Neurocomputing 360 (2019) 138–150

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Superpixel-based spatial-spectral dimension reduction for hyperspectral imagery classification

Huilin Xu[a], Hongyan Zhang[a,*], Wei He[b,*], Liangpei Zhang[a]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China
[b] Geoinformatics Unit, RIKEN AIP, Japan

ARTICLE INFO

Article history:
Received 14 February 2019
Revised 24 May 2019
Accepted 4 June 2019
Available online 14 June 2019

Communicated by Dr. Lu Xiaoqiang

Keywords:
Hyperspectral image
Dimension reduction
Spatial-spectral
Superpixel

ABSTRACT

Dimension reduction (DR) is a useful preprocessing technology for hyperspectral image (HSI) classification. This paper presents an HSI DR method named superpixel-based spatial-spectral dimension reduction (SSDR), which integrates the spatial and spectral similarity. The HSI is first segmented into non-overlapping superpixels, where pixels belonging to the same superpixel have strong correlations, and should be preserved after DR. We then apply the superpixel-based linear discriminant analysis (SPLDA) method, which learns a superpixel-guided graph to capture the spatial similarity. Pixels from the same label also have strong spectral correlations; thereby, we also construct a label-guided graph to explore the spectral similarity. These two graphs are finally integrated to learn the discriminant projection. The classification results on two widely used HSIs demonstrate the advantage of the proposed algorithms compared to the other state-of-the-art DR methods.

1. Introduction

Hyperspectral images (HSIs) can be simultaneously acquired with diagnostic information in the spectral domain and land-cover distribution information in the spatial domain. Compared with multispectral images, HSIs have a better ability to distinguish different materials, which has been extensively exploited in many applications, such as classification [1–3], clustering [4], target detection [5,6], data unmixing [7–9], anomaly detection [10], and change detection [11,12]. Besides, denoising is an active research topic as a preprocessing technology of HSI applications [13]. Among these applications, classification has attracted a lot of attention in the last decades [14], e.g., in military surveillance, urban and rural planning, land-use and land-cover monitoring, and environmental disaster assessment [15]. However, HSI classification becomes a challenging task because of the high dimensions and the limited number of training samples, which can cause the Hughes phenomenon (the curse of dimensionality) [16,17]. In the meantime, the high dimensionality has brought great pressure to the transmission and storage, which slows down the efficiency of the data processing and increases the computational burden. Therefore, DR has become the key preprocessing technology in many information processing tasks [18–21].

As a preprocessing step before classification, hyperspectral dimension reduction (DR) technology mainly aims at extracting features from the original image for classification, without sacrificing important information. It can then search for a low-dimensional data space to effectively express the high-dimensional data characteristics. DR helps to: (1) eliminate the redundancy between features and improve the computational efficiency; (2) ameliorate the statistical ill-posed problem caused by the limited number of samples; and (3) prevent the Hughes phenomenon and improve the classification accuracy.

Popular DR techniques can be generally categorized as feature selection [22–24] and feature extraction methods. Feature extraction methods are dominant in DR, and can be categorized into local- and global-based methods. Local methods are designed to retain the structural information of local regions, whereas the global methods need to learn the optimal projection under the constraint of all the sample pairs. Local methods include local linear embedding (LLE) [25], local discriminant embedding (LDE) [26], regularized LDE (RLDE) [27], block sparse graph-based discriminant analysis (BSGDA) [28], locality preserving projections (LPP) [29], and so on. LLE is a nonlinear DR algorithm, which is reconstructed by neighboring samples, using the affine reconstruction coefficient to characterize the local adjacency relationship of the sample set. LDE tries to embed the submanifold of each class by optimizing a unified problem. RLDE extends LDE by adding constraints

* Corresponding author.
E-mail addresses: xuhl@whu.edu.cn (H. Xu), zhanghongyan@whu.edu.cn (H. Zhang), wei.he@riken.jp (W. He).

https://doi.org/10.1016/j.neucom.2019.06.023
0925-2312/

that preserve the data diversity. Sparsity graphs have also been applied in the classification of HSIs. For instance, BSGDA is one of the sparsity-preserving graph construction methods, which incorporates class-label information for the supervised discriminant analysis. LPP learns to optimally preserve the neighborhood structure of the input data via a linear projective map. Linear discriminant analysis (LDA) [30], principal component analysis (PCA) [31], and isometric feature mapping (ISOMAP) [32] are typical global DR methods. PCA is one of the most popular DR algorithms, which searches for an orthogonal projection to maximize the variance of the input feature that maintains the overall distribution of the data characteristics during the DR process. LDA aims at finding a projection by maximizing the between-class covariance while minimizing the within-class covariance, so that the similar samples are closer and the heterogeneous samples are further separated in the subspace. ISOMAP is based on classical multidimensional scaling (MDS), and it explores the inherent geometric structure of the data set by establishing the geodesic distances between all the pairs of data points. These methods can also be generally classified into supervised (LDA, LDE, RLDE, and BSGDA), unsupervised (LLE, ISOMAP, PCA, and LPP), and semisupervised methods, according to the *a priori* information [33–35].

