red plus sign are from the same Gaussian distributions. Let the four types of signed pixels be from different superpixels. By the original features, Gaussian #1 in superpixels #1 and #2 cannot be distinguished well from Gaussian #2 in superpixel #3 and Gaussian #3 in superpixel #4. By SPCR, they can be well distinguished as shown in Fig 3(b). As SPCRs are used to generate between-superpixel scatter matrix, the distances between pixels from one class but different superpixels are small. As a result, the between-superpixel scatter matrices corresponding to these distances are small as well. Therefore, enlarging the distance between superpixels from one class is limited and can be neglected, as can be seen from Fig. 3(c) where Gaussian #1 in superpixels #1 and #2 are clustered together well.

TABLE V
CLASSIFICATION ACCURACIES [%] OF DIFFERENT DR METHODS ON PAVIAU WITH 30 RANDOMLY CHOSEN TRAINING SAMPLES PER CLASS

Method	Raw	PCA	LDA	CCPGE	BCGDA	OTVCA	SPLDA	SuperPCA	TSLGDA	SPCRGE
Class1	70.35±4.68	74.41±3.19	59.26±7.02	76.75±3.53	70.38±3.03	89.91±2.77	77.44±3.42	90.44±2.81	78.71±4.51	94.88±4.20
Class2	65.40±8.61	68.31±5.49	70.01±5.29	82.99±5.61	70.25±6.47	88.78±3.59	81.16±5.44	91.16±4.28	86.74±4.97	98.25±1.11
Class3	74.03±4.93	81.83±6.18	60.45±4.64	75.74±5.23	68.92±6.46	93.74±2.68	74.35±4.50	93.18±2.02	83.01±4.04	99.05±2.01
Class4	95.35±1.56	89.39±3.72	88.85±4.13	94.76±2.45	94.98±1.93	91.93±3.30	91.39±3.41	72.96±6.93	92.30±1.84	96.25±2.64
Class5	99.32±0.42	99.70±0.83	98.02±1.52	98.17±0.79	98.86±0.49	98.62±1.30	98.48±0.78	95.05±3.17	99.96±0.10	99.11±0.56
Class6	70.75±6.51	79.56±4.80	65.24±5.74	76.94±4.68	73.31±5.22	96.45±2.61	81.04±4.82	91.35±3.96	86.21±4.73	97.75±1.74
Class7	86.62±2.87	91.54±3.44	58.19±9.91	79.85±4.31	86.74±5.90	97.68±1.49	90.52±1.91	99.22±0.41	95.98±2.23	98.73±3.32
Class8	75.12±6.52	59.58±6.45	51.88±5.95	65.91±6.14	73.05±5.31	91.00±4.54	70.19±5.48	96.06±1.48	72.79±6.05	99.03±0.80
Class9	96.27±1.02	99.78±0.33	98.47±1.14	99.99±0.06	99.80±0.11	95.91±1.63	99.99±0.02	99.80±3.28	95.54±1.92	96.86±2.19
OA	72.56±3.48	74.35±2.57	68.22±3.05	81.07±2.24	74.60±2.71	91.25±1.71	81.25±2.43	90.84±2.17	85.27±1.95	97.64±1.00
AA	81.47±0.95	82.68±1.04	72.26±1.61	83.46±0.66	81.81±1.24	93.78±0.93	84.95±0.98	92.14±2.17	87.58±0.83	97.77±0.74
kappa	65.70±3.76	67.72±2.96	59.91±3.40	75.00±2.51	67.92±3.05	89.00±2.14	75.90±2.83	88.06±2.72	80.85±0.02	97.00±1.31

Results are averaged by 20 trials with standard deviations listed as well.

TABLE VI
CLASSIFICATION ACCURACIES [%] OF DIFFERENT DR METHODS ON MUFFL WITH 150 RANDOMLY CHOSEN SAMPLES PER CLASS

Method	Raw	PCA	LDA	CCPGE	BCGDA	OTVCA	SPLDA	SuperPCA	TSLGDA	SPCRGE
Class1	83.84±1.32	82.61±1.69	84.14±1.23	85.92±1.33	83.81±1.84	87.05±1.81	85.43±1.34	78.28±3.44	87.63±1.36	94.35±1.16
Class2	83.78±1.78	83.05±2.27	80.28±1.76	77.67±2.26	81.88±2.63	89.91±1.98	85.07±1.62	91.77±1.62	81.69±2.58	91.80±2.05
Class3	69.84±1.43	69.15±2.55	72.08±1.92	72.67±2.56	72.94±2.62	79.29±2.21	71.86±1.61	88.38±2.21	74.89±2.49	87.27±2.20
Class4	85.78±1.77	85.43±1.78	85.46±1.35	86.28±1.25	88.22±1.13	92.71±2.17	89.45±0.94	95.11±1.19	91.72±1.96	95.49±1.40
Class5	84.59±1.06	87.59±1.33	76.55±2.71	87.12±1.45	87.72±1.47	88.68±1.41	85.98±1.03	86.96±1.68	82.02±1.79	92.35±0.96
Class6	93.58±1.57	94.67±1.13	91.84±1.85	94.30±1.45	94.79±1.10	99.26±1.00	96.63±0.99	97.58±1.55	99.73±0.44	94.78±2.04
Class7	79.98±2.53	88.90±1.72	81.71±2.28	85.21±1.65	88.93±1.23	91.84±1.47	81.91±2.14	90.14±1.91	88.98±2.71	88.92±1.79
Class8	82.61±1.48	80.47±1.47	75.50±1.87	83.04±1.22	79.09±1.93	88.22±1.10	84.75±1.67	86.33±1.26	92.19±1.45	87.45±1.57
Class9	73.73±2.08	70.88±1.98	70.40±1.82	78.54±1.99	74.32±1.98	65.64±3.40	77.70±1.81	82.53±2.50	69.07±3.20	79.94±1.71
Class10	96.21±3.30	96.06±3.47	98.18±2.23	93.94±2.76	98.18±2.01	89.85±5.00	98.03±1.98	94.39±3.87	90.30±4.85	89.70±4.22
Class11	98.66±0.86	98.24±1.27	98.19±1.23	98.32±0.80	98.45±1.75	97.39±1.59	99.20±1.20	96.26±1.79	99.16±0.70	97.06±1.26
OA	81.74±0.61	81.46±0.75	80.02±0.58	83.26±0.51	82.42±0.80	86.58±0.82	83.55±0.65	83.97±1.45	85.19±0.65	91.66±0.59
AA	84.78±0.54	85.19±0.37	83.12±0.44	85.73±0.37	86.21±0.37	88.17±0.65	86.91±0.41	89.79±0.52	87.03±0.73	90.83±0.54
kappa	76.43±0.72	76.16±0.88	74.19±0.69	78.27±0.60	77.32±0.94	82.59±1.00	78.69±0.78	79.33±1.74	80.78±0.79	88.84±0.78

Results are averaged by 20 trials with standard deviations listed as well.

