

Dimensionality Reduction of Hyperspectral Image Using Spatial Regularized Local Graph Discriminant Embedding

Renlong Hang , *Member, IEEE*, and Qingshan Liu , *Senior Member, IEEE*

Abstract—Dimensionality reduction (DR) is an important pre-processing step for hyperspectral image (HSI) classification. Recently, graph-based DR methods have been widely used. Among various graph-based models, the local graph discriminant embedding (LGDE) model has shown its effectiveness due to the complete use of label information. Besides spectral information, an HSI also contains rich spatial information. In this paper, we propose a regularization method to incorporate the spatial information into the LGDE model. Specifically, an oversegmentation method is first employed to divide the original HSI into nonoverlapping superpixels. Then, based on the observation that pixels in a superpixel often belong to the same class, intraclass graphs are constructed to describe such spatial information. Finally, the constructed superpixel-level intraclass graphs are used as a regularization term, which can be naturally incorporated into the LGDE model. Besides, to sufficiently capture the nonlinear property of an HSI, the linear LGDE model is further extended into its kernel counterpart. To demonstrate the effectiveness of the proposed method, experiments have been established on three widely used HSIs acquired by different hyperspectral sensors. The obtained results show that the proposed method can achieve higher classification performance than many state-of-the-art graph embedding models, and the kernel extension model can further improve the classification performance.

Index Terms—Dimensionality reduction (DR), hyperspectral image (HSI) classification, local graph discriminant embedding (LGDE), spatial regularization.

I. INTRODUCTION

SATELLITE-BORNE and airborne sensors are able to measure the spectrum of solar radiation reflected by the Earth's surface. Compared to multispectral sensors, hyperspectral sensors can provide much richer spectral information. Thus, they have become an important tool for detection or classification of land cover materials with complex compositions. However, new challenges arise when dealing with hyperspectral image (HSI) classification. The first challenge is the curse of dimensionality problem. Since the labeling of training samples is time

consuming, we often encounter hundreds of spectral bands with a small number of training samples, which may easily lead to the Hughes phenomenon [1]. The second challenge is the information redundancy problem. Due to the dense sampling of spectral bands across the electromagnetic spectrum, the high-dimensional spectral features are often correlated. This redundancy information not only increases the cost of computation and storage, but may also degrade the classification performance [2].

Dimensionality reduction (DR) has been widely used to address the aforementioned issues [3]–[7]. The goal of DR is to seek a low-dimensional subspace, in which some desired information of the original HSI can be preserved. The existing DR algorithms can be divided into two major classes: band selection and feature transformation. Band selection retains physically meaningful features for classification [8], [9], but the algorithms are often computationally intensive and not robust in complex scenes [4]. Therefore, feature transformation has attracted more and more attention. For instance, principal component analysis (PCA) attempts to find an orthogonal set of vectors that maximizes the variance of the projected data [10]. Linear discriminative analysis (LDA) generates the best projection by maximizing the between-class scatter matrix while minimizing the within-class scatter matrix [11]–[14]. Different from them, aiming to preserve the statistic properties of the original data, graph-based feature transformation methods attempt to capture the geometric properties of neighboring samples [15].

The construction of graph is an essential step for graph-based feature transformation. It usually consists of two steps: determination of graph adjacency relationships and calculation of edge weights [16]. For adjacency construction, k -nearest neighbors (KNN) and ε -ball neighborhood are the most commonly used methods to obtain a sparse graph [17], [18]. For edge weight calculation, numerous methods adopt a heat kernel function to measure the similarity between data points [19], [20]. In recent years, some other methods have been proposed to construct the graph. In [21], a sparse graph discriminant embedding (SGDE) model [22] was employed to an HSI. It uses sparse representation [23] to automatically learn the adjacency relationships and edge weights. Due to its effectiveness, many related algorithms have been further proposed [24]–[27]. In [26], a weighted sparse graph-based DR model was proposed to integrate both locality and sparsity structures. In [27], a sparse and low-rank graph-based discriminant analysis was presented by combining both

Manuscript received January 5, 2018; revised March 30, 2018 and May 25, 2018; accepted June 5, 2018. Date of publication July 1, 2018; date of current version September 5, 2018. This work was supported in part by the Natural Science Foundation of China under Grant 61532009 and in part by the Natural Science Foundation of Jiangsu Province, China, under Grant 15KJA520001. (Corresponding author: Renlong Hang.)

The authors are with the Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: renlong_hang@163.com; qslu@nuist.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2018.2847042

sparsity and low rankness to maintain global and local structures simultaneously. Different from the sparse graph-based methods, a collaborative graph-based discriminant embedding (CGDE) was introduced to replace the ℓ_1 -norm minimization in constructing the graph by ℓ_2 -norm minimization [28]. Due to the existence of a closed-form solution for collaboration representation, it can be computationally very efficient. In [29], a Laplacian regularized CGDE was proposed to consider the manifold structure of an HSI.

To use the label information of training samples, both SGDE and CGDE models construct the dictionary by collecting the training samples from the same class as the atoms. Therefore, the intraclass structure information can be well considered, but the interclass information may not be fully explored. Different from them, local graph discriminant embedding (LGDE) model attempts to construct intraclass graph as well as interclass graph simultaneously, and aims to maintain the neighbor relations of within-class samples while keeping away neighboring points of between-class samples after embedding [30]. Therefore, LGDE can discover both the geometric and discriminative structures of data. However, besides the original spectral information, spatial structure information also plays an important role for HSI classification [31]–[33]. As summarized in our previous paper [2], the existing methods about the use of spatial information can be roughly divided into three classes: feature-level fusion [34]–[37], decision-level fusion [31], [38], and regularization-based fusion [39]–[41].

In this paper, we propose a regularization method to incorporate the spatial information into the LGDE model naturally. Specifically, we first use an oversegmentation method to divide the original HSI into nonoverlapping superpixels. Then, for each superpixel, an intraclass graph is constructed to describe the spatial structure information. Finally, the constructed superpixel-level intraclass graphs are used as a regularization term for LGDE, resulting in a spatial regularized LGDE (SLGDE). Besides, to sufficiently capture the nonlinear property of an HSI caused by the complex acquisition process as well as the impacts of atmospheric and geometric distortions, we further extend the linear SLGDE model into its kernel counterpart (KSLGDE).

The rest of this paper is outlined as follows. Section II introduces the proposed methods in detail, including the SLGDE and KSLGDE models. Section III presents the three used datasets and the experimental results, followed by the conclusion in Section IV.

II. METHODOLOGY

Fig. 1 shows the flowchart of the proposed SLGDE model. It consists of four steps. First, PCA is applied to the input HSI to obtain the first principal component of it. Second, an oversegmentation method is used to divide the principal component into superpixels, whose shape and size can be adaptively changed according to different spatial structures. Then, for each superpixel, an intraclass graph is constructed to capture its spatial information based on the observation that pixels in a superpixel often belong to the same class. Meanwhile, the training samples with their labels are used to construct an intraclass graph and

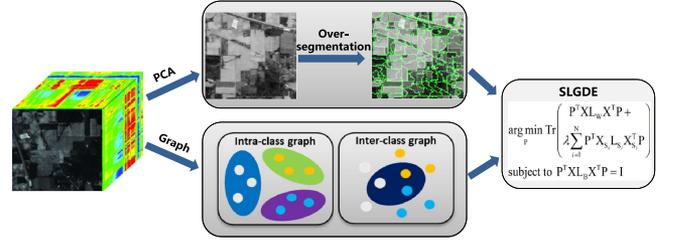


Fig. 1. Flowchart of the proposed SLGDE model.

interclass graph simultaneously. Finally, the superpixel-level intraclass graphs are used as a regularization term to formulate an SLGDE model. In the following sections, we will introduce the SLGDE model and its kernel counterpart in detail.