To sum up, most of these methods are spectral-based, and they only take the spectral information into account [36,37]. However, the spectral information of one specific pixel from one class can be easily mistaken for that of pixels from other classes, which may cause the classification results to appear very noisy [38]. As we know, there is a high probability that nearby pixels will have similar spectral signatures and belong to the same material. Thereby, many HSI classification methods utilize a combined spatial-spectral prior, and have achieved state-of-the-art results [39,40]. From one aspect, the spatial information can be utilized in the classification stage. For example, Benediktsson et al. [41] and Li et al. [42] used extended morphological profiles to explore the spatial information; Fang et al. [38] incorporated multiple kernels into a support vector machine (SVM) classifier, which accounts for the exploration of the spectral and spatial information within superpixels; and Zhou et al. [43] proposed two spatial-spectral composite kernel extreme learning machine (ELM) classification methods, which inherit the advantages of ELM. From another aspect, the postprocessing procedure can also include spatial information exploration, as introduced by Fang et al. [44] and Jiay et al. [45].

Several spatially regularized methods have also been introduced to DR. Fang et al. [46] improved the LLE method by using a new similarity measurement based on both the spectral and spatial features. In addition, the spatial-spectral contextual similarity measure [47] was proposed to preserve the pixel patchwise similarity between the observed pixels and their spatial neighbors. A spatial regularizer was adopted by Ma et al. [48] to enhance the similarity of the manifold coordinates of spatially surrounding pixels. Feng et al. [49] used heterogeneous and homogeneous spectral-spatial neighbors of hyperspectral pixels to define the discriminate spectral-spatial margins (DSSMs), and explored both local and global information of hyperspectral data by maximizing the DSSM of hyperspectral data and casting a low-rank representation (LRR) regularizer. Huang et al. [50] proposed an unsupervised dimensionality reduction algorithm called spatial-spectral manifold reconstruction preserving embedding (SSMRPE), which utilizes the spatial-spectral combined distance (SSCD) for selecting effective spatial-spectral neighbors of HSI pixels to construct a graph. In [27], Zhou et al. proposed a local pixel neighborhood preserving embedding (LPNPE) algorithm, which considers that the pixels from one fixed-size window neighbor will have strong similarity and can be adopted to learn a spatial graph embedding. However, the fixed-size window cannot capture the spatial correlations effectively. It seems that a shape-adaptive window has a better ability to exploit the spatial context, and can be adapted according to the different structures.

In view of this, this paper proposes the superpixel-based linear discriminant analysis (SPLDA) method to construct a superpixel-guided graph, where the superpixel spatial regularizer constrains the spatial neighbors with high spectral similarity to preserve the spatial similarity after DR. Moreover, the superpixel-based spatial-spectral dimension reduction (SSDR) method is proposed to integrate the superpixel-guided graph with the label-guided graph for HSI DR. Specifically, the simple linear iterative clustering (SLIC) [51] superpixel segmentation algorithm is first utilized to segment the HSI into non-overlapping superpixels, and the superpixel-guided graph is then constructed to capture the spatial similarity. Subsequently, a label-guided graph is constructed to preserve the spectral similarity from the same labels. Finally, the two graphs are integrated to learn a discriminant projection. This work is an extension of our previous conference work [52]. Compared to the spectral-related work, i.e., RLDE, SPLDA learns a superpixel-guided graph with spatial information. SPLDA was proposed on the basis of the LPNPE method by exploring the spatial relationship in the adaptive superpixels, rather than the fixed-size window in LPNPE. Furthermore, SSDR is an extension of SPLDA that integrates the superpixel-guided and label-guided graphs to utilize the spatial-spectral information. The main contributions of the proposed algorithms are as follows:

(1) A superpixel-guided graph is learned to preserve the spatial similarity. Compared to a fixed-size window based graph, the superpixel-guided graph is shape-adaptive and more effective at exploiting the spatial context.
(2) The superpixel-guided graph and the label-guided graph are integrated to simultaneously explore the spatial and spectral similarity, to learn a discriminant projection.
(3) The experimental results obtained with two HSI data sets confirm the advantage of the proposed methods for small sample size classification.

The rest of this paper is organized as follows. In Section 2, we briefly review the theories of SLIC superpixel segmentation, graph embedding (GE), LDA, and LDE. Detailed descriptions of the proposed methods are provided in Section 3, and Section 4 provides the experimental results and discussions. Finally, our conclusions are drawn in Section 5.

2. Related works

2.1. Simple linear iterative clustering (SLIC) superpixel segmentation

Pixels within a local domain are likely to share similar spectral curves, which means that local domains have the property of homogeneity. In order to exploit the spatial neighborhood structure, superpixel segmentation is introduced to generate homogeneous regions. Compared to a fixed-size local region, such as a square window, superpixel segmentation can segment images adaptively, according to the spatial features. Moreover, the spectral and spatial prior information can be taken into account through the spectral constraint of the superpixel regions.

SLIC, as proposed by Achanta et al. [51], is one of the most widely used segmentation methods, due to its simplicity and effectiveness. SLIC adopts a K-means clustering approach to generate superpixels. To start SLIC, in detail, the color space of the image should first be transformed into the CIELAB color space. The k initial cluster centers are then distributed on a regular grid spaced S pixels apart, where the grid interval is $\sqrt{N/K}$, N is the total number of pixels of the image, and K is the number of superpixels set in advance. The centers are moved to the lowest gradient position

in a 3×3 neighborhood, in order to avoid centering a superpixel on an edge or a noisy pixel.

To speed up the algorithm, SLIC limits the search region to $2S \times 2S$, instead of the whole image, which can significantly reduce the number of distance calculations, resulting in a speed advantage compared to conventional K-means clustering.