TABLE VII
CLASSIFICATION ACCURACIES [%] OF DIFFERENT DR METHODS ON HOUSTON2013 WITH SPECIFIED TRAINING SAMPLES

Method	Raw	PCA	LDA	CCPGE	BCGDA	OTVCA	SPLDA	SuperPCA	TSLGDA	SPCRGE
Class1	83.10	83.10	81.77	81.39	82.81	80.72	83.00	80.91	83.10	83.10
Class2	82.52	83.65	95.96	83.27	92.67	79.04	82.71	67.67	95.21	81.02
Class3	99.60	99.60	100.00	100.00	100.00	99.21	99.80	100.00	100.00	100.00
Class4	97.63	97.54	97.06	90.25	98.58	84.75	92.14	49.43	96.31	98.01
Class5	98.77	99.15	98.11	98.39	99.34	98.20	99.34	99.91	100.00	100.00
Class6	99.30	99.30	98.60	95.80	99.30	94.41	99.30	79.02	99.30	97.90
Class7	89.27	85.07	79.20	76.31	81.72	89.93	84.42	52.43	86.47	92.63
Class8	64.67	45.39	64.10	54.32	63.82	49.76	80.06	85.85	92.50	80.72
Class9	66.67	72.80	67.61	69.41	67.61	85.08	76.58	47.03	76.86	88.57
Class10	56.56	56.95	72.20	69.69	74.32	67.37	70.37	61.78	38.03	99.61
Class11	74.76	80.46	55.41	92.50	74.00	74.10	80.27	74.10	74.00	94.97
Class12	79.15	79.44	53.31	74.06	76.66	80.02	81.75	85.40	84.73	97.89
Class13	75.44	75.09	64.56	68.07	74.04	58.25	73.33	57.19	79.65	75.44
Class14	99.60	99.60	99.19	98.38	99.60	100.00	99.60	100.00	100.00	98.38
Class15	97.67	98.10	98.31	98.31	98.73	97.25	98.73	100.00	93.23	100.00
OA	81.47	80.65	78.79	80.94	83.06	80.62	84.68	73.17	84.40	92.13
AA	84.31	83.68	81.69	83.34	85.55	82.54	86.76	76.05	86.63	92.55
kappa	79.92	79.06	76.98	79.31	81.62	79.13	83.38	70.89	83.00	91.45

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the proposed SPCRGE is utilized to validate the performance. First, four urban HSI datasets are introduced, including PaviaU, MUFFL, Houston2013, and Houston2018. Second, the impact of the number of superpixels, two tunable parameters λ and β , and reduced dimension of SPCRGE on the four datasets is quantified by the classical SVM classifier.

Finally, based on the SVM classifier, experiments of SPCRGE on the four HSI datasets are carried out in comparison with the traditional methods, PCA [17], LDA [18], and other state-of-the-art methods, CCPGE [30], BCGDA [29], OTVCA [35], SPLDA [40], SuperPCA [41], and TSLGDA [34]. Among these DR methods, PCA, CCPGE, OTVCA, SPLDA, SuperPCA, and the proposed SPCRGE are unsupervised without using any label

TABLE VIII
CLASSIFICATION ACCURACIES [%] OF DIFFERENT DR METHODS ON HOUSTON2018 WITH 300 RANDOMLY CHOSEN SAMPLES PER CLASS

Method	Raw	PCA	LDA	CCPGE	BCGDA	OTVCA	SPLDA	SuperPCA	TSLGDA	SPCRGE
Class1	99.47±0.18	98.98±0.41	96.84±0.76	97.75±0.48	97.60±0.57	87.49±2.57	95.84±1.22	88.57±1.12	92.56±1.04	90.20±2.30
Class2	94.10±0.62	93.07±0.82	90.49±0.64	91.45±0.96	92.16±0.87	77.84±2.32	88.84±1.39	87.90±1.62	85.54±1.33	87.01±2.36
Class3	100.00±0	100.00±0	100.00±0	100.00±0	100.00±0	99.90±0.13	100.00±0	100.00±0	100.00±0	100.00±0
Class4	96.95±0.52	95.61±0.75	96.84±0.45	96.61±0.41	96.63±0.49	95.65±0.59	94.83±0.99	96.24±1.11	96.94±0.39	95.28±1.46
Class5	91.22±1.06	87.69±1.48	89.43±0.80	86.58±1.56	89.37±0.85	81.63±0.77	88.36±1.00	95.61±0.98	88.68±1.33	94.99±1.13
Class6	98.29±0.32	97.84±0.33	97.46±1.85	97.19±1.45	96.95±1.10	99.90±1.00	97.19±0.99	99.96±1.55	99.30±0.44	99.92±0.17
Class7	99.29±0.68	99.32±0.53	99.29±0.46	99.66±0.35	99.44±0.56	99.55±0.47	99.85±0.25	99.17±0.82	99.81±0.19	99.40±0.66
Class8	75.77±1.52	70.88±2.08	71.57±1.05	71.96±1.86	72.46±1.87	87.63±0.61	74.12±1.72	88.47±1.21	81.62±0.85	93.93±1.05
Class9	68.48±1.65	62.68±2.38	62.41±1.27	58.86±1.97	59.04±2.16	82.53±0.75	69.52±1.50	75.32±1.69	71.60±3.01	89.29±1.48
Class10	42.69±1.47	39.09±1.58	37.42±1.64	35.42±1.88	35.70±2.11	39.69±1.48	38.44±1.34	57.30±2.21	37.10±1.85	61.76±1.63
Class11	57.01±2.22	53.36±1.75	49.10±2.09	52.04±1.50	52.27±1.50	39.69±1.43	51.53±1.47	51.32±1.83	46.02±1.44	58.18±1.89
Class12	65.01±1.48	59.87±1.15	53.95±1.04	57.10±1.34	57.81±1.68	60.34±2.20	70.28±1.84	91.95±1.31	60.89±1.57	92.58±2.02
Class13	48.43±1.56	46.45±1.71	37.43±1.83	38.65±1.98	38.10±1.34	49.06±2.92	48.36±2.49	72.26±1.53	49.61±2.12	71.25±2.27
Class14	83.33±0.78	83.24±1.00	77.66±1.24	82.45±0.67	81.46±0.56	92.62±1.14	85.15±1.20	98.12±0.78	81.40±1.33	96.99±0.61
Class15	96.26±0.40	95.86±0.52	86.64±1.32	93.91±0.92	93.03±0.94	86.99±1.94	92.55±0.59	93.96±0.23	95.49±0.80	99.40±0.35
Class16	90.68±0.38	89.22±0.60	84.06±1.03	88.26±0.66	87.20±0.75	50.87±1.67	89.08±0.93	96.78±1.04	89.74±1.07	90.11±2.45
Class17	99.45±0.42	98.70±0.72	85.34±2.57	97.60±1.51	98.01±0.94	100.00±0	96.10±1.81	100.00±0	99.90±2.60	100.00±0
Class18	84.09±1.24	79.40±1.44	79.08±1.12	79.52±1.16	79.45±1.14	80.51±2.59	93.11±1.02	89.83±1.39	90.35±0.43	80.53±1.79
Class19	92.50±0.86	88.82±0.75	83.71±1.77	84.80±0.54	86.84±0.47	89.45±1.50	89.51±1.05	98.25±0.57	95.62±0.89	95.53±1.36
Class20	97.19±0.58	96.11±0.56	93.94±0.92	93.51±0.70	94.16±0.55	99.93±0.10	96.76±0.57	99.97±0.02	99.73±0.19	99.97±0.02
OA	69.69±0.73	65.66±1.07	63.70±0.78	62.64±1.09	62.80±1.07	73.17±0.2	68.66±0.72	76.81±0.84	69.76±1.53	84.02±0.67
AA	84.01±0.14	81.81±0.16	78.63±0.28	80.17±0.23	80.38±0.17	80.07±0.27	82.47±0.19	89.05±0.17	82.80±0.29	89.82±0.19
kappa	63.32±0.73	59.02±1.05	56.73±0.79	55.86±1.06	56.02±1.02	66.54±0.43	62.03±0.75	71.29±0.93	63.18±1.61	79.67±0.78