A. Spatial Regularized LGDE

Suppose we randomly choose m training pixels $\{x_i | x_i \in \mathcal{R}^d\}_{i=1}^m$ and their corresponding class labels $\{y_i\}_{i=1}^m$ from a given HSI, where d is the number of spectral bands. The goal of LGDE is to learn a low-dimensional embedding that characterizes the local geometrical and discriminative properties of the original pixels [30]. To achieve this goal, LGDE needs to construct two undirected graphs over the whole training pixels: the intraclass graph G_ω and the interclass graph G_b . To construct G_ω , we consider each pair of pixels x_i and x_j from the same class, i.e., $y_i = y_j$, and add an edge between them if x_j belongs to the K_ω -nearest neighbors of x_i . The edge weight w_{ij} can be calculated by a heat kernel as follows:

$$w_{ij} = \exp(-\|x_i - x_j\|^2/t) \quad (1)$$

where t is a parameter of the heat kernel. By default, $w_{ij} = 0$ if there is no edge between x_i and x_j . Likewise, we can construct G_b by considering each pair of x_i and x_j from different classes, i.e., $y_i \neq y_j$, and connecting them if x_j belongs to the K_b -nearest neighbors of x_i . The edge weight w'_{ij} can also be calculated from (1).

Assume the low-dimensional embedding for a given pixel x_i is $z_i = P^T x_i$. LGDE attempts to maintain the neighbor relationships of pixels from the same class while keeping away neighboring pixels from different classes, which can be formulated as follows:

$$\begin{aligned} \arg \min_P & \sum_{i=1}^m \sum_{j=1}^m \|P^T x_i - P^T x_j\|^2 w_{ij} \\ \text{s.t.} & \sum_{i=1}^m \sum_{j=1}^m \|P^T x_i - P^T x_j\|^2 w'_{ij} = 1. \end{aligned} \quad (2)$$

The traditional LGDE model can make full use of spectral features to learn the projection function, but the spatial structure information may also be beneficial to DR or classification of an HSI. Therefore, we propose a modified LGDE model named SLGDE. Since the objects in a given HSI often exist in different sizes and shapes, using spatial regions with a fixed scale to extract features easily loses useful information. Inspired from [42]–[44], we adopt superpixel to address this issue. Each

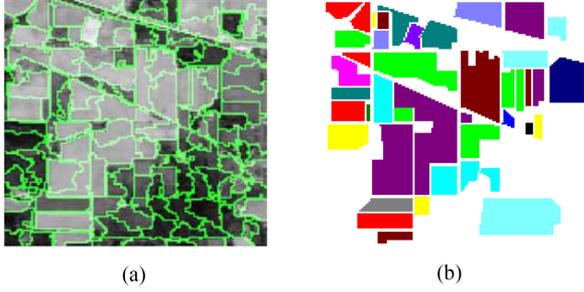


Fig. 2. Example of (a) oversegmentation map and (b) ground-truth map on the Indian Pines dataset.

superpixel is a local region, whose size and shape can be adaptively adjusted according to local structures.

The superpixels can be created by using an efficient oversegmentation method on an HSI. Similar to [43] and [44], we also use the oversegmentation method in [45] to generate L nonoverlapping superpixels. To reduce the computation cost, we apply PCA on the original HSI to obtain the first principal component corresponding to the highest eigenvalue. Since the first principal component contains the most important information in terms of variation for the whole HSI, we use it as the base image for the oversegmentation.

An oversegmentation example can be seen in Fig. 2(a). Compared to the ground-truth map [see Fig. 2(b)] on the same dataset, we can observe that pixels in a superpixel often belong to the same class. Based on this observation, an intraclass graph can be constructed to capture this structure information. Similar to G_ω , each pair of x_i and x_j in the ℓ th superpixel $\{S_\ell\}_{\ell=1}^L$ is connected if x_i belongs to the K_s -nearest neighbors of x_j , and the edge weight \tilde{w}_{ij} can also be calculated according to (1). These superpixel-level intraclass graphs are then used as a regularization term of LGDE, resulting in the SLGDE model

$$\begin{aligned} \arg \min_P & \sum_{i=1}^m \sum_{j=1}^m \|P^\top x_i - P^\top x_j\|^2 w_{ij} \\ & + \lambda \frac{1}{L} \sum_{\ell=1}^L \sum_{x_i \in S_\ell} \sum_{x_j \in S_\ell} \|P^\top x_i - P^\top x_j\|^2 \tilde{w}_{ij} \\ \text{s.t.} & \sum_{i=1}^m \sum_{j=1}^m \|P^\top x_i - P^\top x_j\|^2 w'_{ij} = 1 \end{aligned} \quad (3)$$

where λ is a regularization parameter. If $\lambda = 0$, SLGDE reduces to LGDE.

According to the weight of each edge, we can derive the affinity matrices of intraclass graph W , interclass graph W' , and superpixel-level intraclass graph \tilde{W}_ℓ , where the i th row and the j th column elements in W , W' , and \tilde{W}_ℓ are w_{ij} , w'_{ij} , and \tilde{w}_{ij} , respectively. Their corresponding graph Laplacian matrices are $L = D - W$, $L' = D' - W'$, and $\tilde{L}_\ell = \tilde{D}_\ell - \tilde{W}_\ell$, where D , D' , and \tilde{D}_ℓ are diagonal matrices whose entry equals to the sum of its corresponding row in W , W' , and \tilde{W}_ℓ , respectively. To simplify the computation, (3) is often rewritten in a matrix

manner as

$$\begin{aligned} \arg \min_P & \text{tr}(P^\top X L X^\top P + \lambda \frac{1}{L} \sum_{\ell=1}^L P^\top X_\ell \tilde{L}_\ell X_\ell^\top P) \\ \text{s.t.} & \text{tr}(P^\top X L' X^\top P) = 1 \end{aligned} \quad (4)$$

where $\text{tr}(\cdot)$ denotes the trace of matrix, $X = [x_1, x_2, \dots, x_m]$ is the matrix representation of training pixels, and X_ℓ represents the pixel set whose columns correspond to pixels in the ℓ th superpixel.

The columns of the optimal P in (4) are the generalized eigenvectors corresponding to the smallest eigenvalues in the following equation:

$$X L X^\top P + \lambda \frac{1}{L} \sum_{\ell=1}^L X_\ell \tilde{L}_\ell X_\ell^\top P = \gamma X L' X^\top P \quad (5)$$

where γ is the Lagrange multiplier. Once we have learned the optimal projection matrix P , nearest neighbor classifications become straightforward. For any test pixel $x_t \in \mathbb{R}^d$, we can compute its low-dimensional representation $z_t = P^\top x_t$. Its label is then predicted as y_{i^*} provided that $z_{i^*} = P^\top x_{i^*}$ minimizes $\|z_i - z_t\|$.

B. Kernel Extension of SLGDE

As discussed in [46], hyperspectral data exist nonlinear distribution characteristics caused by multiple sources. One of the significant sources, especially in the application of land-cover classification, stems from the nonlinear nature of scattering, as described in the bidirectional reflectance distribution function [47]. Other effects that contribute to nonlinearities include multiple scattering within a pixel and the heterogeneity of subpixel constituents [48], [49]. However, SLGDE is a linear model, which may ignore the inherent nonlinear properties of hyperspectral data. To address this issue, we extend the SLGDE model into its kernel counterpart.