The distance measure D between cluster center C_k and pixel i is formulated as follows:

$$\begin{aligned} d_{lab} &= \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \\ d_{xy} &= \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \\ D &= \sqrt{(d_{lab})^2 + \left(\frac{d_{xy}}{S}\right)^2 m^2}, \end{aligned} \tag{1}$$

where d_{lab} is the color distance and d_{xy} is the position distance. m balances the relative importance between the color similarity and spatial proximity. As m gets bigger, the resulting superpixels become more compact and have a lower area to perimeter ratio. As m gets smaller, the resulting superpixels adhere more tightly to image boundaries, but have a less regular size and shape [51].

According to D, each pixel can be associated with the nearest cluster center. Therefore, SLIC updates the new cluster center location and then computes the residual error, until convergence.

2.2. Local discriminant embedding (LDE)

The traditional DR methods can be unified into a common graph embedding framework for DR [53], which provides an undirected weighted graph to describe the geometric properties of the data set. We begin by defining the notation and the problem statement. Vectors are written in lowercase and matrices are shown in capitals. For instance, x stands for a vector and X represents a matrix. Specifically, for data $X = \{x_1, x_2, \ldots, x_m\} \in R^{n \times m}$, where $x_i \in R^n$ denotes the ith vertex with n features, we let W be the weight matrix, with W_{ij} standing for the distance between vertices x_i and x_j. The key point of graph embedding is to find a projection which can preserve the similarity between vertices, measured by the weight matrix graph W. We assume that the projection matrix $P \in R^{n \times d} (d \ll n)$ is expected to transform the data X into the low-dimensional space by a linear function, i.e.,$Y = P^T X$. $Y = \{y_1, y_2, \ldots, y_m\}$ is the low-dimensional embedding of all the data. The objective function of the graph embedding is then defined as:

$$P^* = \arg\min_{P^T X X^T P = I} \sum_{i,j} \|P^T X_i - P^T X_j\|^2 W_{ij}, \tag{2}$$

where the constraint $P^T X X^T P = I$ removes an arbitrary scaling factor in the embedding. According to Yan et al. [53], (2) can be derived as follows:

$$\begin{aligned} P^* &= \arg\min_{P^T X X^T P = I} tr(P^T X L X^T P) \\ &= \arg\min \frac{tr(P^T X L X^T P)}{tr(P^T X X^T P)}, \end{aligned} \tag{3}$$

where $L = D - W$ is the Laplacian matrix [54], and D is a diagonal matrix with the sums of the ith row of W_{ij} by $D_{ii} = \sum_j W_{ij}$. The optimal projection $P^* = [p_1, p_2, \ldots . p_d]$ that minimizes the objective function is given by the minimum eigenvalue:

$$X L X^T p_d = \lambda_d X X^T p_d. \tag{4}$$

LDE is conducted under the graph embedding framework, and it tries to learn a label-guided graph. In effect, LDE seeks to dissociate the submanifold of each class from one another [26]. The procedure of LDE can be characterized by the following three steps.

Firstly, the label-guided neighborhood graph construction. Two graphs are constructed, i.e., the within-class graph G_w and the between-class graph G_b. To construct G_w, for each pair of points x_i and x_j, an edge is added between them if they are from the same class and x_j is the k-nearest neighbor of x_i. To construct G_w, an edge is added if x_i and x_j are from different classes and x_j is the k-nearest neighbor of x_i.

Secondly, the affinity weight matrix computation. To compute the affinity matrix W_w of G_w, each element $w_{w,\,ij}$ of W_w is the weight of the edge between x_i and x_j, and is defined as:

$$w_{w,ij} = exp[-\|x_i - x_j\|^2 / t]. \tag{5}$$

Similarly, the affinity weight W_b of G_b is computed in the same way.

Finally, the projection vector optimization. The objective function of LDE is given as follows:

$$\arg\min \frac{tr\{P^T X L_w X^T P\}}{tr\{P^T X L_b X^T P\}}, \tag{6}$$

where $L_b = D_b - W_b$ is the between-class graph Laplacian matrices, $L_w = D_w - W_w$ is the within-class graph Laplacian matrices, and $D_b(D_w)$ is a diagonal matrix whose entries are column sums of $W_b(W_w)$.

2.3. Linear discriminant analysis (LDA)

LDA [30] constructs a within-class scatter matrix and a between-class scatter matrix. The within-class scatter matrix S_w characterizes the intraclass compactness, while the between-class scatter matrix S_b describes the interclass separability, which are shown as follows:

$$S_w = \sum_{k=1}^{c} \left(\sum_{i=1}^{n_k} (x_i^{(k)} - \mu^{(k)})(x_i^{(k)} - \mu^{(k)})^T \right) \tag{7}$$

$$S_b = \sum_{k=1}^{c} n_k (\mu^{(k)} - \mu)(\mu^{(k)} - \mu)^T, \tag{8}$$

where $x_i^{(k)}$ is the ith sample in the kth class, $\mu^{(k)}$ is the mean of the kth class, n_k is the number of samples in the kth class, and μ is the mean vector of the entire training set.

To enhance intraclass compactness and interclass separability, LDA seeks the projection direction in which the ratio of the between-class covariance to within-class covariance is maximized. The optimal projection matrix P can be obtained with the following optimization problem:

$$\begin{aligned} P &= \arg\min_p \frac{tr(P^T S_w P)}{tr(P^T S_b P)} \\ &= arg\min_p \frac{tr(P^T S_w P)}{tr(P^T (S_b + S_w) P)}, \end{aligned} \tag{9}$$

where $S_b + S_w = X X^T$.