Results are averaged by 20 trials with standard deviations listed as well.

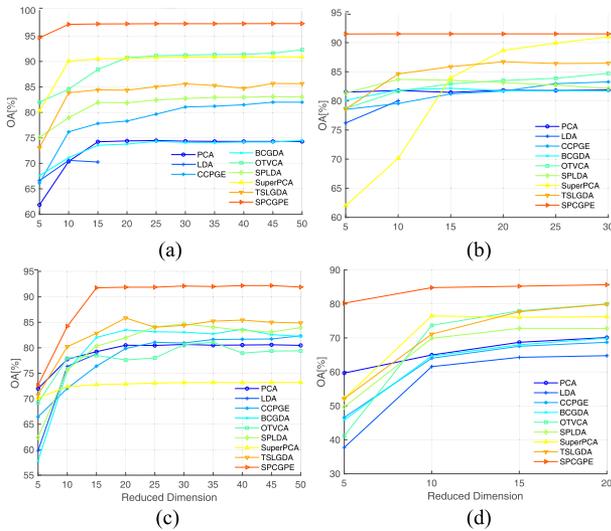


Fig. 6. OA [%] versus reduced dimension by several DR methods on the HSI datasets. (a) PaviaU. (b) MUFFL. (c) Houston2013. (d) Houston2018.

information while LDA, BCGDA, and TSLGDA are supervised with label information of training samples. In addition, PCA, LDA, CCPGE, BCGDA do not use any spatial information while TSLGDA utilizes spatial information via tensor, OTVCA via total variation minimization, and SPLDA, SuperPCA, and SPCRGE via superpixel segmentation.

A. Datasets

1) *PaviaU*: The first experimental urban HSI dataset is called PaviaU, which was acquired by the ROSIS sensor over the University of Pavia in 2002. It consists of 610×340 pixels with spatial resolution of 1.3m after some meaningless pixels are discarded and 103 spectral bands covering the wavelength

range $0.43 - 0.86 \mu\text{m}$. It contains nine classes and the numbers of training and testing samples of each class are listed in Table I.

2) *Muffl*: The second urban HSI dataset is called MUFFL, which was generated by the CASI-1500 sensor over the University of Southern Mississippi in November 2010. It consists of 325×220 pixels with spatial resolution of 0.54×1 m and 64 spectral bands in the wavelength range $0.38 - 1.05 \mu\text{m}$. It contains 11 classes and the numbers of training and testing samples of each class are listed in Table II.

3) *Houston2013*: The third urban HSI dataset is called Houston2013, which was gathered by the CASI-1500 sensor over the University of Houston campus and neighboring areas in June 2012. It consists of 349×1905 pixels with spatial resolution of 2.5 m and 144 spectral bands in the wavelength range $0.38 - 1.05 \mu\text{m}$. It contains 15 classes and the numbers of training and testing samples of each class are listed in Table III.

4) *Houston2018*: The fourth urban HST dataset is called Houston2018, which was captured by the ITRES CASI-1500 sensor over the University of Houston campus in the February of 2017. It consists of 601×2384 pixels with spatial resolution of 1 m and 48 spectral bands in the wavelength range $0.38 - 1.05 \mu\text{m}$. It contains 20 classes and the numbers of training and testing samples of each class are listed in Table IV.

B. Parameters Tuning

Classification performance based on the proposed SPCRGE is related to four parameters, i.e., the number of superpixels, regularization parameters λ and β of each superpixel, and reduced dimension. Before performing parameter-tuning experiments, some intuitive guidelines are listed as follows:

a) The best number of superpixels has relation to scenario complexity, and more complex the scene, the greater the number of superpixels.