Suppose we map the training set X into a higher dimensional Hilbert space \mathcal{H} via a nonlinear mapping $\Phi: \mathbb{R}^d \rightarrow \mathcal{H}$. Then, the original input data can be represented as $\Phi(X)$ in \mathcal{H} . According to the *representer theorem* [50], the projection P can be expressed as a combination of mapped training pixels, i.e., $P = \sum_{i=1}^m \alpha_i \Phi(x_i) = \Phi(X)\alpha$, where $\alpha = [\alpha_1, \dots, \alpha_m]^\top$. Based on them, the objective function of KSLGDE can be formulated as

$$\begin{aligned} \arg \min_\alpha & \text{tr}(\alpha^\top \Phi(X)^\top \Phi(X) L \Phi(X)^\top \Phi(X) \alpha) \\ & + \lambda \frac{1}{L} \sum_{\ell=1}^L \alpha^\top \Phi(X)^\top \Phi(X_\ell) \tilde{L}_\ell \Phi(X_\ell)^\top \Phi(X) \alpha \\ \text{s.t.} & \text{tr}(\alpha^\top \Phi(X)^\top \Phi(X) L' \Phi(X)^\top \Phi(X) \alpha) = 1. \end{aligned} \quad (6)$$

Due to the high dimensionality of the feature space \mathcal{H} , it is computationally not feasible to directly implement any algorithm in it. However, kernel-based learning algorithms use an effective kernel trick to implement dot products in the feature space by employing a kernel function $K(\cdot, \cdot)$,

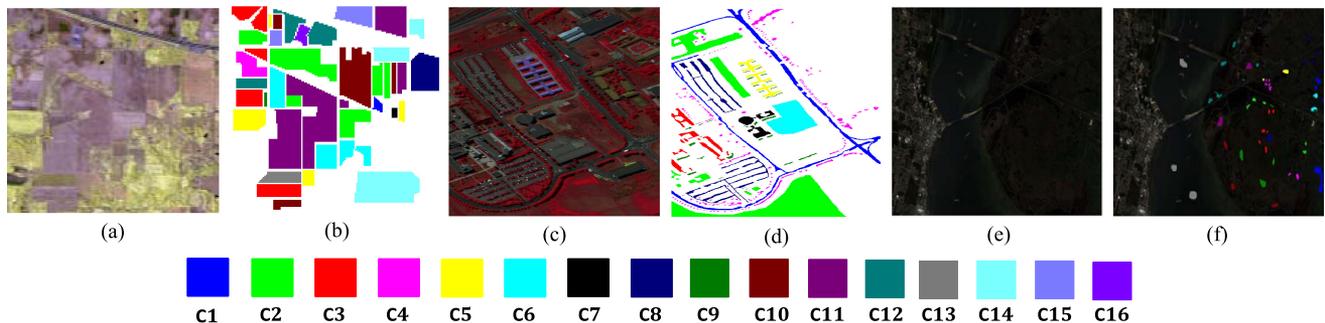


Fig. 3. RGB composite images and ground-truth maps on the three datasets. (a), (b) IP dataset. (c), (d) PUS dataset. (e), (f) KSC dataset.

TABLE I
NUMBERS OF PIXELS FOR THE TRAINING/TESTING AND THE TOTAL NUMBER OF PIXELS IN EACH CLASS IN THE IP GROUND-TRUTH MAP

Class		Sample			Class		Sample		
Label	Name	Total	Training	Testing	Label	Name	Total	Training	Testing
C1	Alfalfa	46	5	41	C9	Oats	20	2	18
C2	Corn-notill	1428	143	1285	C10	Soybean-notill	972	97	875
C3	Corn-mintill	830	83	747	C11	Soybean-mintill	2455	246	2209
C4	Corn	237	24	213	C12	Soybean-clean	593	59	534
C5	Grass-pasture	483	48	435	C13	Wheat	205	21	184
C6	Grass-trees	730	73	657	C14	Woods	1265	127	1138
C7	Grass-pasture-mowed	28	3	25	C15	Buildings-Grass-Trees-Drives	386	39	347
C8	Hay-windrowed	478	48	430	C16	Stone-Steel-Towers	93	9	84

which can be described as $K(x_i, x_j) = \Phi(x_i)^\top \Phi(x_j)$. Commonly used kernel functions include the polynomial kernel $K(x_i, x_j) = (x_i^\top x_j + 1)^t$ and the Gaussian radial basis kernel $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/t)$, where $t > 0$ is the parameter of kernel functions. Therefore, (6) can be rewritten as

$$\begin{aligned} \arg \min_{\alpha} \quad & \text{tr}(\alpha^\top K_t L K_t^\top \alpha + \frac{\lambda}{L} \sum_{\ell=1}^L \alpha^\top K_\ell \tilde{L}_\ell K_\ell^\top \alpha) \\ \text{s.t.} \quad & \text{tr}(\alpha^\top K_t L' K_t^\top \alpha) = 1. \end{aligned} \quad (7)$$

Similar to SLGDE, the optimal α of (7) is the solution of the following generalized eigenvalue problem:

$$K_t L K_t^\top \alpha + \frac{\lambda}{L} \sum_{\ell=1}^L K_\ell \tilde{L}_\ell K_\ell^\top \alpha = \gamma K_t L' K_t^\top \alpha. \quad (8)$$

To test a new pixel $x_t \in \mathfrak{R}^d$ by nearest neighbor classification with KSLGDE, we need to compute its projection as $z_t = P^\top x_t = \sum_{i=1}^m \alpha_i K(x_i, x_t)$, and then find its nearest neighbor in the low-dimensional embedding space.

III. EXPERIMENTS

A. Dataset

We test the proposed methods on three HSIs, which are widely used to evaluate DR and classification algorithms.

- 1) Indian Pines (IP) dataset: The first dataset was acquired by the AVIRIS sensor over the Indian Pine test site in northwestern Indiana, USA, on June 12, 1992. The original dataset contains 224 spectral bands. We utilize 200

of them after removing four bands containing zero values and 20 noisy bands affected by water absorption. The spatial size of the image is 145×145 pixels, and the spatial resolution is 20 m. The false-color composite image and the ground-truth map are shown in Fig. 3(a) and (b), respectively. The number of the totally labeled pixels, the training pixels, and the testing pixels for each class are reported in Table I.

- 2) Pavia University Scene (PUS) dataset: The second dataset was acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy, on July 8, 2002. The original image was recorded with 115 spectral channels ranging from $0.43 \mu\text{m}$ to $0.86 \mu\text{m}$. After removing noisy bands, 103 bands are used. The image size is 610×340 pixels with a spatial resolution of 1.3 m. A three-band false-color composite image and the ground-truth map are shown in Fig. 3(c) and (d), respectively. In the ground-truth map, there are classes of land covers with more than 1000 labeled pixels for each class. The number of pixels for training and testing, as well as the total number of pixels in each class, is listed in Table II.
- 3) Kennedy Space Center (KSC) dataset: The third dataset was acquired by the AVIRIS sensor over Kennedy Space Center (KSC), Florida, USA, on March 23, 1996. It contains 224 spectral bands. We utilize 176 bands of them after removing bands with water absorption and low signal-to-noise ratio. The spatial size of the image is 512×614 pixels, and the spatial resolution is 18 m. Discriminating different land covers in this dataset is difficult due to the similarity of spectral signatures among certain vegetation

TABLE II
NUMBERS OF PIXELS FOR THE TRAINING/TESTING AND THE TOTAL NUMBER OF PIXELS IN EACH CLASS IN THE PUS GROUND-TRUTH MAP

Class		Sample			Class		Sample		
Label	Name	Total	Training	Testing	Label	Name	Total	Training	Testing
C1	Asphalt	6641	548	6093	C6	Bare Soil	5029	532	4497
C2	Meadows	18649	540	18109	C7	Bitumen	1330	375	955
C3	Gravel	2099	392	1707	C8	Self-Blocking Bricks	3682	514	3168
C4	Trees	3064	524	2540	C9	Shadows	947	231	716
C5	Painted metal sheets	1345	265	1080					