3. Spatial-spectral dimension reduction

In this section, we first present the SPLDA method, which builds a superpixel-guided graph to exploit the spatial information. SPLDA utilizes the prior assumption that pixels from the same superpixel have strong correlations, and tries to find an optimal projection which can preserve the local homogeneity. Subsequently, we build a label-guided graph to learn the spectral similarity. The proposed SSDR integrates the superpixel-guided graph and label-guided graph to simultaneously capture the structural similarity underlying the spectral and spatial dimensions.

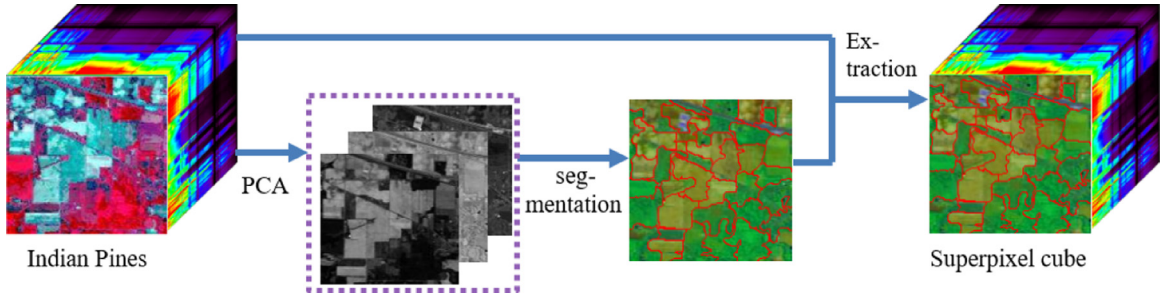

Fig. 1. The generation process for superpixels in an HSI.

3.1. The generation process for superpixels in the HSI

Differing from single-band gray or three-band color images, HSIs usually have tens or hundreds of spectral bands. On this basis, applying SLIC to an HSI directly will result in low computational efficiency. Hence, PCA is first utilized to extract the first three principal components of the HSI as the input data. The input image is then segmented into K homogeneous superpixels, with K set in advance. Finally, the location indices of each superpixel boundary are mapped to the original HSI space to generate the non-overlapping 3-D superpixels [38]. The flowchart of superpixel generation is illustrated in Fig. 1.

3.2. Superpixel-based linear discriminant analysis (SPLDA)

The HSI image can be denoted as $X = [x_1, \ldots, x_M]$, with $X_i \in R^N$, and M is the number of pixels. $Y = [y_1, y_2, \ldots, y_M]$ stands for the dimension-reduced points, with $Y = P^T X$. The HSI is generated into K superpixels with different shapes and sizes via SLIC. Considering one sample x_i, the superpixel neighboring pixels corresponding to x_i are represented as $x_i^{(j)} (j = 1, 2, \ldots, k_i)$, where k_i stands for the pixel number in one superpixel. For each pixel x_i, we can obtain the superpixel-guided graph W_{ii_j} to capture the similarity between x_i and $x_i^{(j)}$ in the same superpixel:

$$W_{ii_j} = \frac{exp\{-\gamma_0 \|x_i - x_i^{(j)}\|^2\}}{\sum_{j=1}^{k_i} exp\{-\gamma_0 \|x_i - x_i^{(j)}\|^2\}}. \tag{10}$$

The pixels from the same superpixel will have close similarity [55], and this relationship should be retained after the DR. In other words, the close relationship of the pixels from the same superpixel should remain after DR. To preserve this intrinsic graph, the optimization model is formulated as:

$$P_{opt} = \arg\min_{P^T X X^T P = I} \sum_{i=1}^{M} \sum_{j=1}^{k_i} \|y_i - y_i^{(j)}\|^2 W_{ii_j}, \tag{11}$$

where the constraint $P^T X X^T P = I$ is adopted to scale the embedding. The objective function (11) can be easily converted to:

$$\begin{aligned} P_{opt} &= \arg\min_{P^T X X^T P = I} tr(P^T S_w^{SPLDA} P) \\ &= \arg\min \frac{tr(P^T S_w^{SPLDA} P)}{tr(P^T S_b^{SPLDA} P)}, \end{aligned} \tag{12}$$

where $S_w^{SPLDA} = \sum_{i=1}^{M} (\sum_{j=1}^{k_i} W_{ii_j} (x_i - x_i^{(j)})(x_i - x_i^{(j)})^T)$ is the superpixel-guided scatter matrix, and $S_b^{SPLDA} = XX^T$ is the dissimilarity scatter matrix. Similar to LDA, the objective model (12) tries to find a discriminant low-dimensional space by minimizing the superpixel neighborhood interclass preserving scatter while, at the same time, maximizing the distance between the means of the samples. After being transformed into a projection space, data points in the same superpixel maintain their intrinsic neighborhood relations, whereas neighboring superpixels no longer stick to one another. A schematic diagram of SPLDA is provided in Fig. 2.

Eq. (12) can be easily solved by the generalized eigenvalue decomposition approach:

$$S_w^{SPLDA} p_d = \lambda_d S_b^{SPLDA} p_d, \tag{13}$$

where λ_d is the dth eigenvalue, and the projection matrix is $P = [p_1, p_2, \ldots, p_d]$, where p_d is the eigenvector corresponding to the dth smallest nonzero eigenvalue.

3.3. Spatial-spectral dimension reduction (SSDR)

As mentioned in Section 3.2, SPLDA can fully use the spatial information in the HSI and find a discriminant projection. However, SPLDA fails to take the label information of the samples into account. Inspiringly, as introduced in Section 2.2, the label-guided graphs, i.e., G_w and G_b, are constructed, and the related affinity weights are W_w and W_b, respectively. The proposed SSDR tries to integrate the superpixel-guided graph and label-guided graph to utilize the spatial and spectral similarity concurrently.