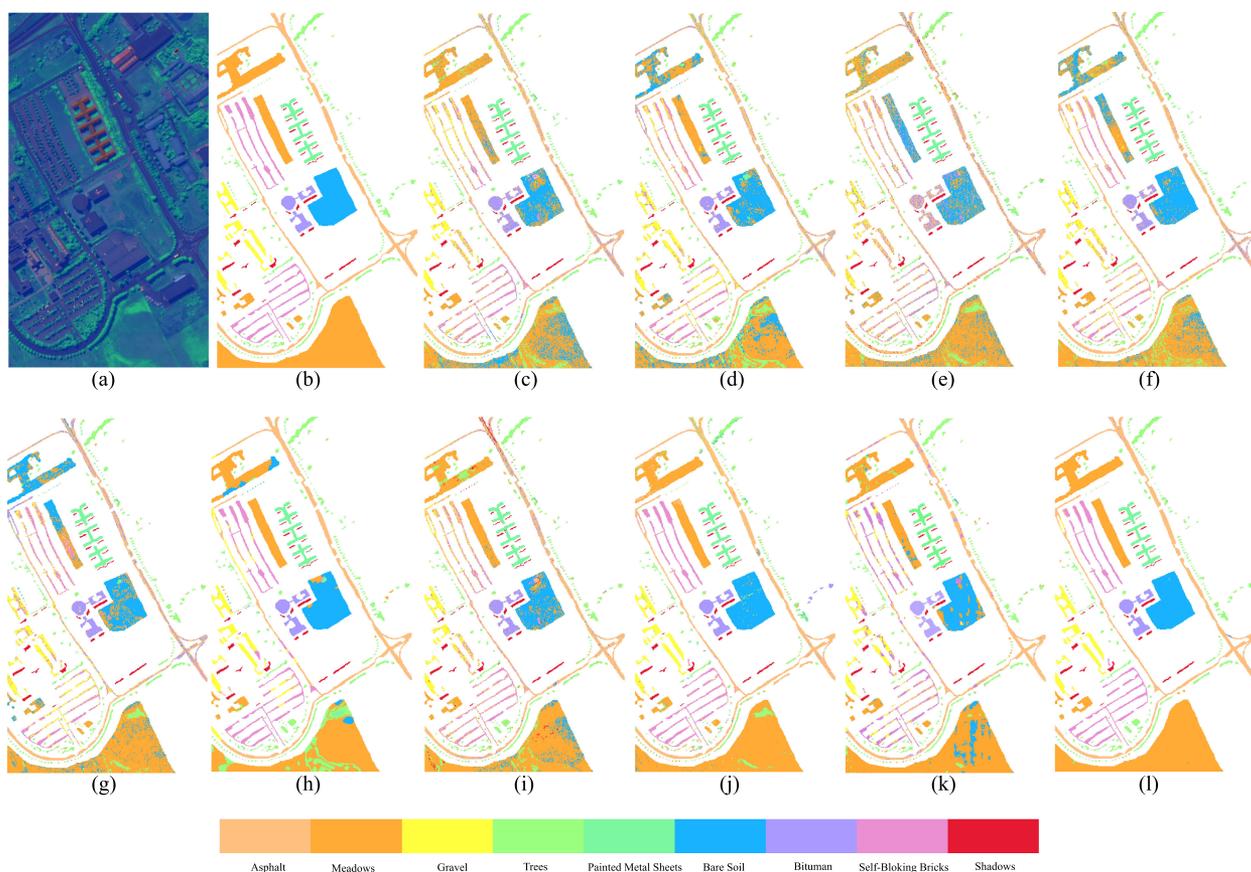


Fig. 7. Classification maps by different DR methods on PaviaU. (a) Pseudocolor map. (b) Groundtruth. (c) Raw (72.56%). (d) PCA (74.35%). (e) LDA (68.22%). (f) CCPGE (81.07%). (g) BCGDA (74.60%). (h) OTVCA (91.25%). (i) SPLDA (81.25%). (j) SuperPCA (90.84%). (k) TSLGDA (85.27%). (l) SPCRGE (97.64%).

b) Regularization parameter λ controls the degree of closeness to principal directions and β controls the degree of closeness of pixels in one superpixel. Therefore, they are related to scenario complexity as well. The regularization parameters decrease with the increase of the complexity to extract commonality and maintain individuality to some extent.

Based on guideline *a*), the number of superpixels is set in the range 50–1000 at every step of 50 for PaviaU and MUFFLE and set in the range 1000–10 000 at every step of 1000 for Houston2013 and Houston2018. Fig. 4 illustrates the overall accuracy (OA [%]) of the proposed SPCPGE as a function of the number of superpixels. The results in PaviaU, MUFFLE, and Houston2018 are averaged by 20 trials of randomly chosen training examples. It can be seen that the OA first increases and then decreases with the increase of superpixel number for MUFFLE and Houston2013, but the OAs for PaviaU and Houston2018 oscillate with the increase of superpixel number and are acceptable in a wide range of superpixel number. The best superpixel numbers for PaviaU, MUFFLE, Houston2013, and Houston2018 are 400, 400, 6000, and 3000, respectively. This is not surprising as Houston2013 and Houston2018 consist of more pixels and are more complex.

For simplicity, the regularization parameters λ and β take the same values for all superpixels in a dataset. For all datasets, regularization parameter λ is chosen from set $\{0.1, 0.5, 1, 2, 5,$

$10, 50\}$ and the other parameter β is chosen from set $\{0, 0.1, 0.5, 1, 5, 10, 50\}$. Fig. 5 illustrates the OA as a function of λ and β for the four datasets. For PaviaU, the best classification results are in a range $\lambda > 5$ or $\beta > 5$. For MUFFLE, the best classification results are in a range $\lambda > 5$ and seemingly have less relation to parameter β . However, for Houston2013 and Houston2018, acceptable classification results are in a wide range of λ and β and the best classification results are around $\lambda = 2$ and $\beta = 0$, and $\lambda = 10$ and $\beta = 0.5$, respectively. After the fine tuning of the best parameters λ and β , the best combinations of the parameters $\{\lambda, \beta\}$ are chosen as $\{5, 10\}$, $\{10, 10\}$, $\{1.9, 0\}$, and $\{50, 0.5\}$ for the four datasets in subsequent experiments, respectively. It conforms to guideline *b*) that the best parameters λ and β of Houston2013 and Houston2018 are less than that of PaviaU and MUFFLE, since Houston2013 and Houston2018 are more complex.