TABLE III
NUMBERS OF PIXELS FOR THE TRAINING/TESTING AND THE TOTAL NUMBER OF PIXELS IN EACH CLASS IN THE KSC GROUND-TRUTH MAP

Class		Sample			Class		Sample		
Label	Name	Total	Training	Test	Label	Name	Total	Training	Test
C1	Scrub	761	76	685	C8	Graminoid marsh	431	43	388
C2	Willow swamp	243	24	219	C9	Spartina marsh	520	52	468
C3	Cabbage palm hammock	256	26	230	C10	Cattail marsh	404	40	364
C4	Cabbage palm/oak hammock	252	25	227	C11	Salt marsh	419	42	377
C5	Slash pine	161	16	145	C12	Mud flats	503	50	453
C6	Oak/broadleafhammock	229	23	206	C13	Water	927	93	834
C7	Hardwood swamp	105	11	94					

types. For classification purposes, 13 classes representing the various land-cover types that occur in this environment are defined. Fig. 3(e) and (f) show a three-band false-color composite image and the ground-truth map, respectively. Table III reports the number of totally labeled pixels and the number of pixels for training and testing in each class.

B. Experimental Setup

We compare the proposed SLGDE and KSLGDE models with several DR methods, including PCA, LDA, spectral-spatial LDA (SSLDA) [41], locality preserving projection (LPP) [18], CGDE [28], SGDE [21], and LGDE [30]. Besides, we also directly use the raw pixels as a benchmark. For LDA and SSLDA, the reduced dimensionality is $NC - 1$, where NC represents the number of classes in each dataset. For the other DR algorithms, the reduced dimensionality ranges from 5 to 50 with a stepsize 5. For CGDE, the optimal regularization parameter in collaboration representation is selected from $\{0.01, 0.1, 0.1, 1, 10\}$. For SGDE, the number of nonzero atoms is empirically set to 50, because it can achieve a satisfying result. For SLGDE and KSLGDE algorithms, the optimal number of superpixels L is chosen from a candidate set $\{100, 500, 1000, 1500, 2000, 2500\}$, while the optimal regularization parameter λ is selected from $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$. For simplicity, we empirically set $K_w = K_b = K_s = 5$ and $t = 1$. For KSLGDE, we choose Gaussian radial basis kernel as the kernel function, and empirically set its parameter to 1.

For the IP and KSC datasets, we randomly select 10% pixels from each class as the training set, and use the remaining pixels as the testing set, which are demonstrated in Tables I and III, respectively. For the PUS dataset, as shown in Table II, we randomly choose 3921 pixels as the training set and the rest of the pixels as the testing set. The training set is used to learn the

mapping functions for all of the DR methods. For classification purposes, the extracted features are fed into the nearest neighbor classifier.

In order to reduce the effects of random selection, all the algorithms are repeated ten times and the average results are reported. The classification performance is evaluated by the overall accuracy (OA), the average accuracy (AA), the per-class accuracy, and the Kappa coefficient κ . OA defines the ratio between the number of correctly classified pixels to the total number of pixels in the testing set, AA refers to the average of accuracies in all classes, and κ is the percentage of agreement corrected by the number of agreements that would be expected purely by chance.

C. Parameter Selection

There exist two important parameters in the proposed SLGDE and KSLGDE methods. They are the regularization parameter λ and the number of superpixels L . To evaluate their effects on the classification performance, we fix the other parameters and change λ from 10^{-3} to 10, and L from 100 to 2500. The reduced dimensionality is set to 30 and the bold values correspond to the best results. The changes of OAs achieved by SLGDE and KSLGDE on the IP dataset are shown in Fig. 4, where the x -axis represents λ , the y -axis represents L , and the z -axis represents OA. From this figure, it can be observed that as λ and L increase, OA initially increases and then decreases. Therefore, the optimal parameter values on the IP dataset are set as $\lambda = 0.1$, $L = 500$ for SLGDE and $\lambda = 0.1$, $L = 2000$ for KSLGDE, because they can achieve the maximal OA value. Similarly, Figs. 5 and 6 demonstrate the changes of OAs on the PUS dataset and the KSC dataset, respectively. From these three-dimensional diagrams, we can find the maximal OA value and set the corresponding parameter values as the optimal ones. Specifically, the optimal

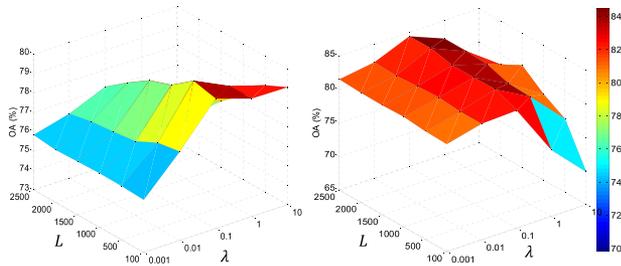


Fig. 4. OAs of SLGDE (left) and KSLGDE (right) with different L and λ values on the IP dataset.

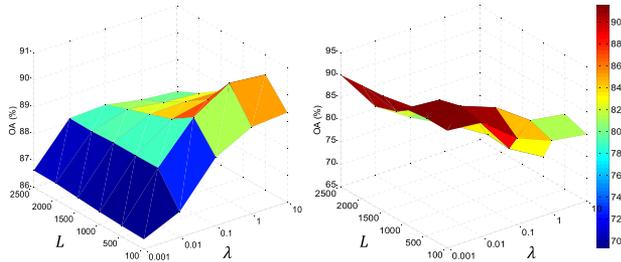


Fig. 5. OAs of SLGDE (left) and KSLGDE (right) with different L and λ values on the PUS dataset.

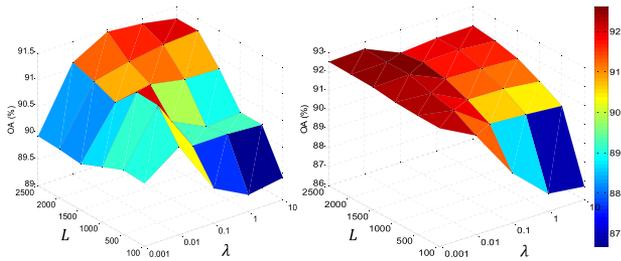


Fig. 6. OAs of SLGDE (left) and KSLGDE (right) with different L and λ values on the KSC dataset.

values are $\lambda = 1, L = 500$ for SLGDE and $\lambda = 10^{-2}, L = 500$ for KSLGDE on the PUS dataset. Besides, the optimal values are $\lambda = 10^{-2}, L = 500$ for SLGDE and $\lambda = 10^{-2}, L = 1500$ for KSLGDE on the KSC dataset.