Firstly, on the basis of the label-guided weight matrices W_w and W_b, we extend LDE to regularized LDE (RLDE) and reformulate (6) as the LDA version:

$$\arg\min \frac{tr(P^T S_w^{RLDE} P)}{tr(P^T S_b^{RLDE} P)}, \tag{14}$$

where $S_w^{RLDE} = [(1-\alpha) S_w + \alpha J_{reg}]$ with $S_w = X L_w X^T, J_{reg} = \sum_j S_{w,ij}$, and $S_b^{RLDE} = X L_b X^T$. The regularization term $J_{reg} = \sum_j S_{w,ij}$ is a diagonal matrix, which can overcome the singularity when training samples are limited [27].

Secondly, by setting $S_w^{SSDR} = [(1-\beta) S_w^{SPLDA} + \beta S_w^{RLDE}]$ and $S_b^{SSDR} = [(1-\beta) S_b^{SPLDA} + \beta S_b^{RLDE}]$, we fuse SPLDA (12) and RLDE (14) into one objective function:

$$\arg\min \frac{tr(P^T S_w^{SSDR} P)}{tr(P^T S_b^{SSDR} P)}, \tag{15}$$

in which α and β are two tradeoff parameters that control the relative importance of the spatial information, spectral information, and regularization terms with $0 \le \alpha, \beta \le 1$. In this way, the superpixel spatial structure and the labeled spectral information are integrated to learn a discriminative low-dimensional subspace. The optimization (15) can be efficiently solved by the minimum eigenvalue solution to the generalized eigenvalue problem (4).

4. Experimental results and discussions

In order to demonstrate the effectiveness of the proposed algorithms, we conducted experiments on two widely used HSI data sets. This section includes three parts: (1) an introduction to the

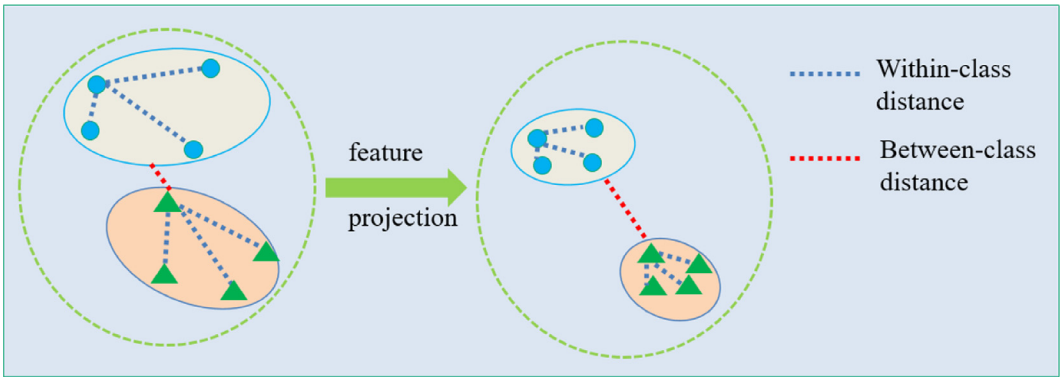

Fig. 2. Schematic diagram of SPLDA.

experimental data sets and the experimental setup; (2) a parameter analysis for the proposed methods; and (3) a quantitative assessment of the proposed algorithms for HSI classification by a comparison with other existing state-of-the-art methods: PCA [31], local Fisher discriminant analysis (LFDA) [56], nonparametric weighted feature extraction (NWFE) [57], LDE [26], RLDE [27], and LPNPE [27].

4.1. Experimental data sets and setup

(1) The Indian Pines data set: This data set was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in 1992 over the Indian Pines test site in North-western Indiana. The image contains 145×145 pixels at a spatial resolution of 20 m. The AVIRIS sensor generates 220 bands across the 0.4 – 2.5 μm spectral range. In total, 200 bands of the image were retained for the experiments after removing the 20 water absorption bands. The total number of labeled samples was 10249 from 16 classes. The false-color image (R:50, G:27, B:17) is shown in Fig. 7(a).

(2) The Pavia University data set: This data set was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor. It is of a spatial size of 610×340 pixels and contains 103 spectral bands, after removing the noisy and water absorption bands. The available ground truth contains nine classes. The false-color image (R:102, G:56, B:31) is shown in Fig. 10(a).

This paper utilizes the overall accuracy (OA), average accuracy (AA), and kappa coefficient to evaluate the effectiveness of the proposed SPLDA and SSDR algorithms. Two simple classifiers were adopted to evaluate the DR performance: *k*-nearest neighbor (KNN) and SVM. The parameters in SVM, i.e., the radial basis function (RBF) kernel parameter and the penalty factor, were selected by 10-fold cross validation. We compared the proposed SPLDA and SSDR algorithms with six representative DR methods. The comparison algorithms were as follows, among which the first five methods are spectral feature based DR methods, and the last method takes the spatial structure information into consideration.

(1) Principal component analysis (PCA): a classic method of global DR [31], which aims to find a subspace of orthogonal projections in accordance with maximizing the variance of the input feature matrix.
(2) Local fisher discriminant analysis (LFDA) [56]: uses local neighborhood information to construct the weighted between-class and within-class scatter matrices, and then finds a more discriminative subspace.
(3) Nonparametric weighted feature extraction (NWFE) [57]: is based on a nonparametric extension of the scatter matrices. The main idea is putting different weights on each sample to compute the "weighted means", and defining new nonparametric between-class and within-class scatter matrices to obtain more features.
(4) Local discriminant embedding (LDE): a manifold learning method [26], which attempts to maintain the original neighborhood relations for the neighboring data points of the same class and exclude neighboring points of different classes after the embedding.
(5) Regularized LDE (RLDE) [27]: extends LDE by adding constraints that preserve the data diversity, and overcomes the singularity in the case of limited training samples.
(6) Local pixel neighborhood preserving embedding (LPNPE) [27]: the method takes the spatial information that minimizes the scatter matrix in a fixed neighborhood window into consideration, and maximizes the total scatter matrix.