Fig. 6 depicts the OA as a function of reduced dimension for different DR methods on the four HSI datasets. With the increase of reduced dimension, the OAs for all DR methods first increase and then plateau except for LDA. Particularly, the proposed SPCRGE starts to plateau at a low reduced dimension for all datasets and is always better than the other DR methods in terms of OA. This means that SPCRGE can obtain sufficiently discriminative features in a extremely low-dimensional space. In addition, the classification results for all DR methods achieve

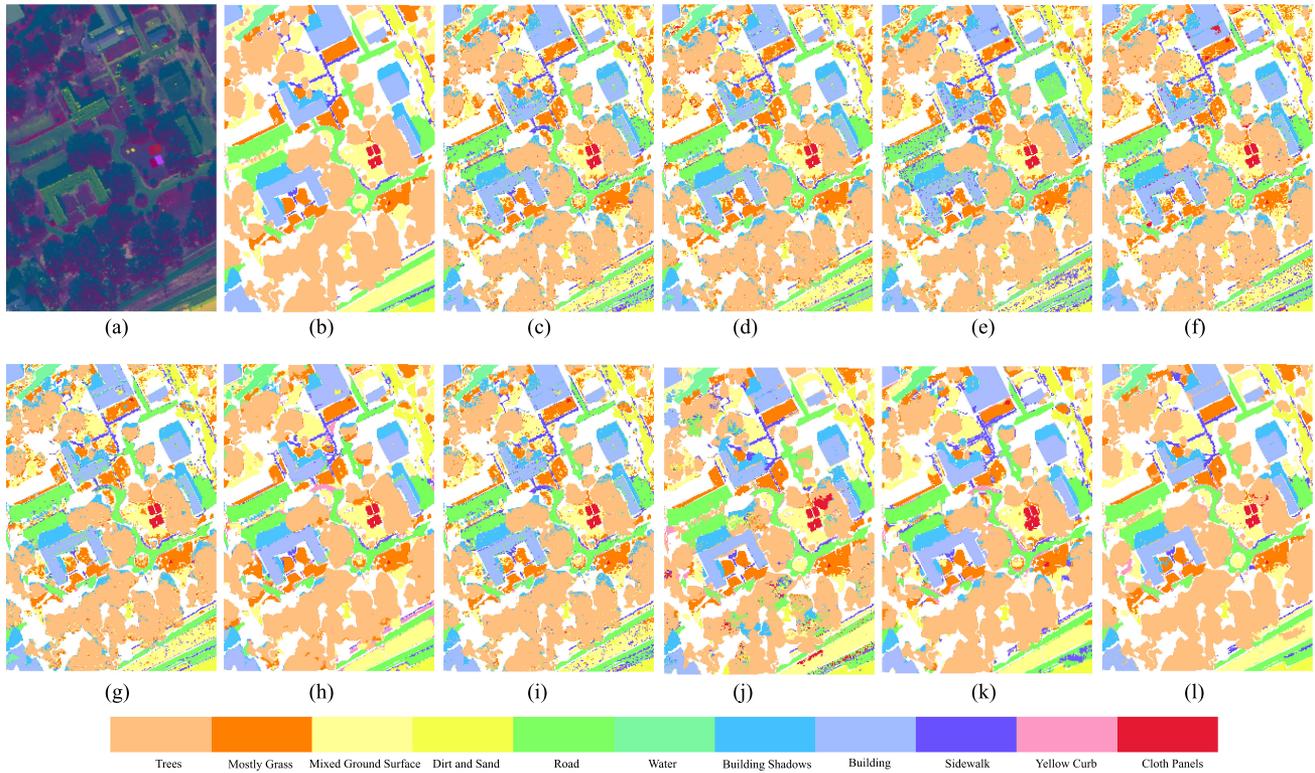


Fig. 8. Classification maps by different DR methods on MUFFL. (a) Pseudocolor map. (b) Groundtruth. (c) Raw (81.74%). (d) PCA (81.46%). (e) LDA (80.02%). (f) CCPGE (83.26%). (g) BCGDA (82.42%). (h) OTVCA (86.58%). (i) SPLDA (83.55%). (j) SuperPCA (83.97%). (k) TSLGDA (85.19%). (l) SPCRGE (91.66%).

acceptable results at reduced dimension 30 on PaviaU and Houston2013, 15 on MUFFL, and 10 on Houston2018 so in the following experiments, reduced dimensions for all DR methods are set as these numbers.

C. Classification Performance

To validate the effectiveness of the proposed SPCRGE, four experiments are conducted in comparison with the aforementioned DR methods. Classification performances of different DR methods on PaviaU, MUFFL, Houston2013, and Houston2018 are listed in the Tables V, VI, VII, and VIII, respectively, and classification results without DR are also listed as baselines. In the experiments on PaviaU, MUFFL, and Houston2018, the training examples are randomly chosen and results are averaged by 20 trials with standard deviations listed as well. For Houston2013, training examples are specified, so no standard deviation is listed. The reduced dimensions are fixed as 30 on PaviaU and Houston2013, 15 on MUFFL, and 10 on Houston2018 except for LDA whose reduced dimension is equal to the number of class minus one, i.e., 8 on PaviaU, 10 on MUFFL, 14 on Houston2013, and 19 on Houston2018.

In Tables V, VI, VII, and VIII for all experiments on the four datasets, spectral-spatial DR methods are generally better than spectral-only DR methods whether they are supervised or unsupervised. This indicates that spatial information is important for urban HSI classification. For spectral-only DR methods, CCPGE and BCGDA are better than PCA and LDA, and LDA performs

the worst in all experiments (except in the Houston2018 experiment where its reduced dimension can be 19) because its reduced dimension is limited to the number of class minus one. For spectral-spatial DR methods, OTVCA performs better than SPLDA, TSLGDA, and SuperPCA on PaviaU and MUFFL, while TSLGDA and SPLDA outperforms the other two on Houston2013. It is reasonable since TSLGDA is a tensor-based spectral-spatial with the help of the label information of the relatively sufficient training samples in Houston2013. However, for the scenario of limited training examples on PaviaU and MUFFL, the label information may be not comparable to the well-utilized spatial information brought by all pixels. This argument can be also supported by the proposed SPCRGE as it outperforms all the other aforementioned DR methods by approximately 6.4% on PaviaU (97.64%), 5% on MUFFL (91.66%), 7.5% on Houston2013 (92.13%), and 7.2% on Houston2018 (84.02%) over the second highest methods in terms of OA.

To visually demonstrate the effectiveness of SPCRGE, classification maps of all the aforementioned DR methods are illustrated in Figs. 7, 8, 9, and 10. Obviously, the proposed SPCRGE produces the most accurate and spatially smoothest classification maps with less mislabeled pixels, which are consistent with the results listed in Tables V–VIII.