D. Performance Comparison

To demonstrate the effectiveness of the proposed SLGDE and KSLGDE models, we quantitatively and qualitatively compare it with the aforementioned DR methods. Fig. 7 shows OA versus reduced dimensionality using different methods on the IP dataset. Several conclusions can be observed from this figure. First of all, as reduced dimensionality increases, OAs achieved by all methods will initially increase and then tend to be stable, because reducing the dimensionality of the original data into a much lower one may lose useful information for classification. Second, compared to LPP and CGDE, SGDE can quickly reach its stable state and achieve similar OA to RAW using only ten components. This can be explained by that sparse representation is capable of learning discriminative features and the adjacency relationships when constructing the graph model. Third, com-

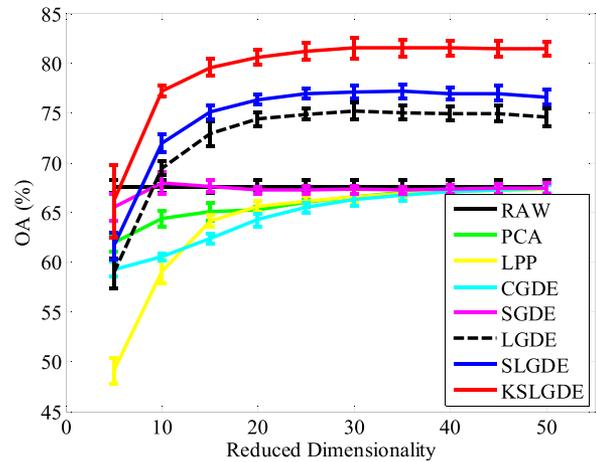


Fig. 7. OA versus reduced dimensionality using different methods on the IP dataset.

pared to SGDE, LGDE achieves higher OAs especially when the reduced dimensionality is larger than ten, because SGDE only explores the information from intraclass pixels to construct the graph model, while LGDE attempts to make use of the information from intraclass and interclass pixels simultaneously, leading to a more discriminative embedding. It is worth noting that the stable OA value of LGDE is obviously larger than RAW. This indicates that the spectral features of HSIs are highly correlated, and an effective DR method can extract the most discriminative features from them. Finally, by incorporating the spatial information into LGDE, SLGDE and KSLGDE can obtain superior performance as compared to LGDE given the same reduced dimensionality, which sufficiently certify the effectiveness of our proposed methods. Besides, due to the use of kernel methods, KSLGDE is able to capture the nonlinear distribution property of an HSI, thus significantly improving the performance of SLGDE.

To further analyze the results in detail, Table IV reports OA, AA, per-class accuracy, and κ achieved by different methods when the reduced dimensionality is 30. From this table, we can observe that KSLGDE and SLGDE achieve the highest value in 12 classes. Specifically, SLGDE can improve OA by about 2% compared to the best comparison method. As an extension of SLGDE, KSLGDE can further boost OA by 4% in comparison with SLGDE. Fig. 8 shows classification maps on the IP dataset using ten different methods. In this figure, different colors correspond to different classes of land covers. To facilitate easy comparisons between methods, only areas for which we have ground-truth are shown in these maps. These maps are consistent with the results listed in Table IV. Due to the use of spatial information, the classification maps of SLGDE and KSLGDE are much smoother than other methods.

Figs. 9 and 10 demonstrate OA versus reduced dimensionality using different methods on the PUS dataset and the KSC dataset, respectively. Again, all methods obtain increasing OAs initially and then get their stable values as the reduced dimensionality increases. Compared to other methods, SLGDE and KSLGDE can achieve larger stable values. Tables V and VI quantitatively report the performance of different methods on the PUS dataset

TABLE IV
OA, AA, PER-CLASS ACCURACY (%), κ , AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE IP DATASET

Label	RAW	PCA	LDA	SSLDA	LPP	CGDE	SGDE	LGDE	SLGDE	KSLGDE
C1	42.92±3.27	42.92±9.69	33.17±13.20	51.71±19.69	46.34±9.60	43.90±9.13	40.49±8.89	20.98±5.62	19.02±9.19	19.44±12.19
C2	54.53±3.03	53.76±3.24	60.16±3.05	65.99±1.18	54.52±2.84	54.32±1.80	55.02±2.31	65.23±2.48	67.50±1.92	75.13±3.20
C3	51.03±1.93	50.14±1.58	49.99±1.39	50.47±1.84	50.28±1.93	51.99±1.11	51.14±1.32	55.88±1.88	59.57±1.85	67.79±3.11
C4	40.00±4.95	38.40±6.39	46.57±4.82	42.91±8.38	38.12±4.65	39.62±5.20	39.62±5.02	47.61±5.79	50.14±6.01	51.27±6.95
C5	83.03±1.82	80.13±1.93	86.34±2.20	87.49±1.95	78.30±2.40	78.71±3.19	81.24±3.27	85.84±3.20	87.03±1.80	88.97±2.34
C6	91.66±1.43	88.86±1.77	93.64±1.93	94.28±1.39	88.83±1.23	88.46±1.99	89.47±1.16	95.95±1.79	97.35±1.44	96.16±1.32
C7	76.80±5.21	75.20±3.35	31.20±11.80	57.60±24.59	75.20±5.21	83.20±8.67	77.60±4.56	73.60±20.51	72.00±20.98	48.00±19.18
C8	94.32±2.27	93.39±2.58	99.02±1.21	97.67±1.23	92.93±2.36	93.63±2.56	94.09±2.44	98.51±0.63	99.49±0.25	98.19±0.60
C9	16.67±11.11	16.67±13.03	6.67±6.09	16.67±13.03	16.67±11.79	20.00±9.30	16.67±10.39	21.11±13.26	16.67±12.42	14.44±6.33
C10	58.58±3.92	57.65±3.00	54.58±2.57	63.73±3.64	58.15±3.26	61.19±4.48	59.13±2.84	59.91±3.81	61.42±4.48	70.29±1.86
C11	70.68±1.05	70.77±0.97	65.72±1.96	71.46±1.56	70.63±1.06	71.41±1.37	70.99±0.99	78.46±1.51	80.54±1.28	88.02±1.52
C12	40.49±1.25	38.39±1.23	65.02±2.30	57.64±4.17	38.13±1.58	37.57±1.79	39.48±1.96	65.13±3.19	69.66±3.16	70.64±3.26
C13	92.93±3.33	91.09±3.56	96.30±3.13	90.76±5.63	91.63±2.51	92.39±3.00	91.52±3.80	96.20±3.61	97.39±2.56	97.39±1.41
C14	87.52±0.96	86.92±1.06	91.30±1.15	93.88±0.94	86.73±1.14	86.26±0.58	86.94±0.96	96.31±1.08	96.98±1.17	93.50±1.42
C15	36.71±3.12	33.31±3.72	56.89±1.03	54.54±5.39	34.35±3.27	33.54±3.17	35.73±3.63	48.36±6.77	50.37±6.47	65.25±7.06
C16	87.14±3.61	87.14±4.24	81.67±8.10	88.57±2.74	87.14±4.24	86.90±3.76	86.90±4.21	85.95±6.70	85.71±6.84	80.71±5.42
AA	67.59±0.61	66.57±0.37	69.87±0.50	72.95±0.92	66.61±0.62	66.28±0.61	67.33±0.38	75.18±0.85	77.12±0.62	81.50±0.77
OA	64.06±1.06	62.80±1.54	63.64±1.62	67.84±2.92	63.00±1.48	62.51±1.07	63.50±1.16	68.44±2.25	69.43±2.36	70.27±1.38
κ	62.94±0.70	61.76±0.42	65.56±0.58	69.05±1.04	61.81±0.71	61.41±0.68	62.65±0.44	71.48±0.97	73.71±0.72	78.70±0.88