The parameters of these methods were set as suggested in the original papers. We randomly selected a certain number of samples for each class as training samples, with the rest used for the testing. In order to reduce the possible bias caused by the random sampling, 10 sets of independent repeated experiments were conducted.

4.2. Parameter analysis

4.2.1. Effect of the number of superpixels

First of all, the effect of the superpixel number on the proposed SPLDA method was analyzed. The number of training samples was randomly selected as 20 and 30 in the Indian Pines data set and Pavia University data set, respectively. The accuracy values were averaged over 10 runs to reduce the possible bias induced by the random sampling. Considering that each superpixel usually contains 10 to 30 pixels, the number of superpixels was selected from 600 to 2000 for the Indian Pines data set, and from 6000 to 20000 for the Pavia University data set.

Fig. 3(a)–(d) report the OA values of the proposed SPLDA method under different numbers of superpixels with the two classifiers (i.e., KNN and SVM) on the two test images. It can be observed that, as the superpixel number varies within a certain range, the OA values of the SPLDA method are stable. It can thus be observed that the proposed SPLDA method is robust to the number of superpixels. Therefore, we set the number of superpixels to 800 in the Indian Pines experiments, and to 18000 in the Pavia University experiments.

4.2.2. Effect of the regularization parameters α and β

The regularization parameters α and β are used to balance the spectral, spatial, and regularized contributions, as introduced in Section 3.3. We fixed the other parameters and focused on

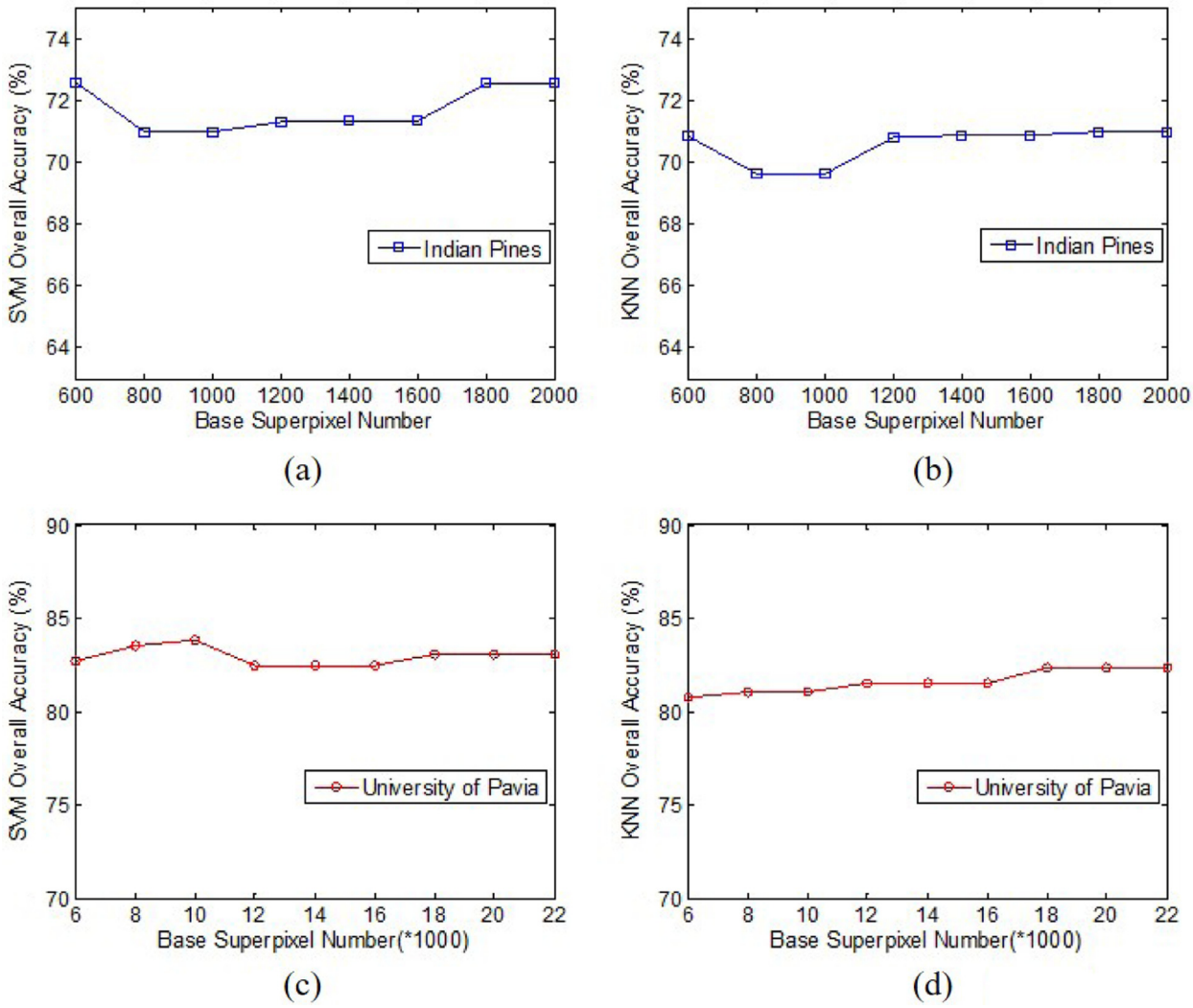

Fig. 3. Classification accuracy with respect to the number of superpixels on the two test data sets: (a) Indian Pines image with SVM classifier; (b) Indian Pines image with KNN classifier; (c) Pavia University image with SVM classifier; (d) Pavia University image with KNN classifier.