Fig. 11 illustrates the classification performances on PaviaU, MUFFL, and Houston2018 with the increase of training ratio. The training ratio represents the ratio of selected training samples to all the labeled samples. As shown in Fig. 11, the training

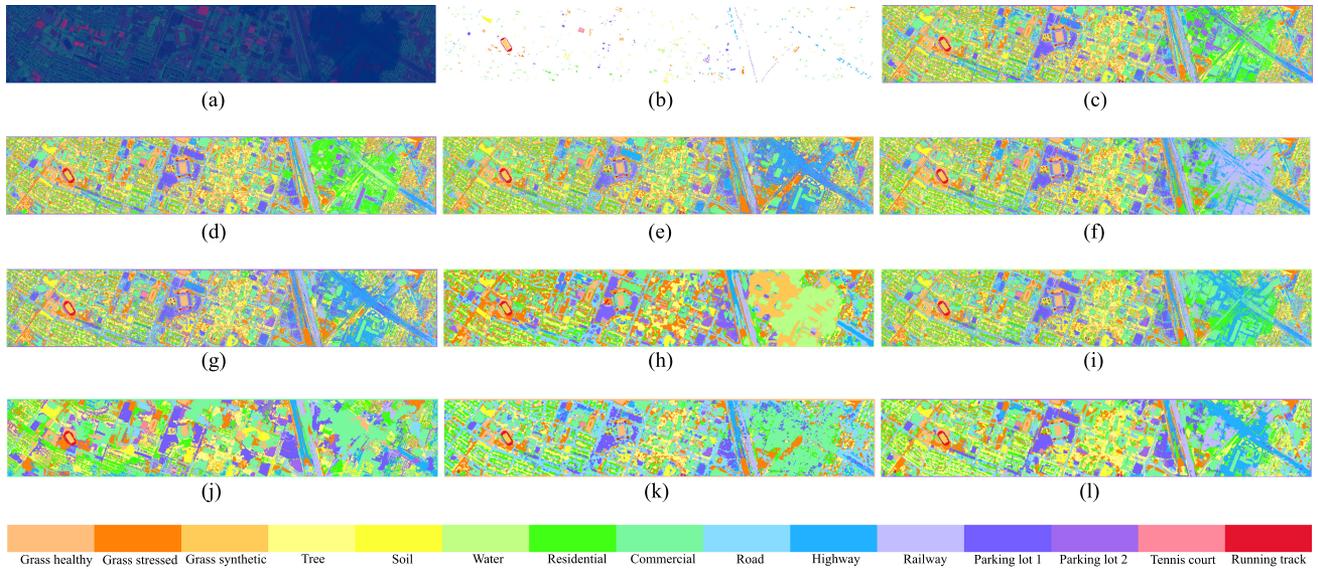


Fig. 9. Classification maps by different DR methods on Houston2013. (a) Pseudocolor map. (b) Groundtruth. (c) Raw (81.47%). (d) PCA (80.65%). (e) LDA (78.79%). (f) CCPGE (80.94%). (g) BCGDA (83.06%). (h) OTVCA (80.62%). (i) SPLDA (84.68%). (j) SuperPCA (73.17%). (k) TSLGDA (84.40%). (l) SPCRGE (92.13%).

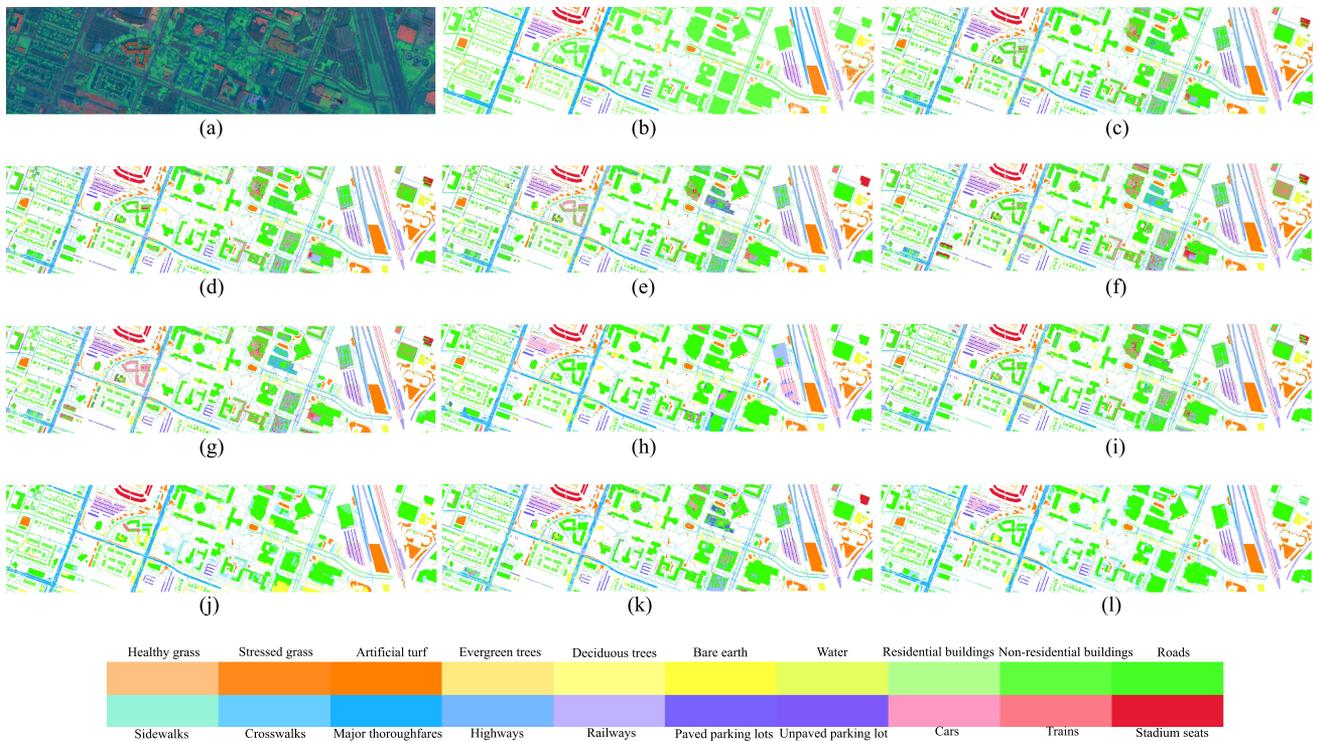


Fig. 10. Classification maps by different DR methods on Houston2018. (a) Pseudocolor map. (b) Groundtruth. (c) Raw (69.69%). (d) PCA (65.66%). (e) LDA (63.70%). (f) CCPGE (62.64%). (g) BCGDA (62.80%). (h) OTVCA (73.16%). (i) SPLDA (68.66%). (j) SuperPCA (76.81%). (k) TSLGDA (69.76%). (l) SPCRGE (84.02%).

ratio is in the range 0.5%–5% on PaviaU and Houston2018, and 1%–10% on MUFFL. From the results, the OAs of the proposed SPCRGE are always higher than the other methods, plateauing above 99% after 2.5% training ratio on PaviaU, reaching above 95% after 9% training ratio on MUFFL and above 93% after 3.5% training ratio on Houston2018. Overall, SPCRGE outperforms the other methods consistently.