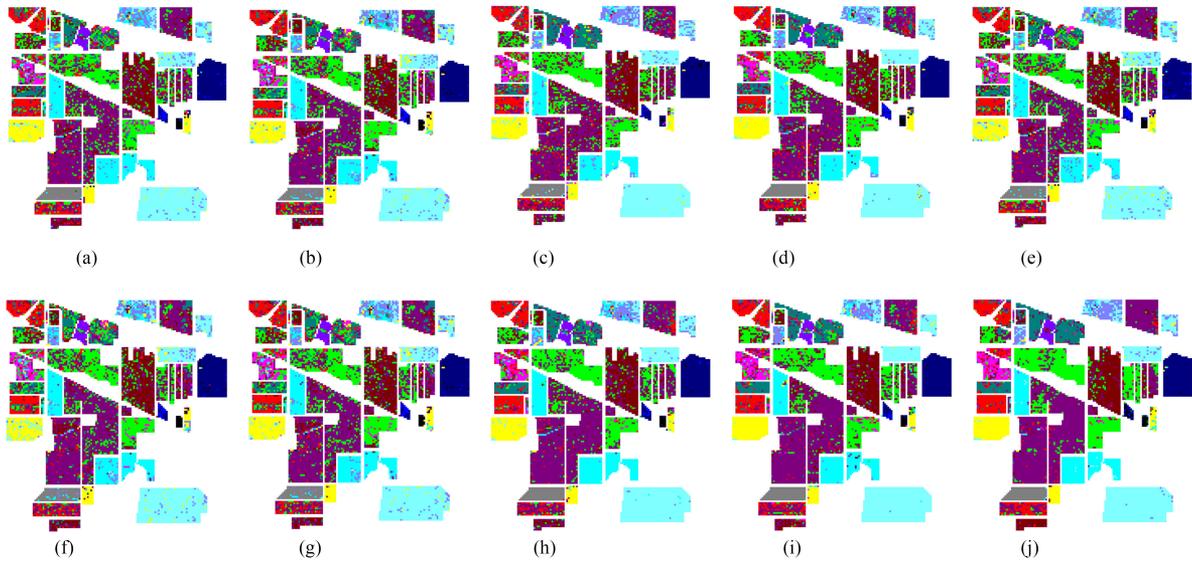


Fig. 8. Classification maps on the IP dataset using ten different methods. (a) RAW. (b) PCA. (c) LDA. (d) SSLDA. (e) LPP. (f) CGDE. (g) SGDE. (h) LGDE. (i) SLGDE. (j) KSLGDE.

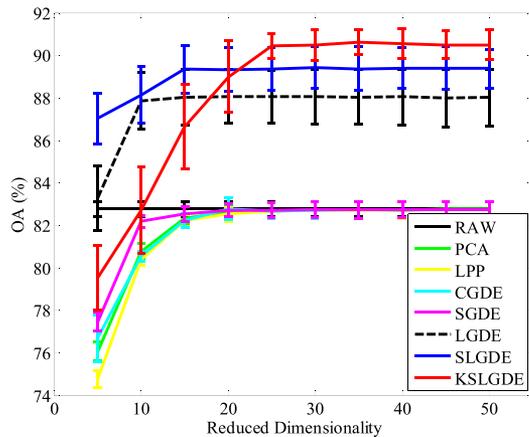


Fig. 9. OA versus reduced dimensionality using different methods on the PUS dataset.

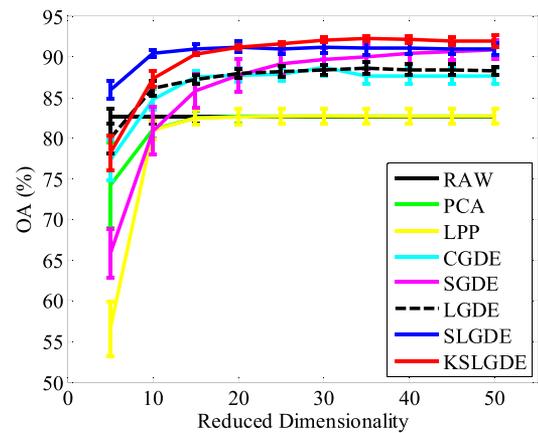


Fig. 10. OA versus reduced dimensionality using different methods on the KSC dataset.

and the KSC dataset, respectively. In these tables, the reduced dimensionality is set to 30 and the bold values correspond to the best results. Similar to Table IV, SLGDE and KSLGDE can achieve the best results in most classes. Besides, Figs. 11 and 12

show classification maps on the PUS and the KSC datasets, respectively, using ten different methods. All the observations can validate the effectiveness of the proposed SLGDE and KSLGDE methods on the PUS and KSC datasets.

TABLE V

OA, AA, PER-CLASS ACCURACY (%), κ , AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE PUS DATASET

Label	RAW	PCA	LDA	SSLDA	LPP	CGDE	SGDE	LGDE	SLGDE	KSLGDE
C1	80.52±1.10	80.55±1.15	83.37±1.21	83.87±0.81	80.59±1.14	80.55±1.13	80.60±1.11	84.59±1.63	83.23±1.71	85.00±1.45
C2	81.89±0.93	81.78±0.91	89.89±0.97	89.62±0.90	81.75±0.93	81.74±0.89	81.79±0.89	88.12±1.95	90.60±1.10	93.69±1.75
C3	74.54±0.72	74.76±0.78	74.31±1.39	74.55±1.58	74.84±0.81	74.77±0.88	74.86±0.80	80.16±1.38	80.59±1.46	84.03±0.95
C4	93.80±0.58	93.76±0.59	95.98±0.91	94.79±0.56	93.72±0.61	93.76±0.57	93.71±0.59	96.43±0.21	96.97±0.51	92.53±2.18
C5	99.20±0.14	99.20±0.14	99.91±0.09	99.98±0.04	99.20±0.14	99.20±0.14	99.20±0.14	99.72±0.09	99.78±0.12	91.61±1.07
C6	79.36±1.39	79.50±1.32	89.90±1.45	88.63±0.85	79.50±1.32	79.48±1.26	79.52±1.30	87.48±1.47	91.24±1.08	87.12±2.68
C7	89.70±1.22	89.57±1.01	89.82±1.78	89.91±1.68	89.53±1.02	89.55±1.05	89.49±1.05	92.73±1.22	93.03±1.00	92.71±2.06
C8	80.88±0.51	80.81±0.71	77.52±1.06	77.32±1.17	80.78±0.68	80.86±0.66	80.89±0.51	84.55±1.15	83.88±1.78	86.30±1.46
C9	100.00±0.00	100.00±0.00	99.86±0.00	99.89±0.06	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	99.94±0.08	99.11±0.31
OA	82.76±0.36	82.73±0.39	88.04±0.18	87.75±0.48	82.72±0.40	82.72±0.40	82.75±0.37	88.05±1.27	89.43±0.99	90.48±0.72
AA	86.65±0.15	86.66±0.18	88.95±0.38	88.73±0.41	86.66±0.18	86.66±0.20	86.67±0.17	90.42±0.67	91.03±0.66	90.23±0.44
κ	77.14±0.42	77.11±0.48	83.94±0.21	83.62±0.62	77.09±0.48	77.09±0.48	77.13±0.45	84.01±1.64	85.82±1.30	87.08±0.93

TABLE VI

OA, AA, PER-CLASS ACCURACY (%), κ , AND STANDARD DEVIATIONS AFTER TEN RUNS PERFORMED BY TEN DIFFERENT METHODS ON THE KSC DATASET