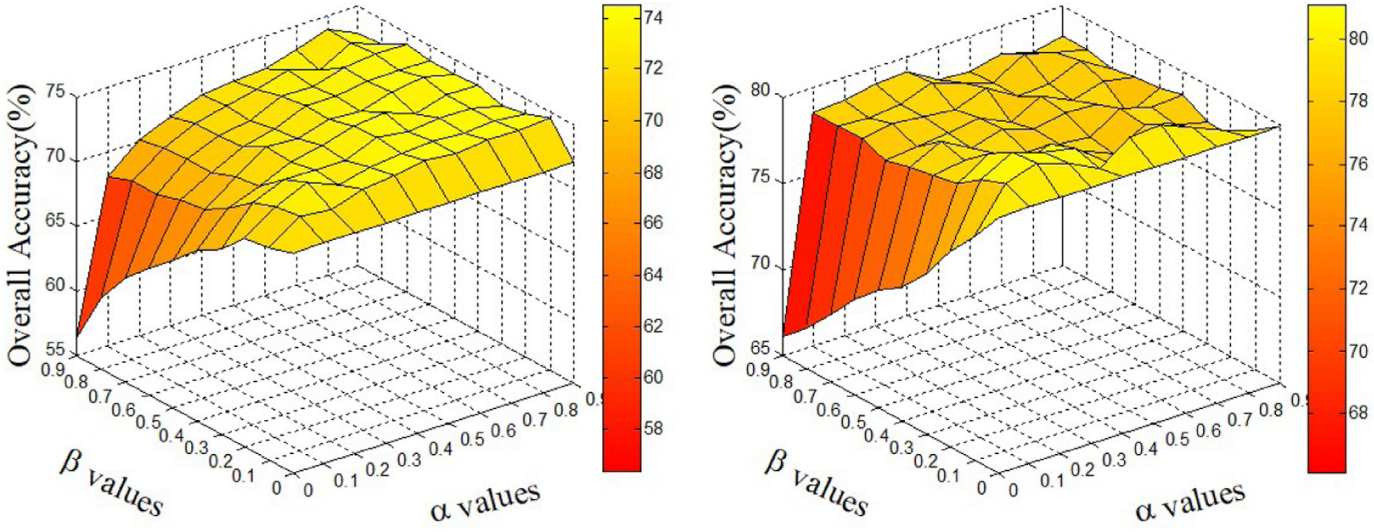

Fig. 4. The changes of the evaluation values with changes of parameters α and β on the two HSI data sets: (a) Indian Pines; (b) Pavia University.