To illustrate the computational complexity of the proposed SPCRGE compared to other DR methods, Table IX shows the computing time of several DR methods on PaviaU by using MATLAB on an Intel(R) Core(TM) i5-7300HQ Central processing unit with 8 GB of RAM. The methods, SPLDA, SuperPCA, and the proposed SPCRGE contain computing time of ERS superpixel segmentation. As shown in Table IX,

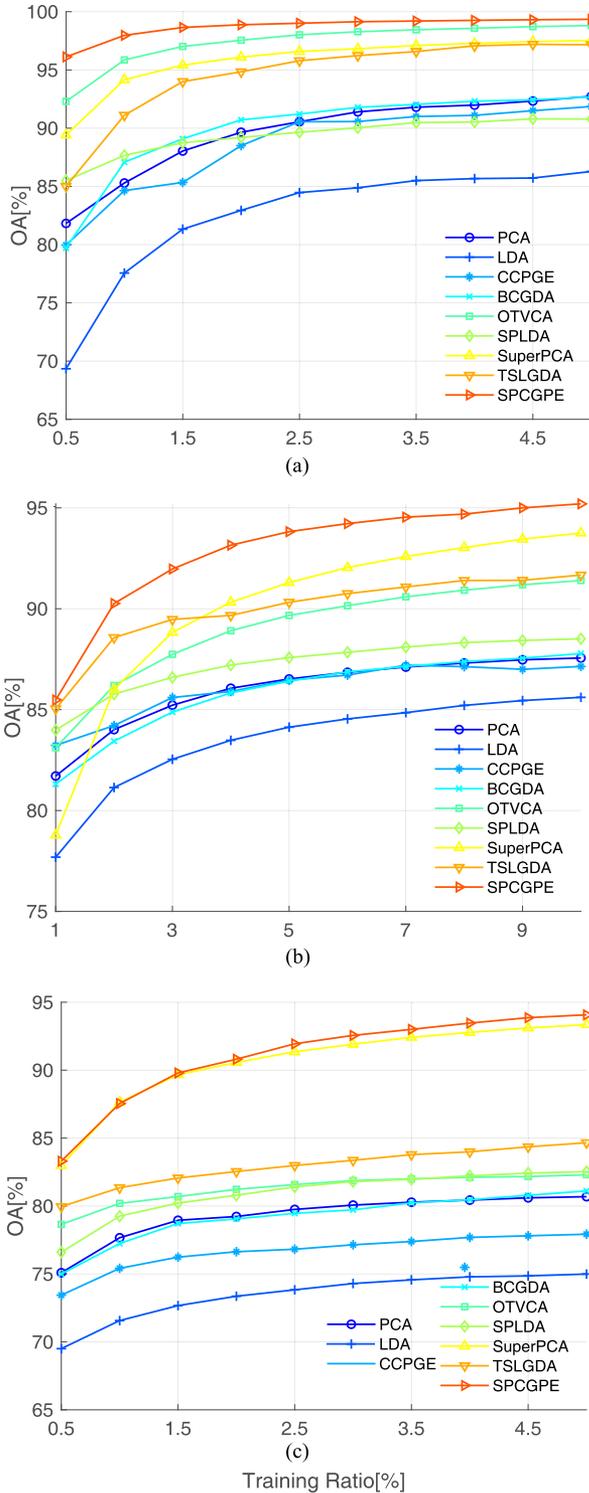


Fig. 11. OA [%] versus training ratio [%] on the HSI datasets. (a) PaviaU. (b) MUFFL. (c) Houston2018.

although its classification performance is poor, LDA is the fastest among all the DR methods. The computing time of SPCRGE is 56.75s, which is acceptable considering about its classification performance. The computing burden of SPCRGE mainly lies on the calculation of SPCRs. OTVCA and TSLGDA are the two slowest DR methods. The heaviest computing burden of

TABLE IX
COMPUTATION TIME (IN SECONDS) OBTAINED THROUGH DIFFERENT METHODS FOR PAVIAU

Algorithm	Computation Time(s)
PCA	1.84
LDA	0.11
CCPGE	72
BCGDA	0.6
OTVCA	2719
SPLDA	9.02
SuperPCA	3.65
TSLGDA	214
SPCRGE	56.75

OTVCA lies on the step of total variation minimization while that of TSLGDA lies on the calculation of sparse and low-rank representation.

V. CONCLUSION

In this article, an unsupervised spatial-spectral DR method called SPCRGE has been proposed for the HSI classification. In SPCRGE, abundant spatial information is utilized by superpixel segmentation. Then, to extract commonality and maintain individuality of the pixels to some extent in each superpixel, Laplacian regularized superpixelwise collaborative representation (SPCR) of each pixel was obtained by solving a generalized Sylvester equation. By reducing the discrepancy between SPCRs and the original spectral features, and reducing the differences between pixels from one superpixel and increasing the differences between pixels from different superpixels simultaneously, a global projection matrix to a low-dimensional space is obtained. In this way, the problem of spectral information redundancy and Hughes phenomenon, and the problem of intraclass difference and interclass similarity in HSI classification can be alleviated simultaneously. Experiments on four urban HSI datasets have demonstrated that the proposed SPCRGE outperforms the existing state-of-the-art DR methods and confirmed its effectiveness.

REFERENCES

- [1] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [2] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.
- [3] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.
- [4] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.
- [5] X. Mei *et al.*, "Spectral-spatial attention networks for hyperspectral image classification," *Remote Sens.*, vol. 11, no. 8, 2019, Art. no. 963.
- [6] W. Sun, J. Peng, G. Yang, and Q. Du, "Correntropy-based sparse spectral clustering for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 484–488, Mar. 2020.
- [7] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