Label	RAW	PCA	LDA	SSLDA	LPP	CGDE	SGDE	LGDE	SLGDE	KSLGDE
C1	90.39±0.38	90.36±0.36	92.12±1.48	93.20±1.90	90.39±0.38	90.37±0.68	93.26±1.83	93.52±2.57	96.15±1.71	96.23±1.22
C2	78.45±6.74	78.45±6.74	88.77±5.21	91.23±2.71	78.45±6.74	86.30±5.80	89.04±2.14	84.20±4.46	87.31±3.93	87.40±4.52
C3	74.61±4.76	74.17±4.48	78.09±5.70	78.52±2.97	74.35±4.04	75.91±3.94	77.13±6.18	73.39±3.54	77.39±6.19	86.52±3.18
C4	46.43±1.45	46.52±1.48	55.68±6.63	60.44±6.61	46.43±1.58	61.85±8.53	64.76±7.56	57.89±9.12	65.11±6.80	72.60±4.61
C5	44.83±3.12	44.83±3.12	63.86±7.89	64.97±4.00	44.97±3.25	66.76±2.94	64.83±4.28	61.38±7.65	65.79±6.20	68.00±5.53
C6	41.94±3.92	41.84±3.86	60.78±5.04	70.87±6.58	41.84±3.86	53.79±10.22	58.06±9.57	49.51±9.71	63.20±7.20	62.23±7.11
C7	63.19±14.69	63.19±14.69	77.45±12.39	78.72±6.06	63.40±14.46	79.57±9.32	75.96±12.24	82.13±10.73	75.74±12.57	80.85±11.08
C8	76.39±3.33	76.55±3.42	92.94±1.98	88.25±2.83	76.55±3.37	92.89±1.51	91.08±2.52	86.70±1.46	89.48±1.82	91.65±1.92
C9	90.90±2.63	90.85±2.69	98.08±0.52	96.20±1.44	90.81±2.70	97.69±0.70	97.56±0.29	93.93±2.73	95.68±1.59	95.73±1.09
C10	83.13±1.66	83.24±1.61	95.60±1.55	94.78±1.15	83.24±1.61	94.45±1.16	94.40±1.70	97.14±1.21	99.34±0.57	97.14±1.04
C11	95.49±0.73	95.49±0.73	97.72±1.00	96.13±1.73	95.49±0.73	92.04±2.71	96.82±1.29	96.66±0.87	97.24±1.02	97.24±0.89
C12	85.47±5.20	85.47±5.25	93.95±2.34	97.13±1.51	85.47±5.25	92.05±2.20	92.54±1.93	93.91±1.40	98.76±0.40	97.79±0.54
C13	99.26±0.87	99.26±0.87	99.23±0.80	99.90±0.10	99.26±0.87	99.26±0.81	99.14±0.74	99.90±0.05	99.95±0.11	99.93±0.16
OA	82.69±0.99	82.68±0.96	89.79±1.03	90.48±0.39	82.70±0.97	88.67±1.04	89.72±1.34	88.40±0.65	91.18±0.76	92.06±0.27
AA	74.65±1.29	74.63±1.23	84.17±1.90	85.41±0.78	74.67±1.19	83.30±1.81	84.20±2.44	82.33±1.53	85.47±1.10	87.18±0.42
κ	80.71±1.10	80.70±1.08	88.63±1.15	89.39±0.44	80.72±1.09	87.37±1.16	88.54±1.50	87.07±0.72	90.17±0.84	91.14±0.30

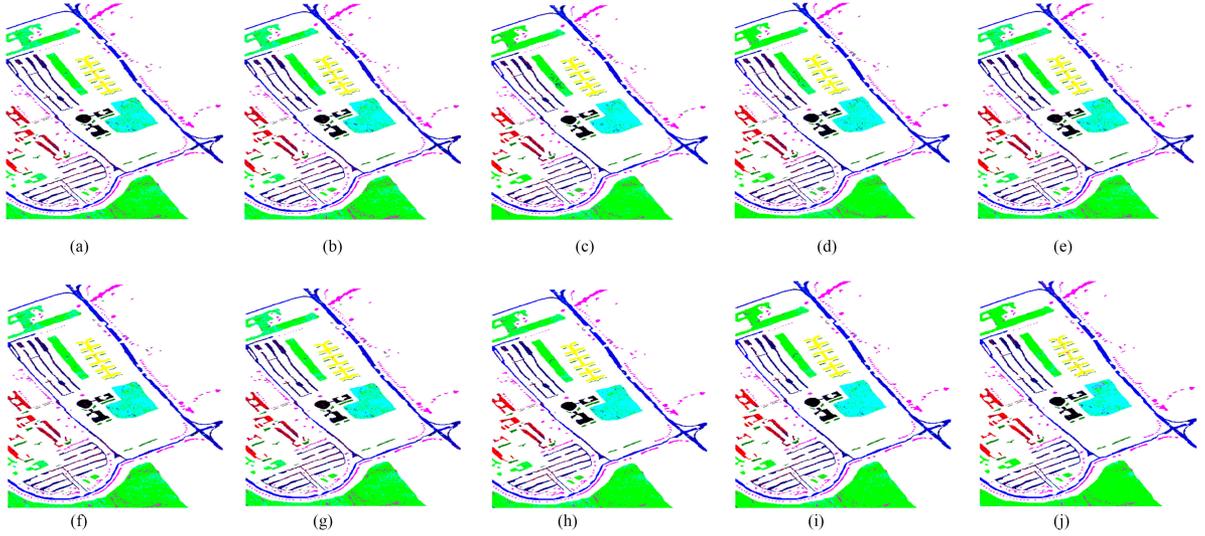


Fig. 11. Classification maps on the PUS dataset using ten different methods. (a) RAW. (b) PCA. (c) LDA. (d) SSLDA. (e) LPP. (f) CGDE. (g) SGDE. (h) LGDE. (i) SLGDE. (j) KSLGDE.

E. Computational Complexity

As shown in (5), the solution of the proposed SLGDE algorithm is to compute the eigenvectors of $(XL'X^T)^{-1}(XLX^T + \frac{\lambda}{L} \sum_{\ell=1}^L X_\ell \tilde{L}_\ell X_\ell^T)$. Assume the dimension of X is $d \times m$, then the computation complexity of $(XL'X^T)^{-1}$ is $O(dm^2 + d^3)$. Since the number of pixels in each superpixel is generally smaller than that of training pixels, the computation complexity of $XLX^T + \frac{\lambda}{L} \sum_{\ell=1}^L X_\ell \tilde{L}_\ell X_\ell^T$ is $O(dm^2)$. Besides, the

computation complexity of the eigenvalue decomposition problem is $O(d^3)$. Therefore, the final complexity of the proposed SLGDE is $O(dm^2 + d^3)$. Similarly, we can derive that the computation complexity of the proposed KSLGDE is $O(m^3)$. In general, the number of training pixels is significantly larger than that of spectral bands (i.e., $m \gg d$), so KSLGDE costs much more time than SLGDE.

To quantitatively compare the computation complexity of different methods, we conduct the aforementioned experiments on

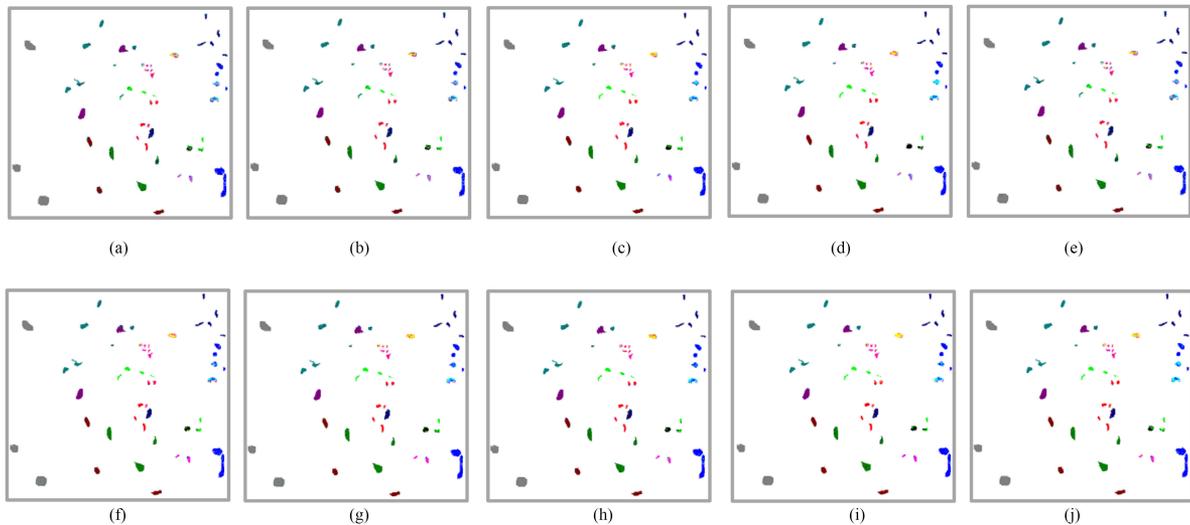


Fig. 12. Classification maps on the KSC dataset using ten different methods. (a) RAW. (b) PCA. (c) LDA. (d) SSLDA. (e) LPP. (f) CGDE. (g) SGDE. (h) LGDE. (i) SLGDE. (j) KSLGDE.

TABLE VII
COMPUTATIONAL TIME (IN SECONDS) OF TEN DIFFERENT METHODS ON THREE DATASETS

Data	RAW	PCA	LDA	SSLDA	LPP	CGDE	SGDE	LGDE	SLGDE	KSLGDE
IP	1.41	0.39	0.42	808.10	0.61	49.14	233.22	0.88	3.22	44.87
PUS	10.78	3.36	0.39	385.05	5.66	277.56	2293.21	8.26	25.22	728.44
KSC	0.35	0.22	0.20	83.51	0.29	44.15	63.11	0.46	38.32	99.18

a computer with an Intel Core i7-4790 CPU 3.60 GHz and 32 GB RAM. The software implementation is performed using MATLAB. Table VII demonstrates the computational time of different models on three datasets. It is worth noting that besides LDA and SSLDA, the reduced dimensionality of other models is set to 30. From this table, we can observe that the proposed SLGDE model is faster than SSLDA, CGDE, and SGDE, but is slower than the other comparison models. Besides, due to the computation of kernel function, KSLGDE is slower than SLGDE.

IV. CONCLUSION

In this paper, we have proposed an SLGDE model and its kernel counterpart KSLGDE for DR and classification of HSIs. To make full use of the spatial information, the original HSI was first divided into nonoverlapping superpixels, which can adjust their sizes and shapes according to local structures. Then, for each superpixel, an intraclass graph was constructed. These constructed intraclass graphs can be naturally incorporated into the LGDE model, resulting in the SLGDE model. Besides, considering the nonlinear distribution property of HSIs, SLGDE was further extended to the KSLGDE model. To validate the effectiveness of the proposed models, we compared them with several DR methods on three widely used HSIs. The obtained results demonstrate that SLGDE obtains superior performance than the comparison methods, and KSLGDE can further improve the performance of SLGDE. Additionally, we also thoroughly evaluated the effects of different parameters on the classification performance of SLGDE and KSLGDE, including the regularization parameter λ and the number of superpixels L .

REFERENCES

- [1] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
- [2] R. Hang, Q. Liu, H. Song, and Y. Sun, "Matrix-based discriminant subspace ensemble for hyperspectral image spatial-spectral feature fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 783–794, Feb. 2016.
- [3] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.
- [4] 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 Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.
- [5] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.
- [6] Y. Fang *et al.*, "Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1712–1716, Oct. 2014.
- [7] D. Liao, Y. Qian, J. Zhou, and Y. Y. Tang, "A manifold alignment approach for hyperspectral image visualization with natural color," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3151–3162, Jun. 2016.
- [8] C. Vaiphasa, A. K. Skidmore, W. F. de Boer, and T. Vaiphasa, "A hyperspectral band selector for plant species discrimination," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 3, pp. 225–235, 2007.
- [9] 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, Oct. 2008.
- [10] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surveying Land Inf. Sci.*, vol. 62, no. 2, pp. 115–122, 2002.
- [11] 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.
- [12] M. Imani and H. Ghassemian, "Feature space discriminant analysis for hyperspectral data feature reduction," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 1–13, 2015.
- [13] R. Hang *et al.*, "Robust matrix discriminative analysis for feature extraction from hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2002–2011, May 2017.

- [14] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.
- [15] 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.
- [16] L. Ma, M. M. Crawford, X. Yang, and Y. Guo, "Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2832–2844, May 2015.
- [17] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. in Neural Inf. Process. Syst.*, 2002, pp. 585–591.
- [18] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 153–160.
- [19] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, 2005, vol. 2, pp. 1208–1213.
- [20] Y. Gao, R. Ji, P. Cui, Q. Dai, and G. Hua, "Hyperspectral image classification through bilayer graph-based learning," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2769–2778, Jul. 2014.
- [21] 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.
- [22] 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.
- [23] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [24] H. Huang, F. Luo, J. Liu, and Y. Yang, "Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding," *ISPRS J. Photogramm. Remote Sens.*, vol. 106, pp. 42–54, 2015.
- [25] Z. Xue, P. Du, J. Li, and H. Su, "Simultaneous sparse graph embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6114–6133, Nov. 2015.
- [26] W. He, H. Zhang, L. Zhang, W. Philips, and W. Liao, "Weighted sparse graph based dimensionality reduction for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 686–690, May 2016.
- [27] 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, Jul. 2016.
- [28] 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.
- [29] 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.
- [30] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 2, pp. 846–853.
- [31] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
- [32] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
- [33] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
- [34] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [35] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
- [36] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
- [37] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.
- [38] L. O. Jiménez, J. L. Rivera-Medina, E. Rodríguez-Díaz, E. Arzuaga-Cruz, and M. Ramírez-Vélez, "Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 844–851, Apr. 2005.
- [39] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.
- [40] X. Jia and J. A. Richards, "Managing the spectral-spatial mix in context classification using Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 311–314, Apr. 2008.
- [41] H. Yuan, Y. Y. Tang, Y. Lu, L. Yang, and H. Luo, "Spectral-spatial classification of hyperspectral image based on discriminant analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2035–2043, Jun. 2014.
- [42] G. Zhang, X. Jia, and J. Hu, "Superpixel-based graphical model for remote sensing image mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5861–5871, Nov. 2015.
- [43] 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, Oct. 2015.
- [44] 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, Aug. 2015.
- [45] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 2097–2104.
- [46] 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, Mar. 2005.
- [47] D. G. Goodin, J. Gao, and G. M. Henebery, "The effect of solar illumination angle and sensor view angle on observed patterns of spatial structure in tallgrass prairie," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 154–165, Jan. 2004.
- [48] D. Roberts, M. Smith, and J. Adams, "Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data," *Remote Sens. Environ.*, vol. 44, no. 2/3, pp. 255–269, May/Jun. 1993.
- [49] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [50] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.



Renlong Hang (M'17) received the Ph.D. degree in meteorological information technology from the Nanjing University of Information Science and Technology, Nanjing, China, in 2017.

He is currently a Lecturer with the School of Information and Control, Nanjing University of Information Science and Technology. His research interests include machine learning and pattern recognition.



Qingshan Liu (M'05–SM'07) received the M.S. degree from Southeast University, Nanjing, China, in 2000, and the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 2003.

From June 2004 to April 2005, he was an Associate Researcher with the Multimedia Laboratory, The Chinese University of Hong Kong, Hong Kong. Then, he was an Associate Professor with the National Laboratory of Pattern Recognition, Chinese Academy of Sciences. From 2010 to 2011, he was an Assistant Research Professor with the Department of Computer Science, Computational Biomedicine Imaging and Modeling Center, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. He is currently a Professor with the School of Information and Control, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include image and vision analysis.

Dr. Liu received the President Scholarship of the Chinese Academy of Sciences in 2003.