- [8] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.
- [9] S. Yang, Z. Shi, and W. Tang, "Robust hyperspectral image target detection using an inequality constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3389–3404, Jun. 2015.
- [10] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognit.*, vol. 48, no. 12, pp. 3904–3916, 2015.
- [11] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [12] R. Tao, X. Zhao, W. Li, H.-C. Li, and Q. Du, "Hyperspectral anomaly detection by fractional Fourier entropy," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.
- [13] A. Ertürk, M.-D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 708–719, Feb. 2016.
- [14] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
- [15] P. Chen, L. Jiao, F. Liu, J. Zhao, Z. Zhao, and S. Liu, "Semi-supervised double sparse graphs based discriminant analysis for dimensionality reduction," *Pattern Recognit.*, vol. 61, pp. 361–378, 2017.
- [16] P. Chen, L. Jiao, F. Liu, Z. Zhao, and J. Zhao, "Adaptive sparse graph learning based dimensionality reduction for classification," *Appl. Soft Comput.*, vol. 82, 2019, Art. no. 105459.
- [17] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [18] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Amsterdam, The Netherlands: Elsevier, 2013.
- [19] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.
- [20] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [21] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [22] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [23] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, vol. 2, 2005, pp. 1208–1213.
- [24] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 153–160.
- [25] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
- [26] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with l^1 -graph for image analysis," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 858–866, Apr. 2010.
- [27] N. H. Ly, Q. Du, and J. E. Fowler, "Sparse graph-based discriminant analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3872–3884, Jul. 2014.
- [28] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 471–478.
- [29] 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, Jun. 2014.
- [30] N. Liu, W. Li, and Q. Du, "Unsupervised feature extraction for hyperspectral imagery using collaboration-competition graph," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1491–1503, Dec. 2018.
- [31] A. Mohan, G. Sapiro, and E. Bosch, "Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 206–210, Apr. 2007.
- [32] 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.
- [33] H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2604–2616, Jun. 2020.
- [34] L. Pan, H.-C. Li, Y.-J. Deng, F. Zhang, X.-D. Chen, and Q. Du, "Hyperspectral dimensionality reduction by tensor sparse and low-rank graph-based discriminant analysis," *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 452.
- [35] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976–6985, Dec. 2016.
- [36] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral-spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, May 2014.
- [37] T. Lu, S. Li, L. Fang, L. Bruzzone, and J. A. Benediktsson, "Set-to-set distance-based spectral-spatial classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7122–7134, Dec. 2016.
- [38] B. Cui, X. Xie, X. Ma, G. Ren, and Y. Ma, "Superpixel-based extended random walker for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3233–3243, Jun. 2018.
- [39] P. Sellars, A. I. Aviles-Rivero, and C.-B. Schönlieb, "Superpixel contracted graph-based learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4180–4193, Jun. 2020.
- [40] H. Xu, H. Zhang, W. He, and L. Zhang, "Superpixel-based spatial-spectral dimensionality reduction for hyperspectral imagery classification," *Neurocomputing*, vol. 360, pp. 138–150, 2019.
- [41] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.
- [42] 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, Feb. 2015.
- [43] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [44] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104.
- [45] F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Inf. Sci.*, vol. 397, pp. 48–68, 2017.
- [46] Y. Guo, L. Jiao, S. Wang, S. Wang, F. Liu, and W. Hua, "Fuzzy superpixels for polarimetric sar images classification," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 2846–2860, Oct. 2018.
- [47] W. Li and Q. Du, "Laplacian regularized collaborative graph for discriminant analysis of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7066–7076, Dec. 2016.
- [48] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2014.



Huan Liu received the B.S. degree from China University of Mining and Technology, Xuzhou, China, and the M.S. degree from Beijing Institute of Technology, Beijing, China, in 2013 and 2016, respectively. He is currently working toward the Ph.D. degree with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China.

His research interests include image processing, image classification, and transfer learning.



Wei Li (Senior Member, IEEE) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Postdoctoral Researcher with the University of California at Davis, Davis, CA, USA. He is a Professor with the School

of Information and Electronics, Beijing Institute of Technology, Beijing, China. His research interests include hyperspectral image (HSI) analysis, pattern recognition, and data compression.

Dr. Li received the 2015 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society for his service for the IEEE JSTARS. He is serving as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He served as a Guest Editor for special issue of the *Journal of Real-Time Image Processing, Remote Sensing*, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS).



Xiang-Gen Xia (Fellow, IEEE) received the B.S. degree in mathematics from Nanjing Normal University, Nanjing, China, the M.S. degree in mathematics from Nankai University, Tianjin, China, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1983, 1986, and 1992, respectively.

He was a Senior/Research Staff Member at Hughes Research Laboratories, Malibu, CA, USA, during 1995–1996. In September 1996, he joined the Department of Electrical and Computer Engineering,

University of Delaware, Newark, Delaware, where he is the Charles Black Evans Professor. His current research interests include space-time coding, MIMO and OFDM systems, digital signal processing, and SAR and ISAR imaging. He is the author of the book *Modulated Coding for Intersymbol Interference Channels* (New York, Marcel Dekker, 2000).

Dr. Xia received the National Science Foundation (NSF) Faculty Early Career Development (CAREER) Program Award in 1997, the Office of Naval Research (ONR) Young Investigator Award in 1998, and the Outstanding Overseas Young Investigator Award from the National Nature Science Foundation of China in 2001. He received the 2019 Information Theory Outstanding Overseas Chinese Scientist Award, The Information Theory Society of Chinese Institute of Electronics. He has served as an Associate Editor for numerous international journals including IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is the Technical Program Chair of the Signal Processing Symp., Globecom 2007 in Washington, DC, USA, and the General Co-Chair of ICASSP 2005 in Philadelphia.



Mengmeng Zhang received the B.S. degree in computer science and technology from the Qingdao University of Science and Technology, Qingdao, China, in 2014, and the Ph.D. degree in control science and engineering from Beijing University of Chemical Technology, Beijing, China, in 2019.

She is currently a Postdoctoral Researcher with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. Her research interests include remote sensing image processing and pattern recognition. Her current research interests

include remote sensing image process and pattern recognition.



Ran Tao (Senior Member, IEEE) was born in 1964. He received the B.S. degree from the Electronic Engineering Institute of PLA, Hefei, China, in 1985, and the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1990 and 1993, respectively.

In 2001, he was a Senior Visiting Scholar with the University of Michigan, Ann Arbor, MI, USA. He is a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. He has three books and more than 100 peer-

reviewed journal articles. His research interests include fractional Fourier transform and its applications, theory, and technology for radar and communication systems.

Dr. Tao is a member of the Wireless Communication and Signal Processing Commission of International Union of Radio Science (URSI). He was a recipient of the National Science Foundation of China for Distinguished Young Scholars in 2006, and the First Prize of Science and Technology Progress in 2006 and 2007, and the First Prize of Natural Science in 2013, both awarded by the Ministry of Education. He was a Distinguished Professor of the Changjiang Scholars Program in 2009. He was a Chief Professor of the Program for Changjiang Scholars and Innovative Research Team in University from 2010 to 2012. He has been a Chief Professor of the Creative Research Groups of the National Natural Science Foundation of China since 2014. He is the Vice Chair of the IEEE China Council and the URSI China Council. He is an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS (SPL).