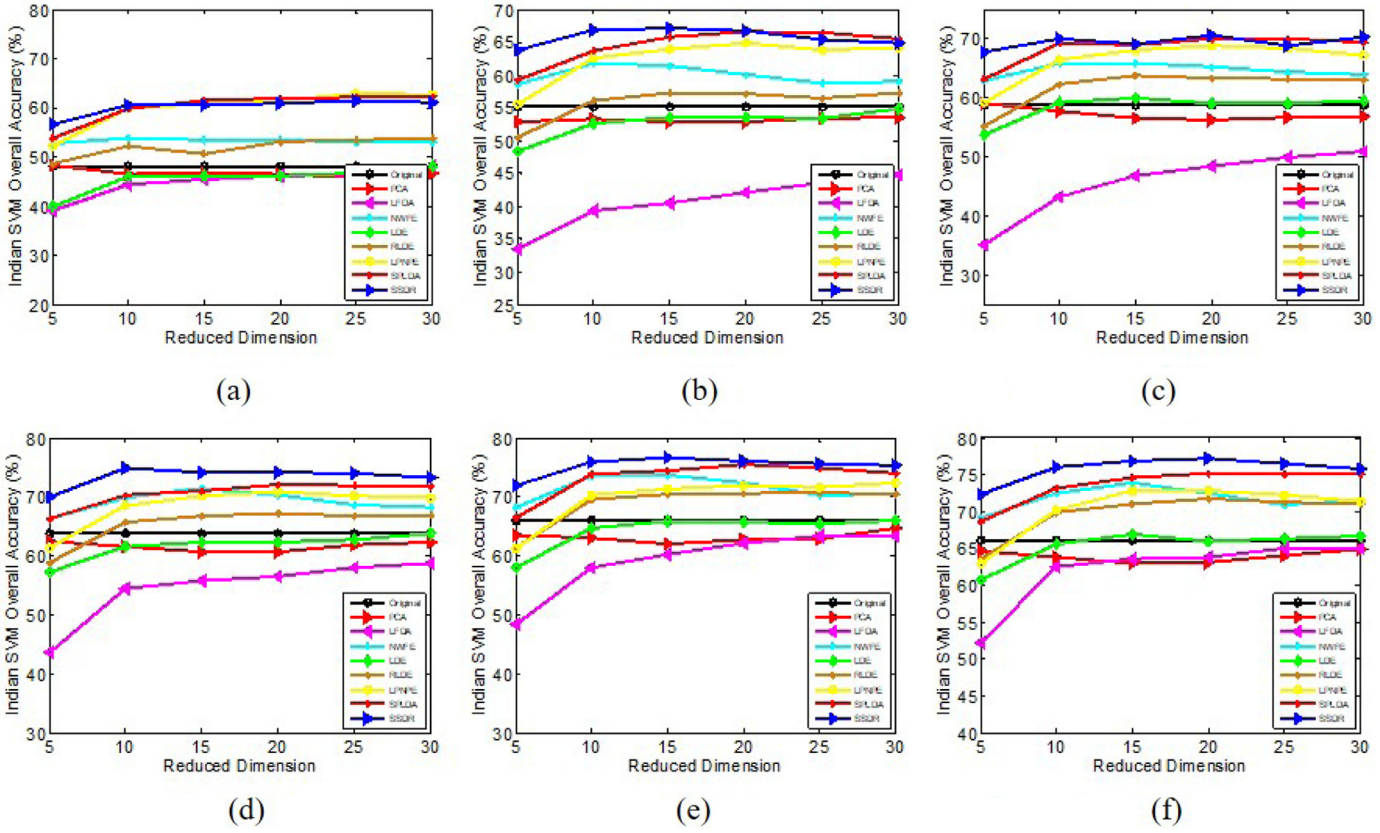

Fig. 5. OAs with respect to: (a)–(f) different sizes of training set (5, 10, 15, 20, 25, 30) and different reduced dimensions (from 5 to 30) for the Indian Pines data set, combined with the SVM classifier.

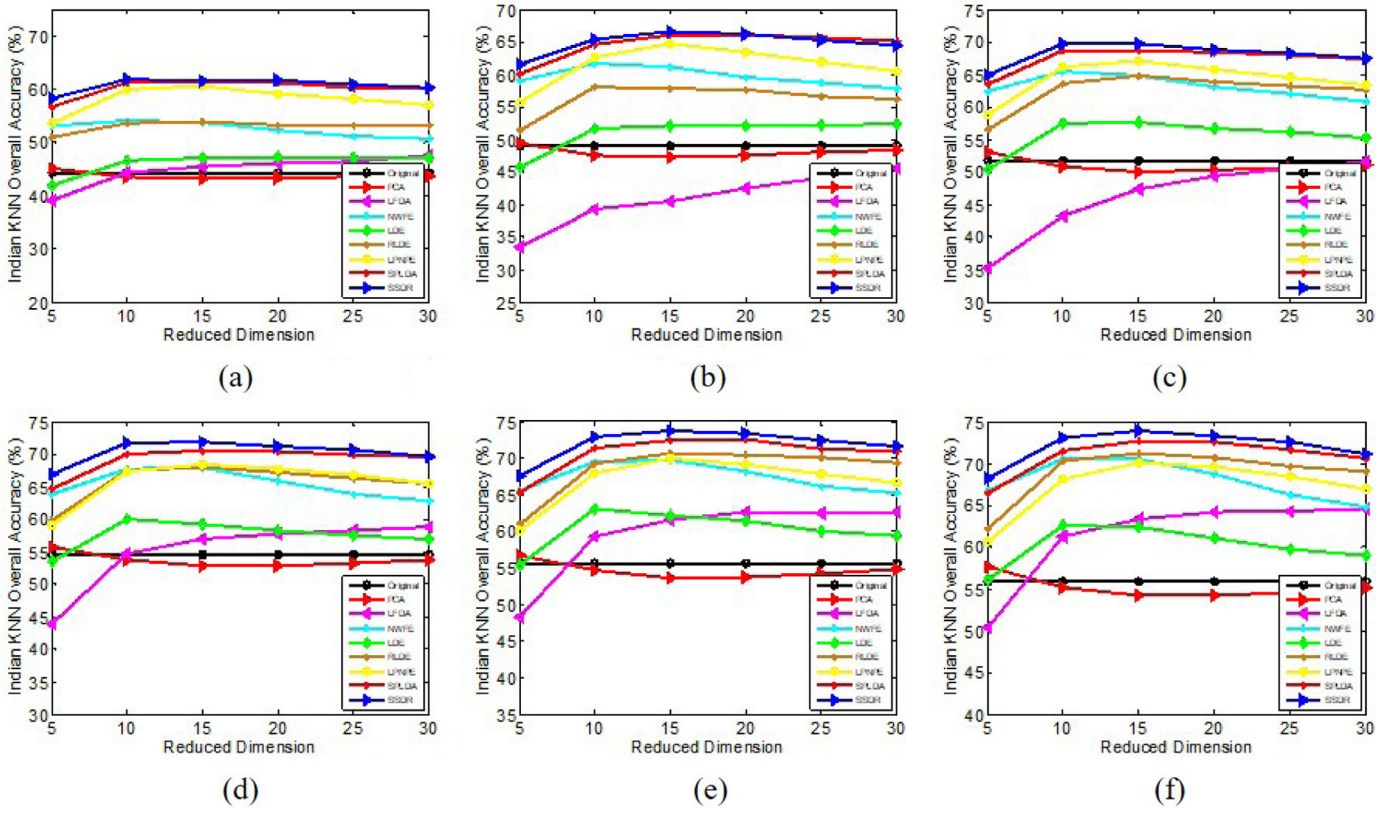

Fig. 6. OAs with respect to: (a)–(f) different sizes of training set (5, 10, 15, 20, 25, 30) and different reduced dimensions (from 5 to 30) for the Indian Pines data set, combined with the KNN classifier.

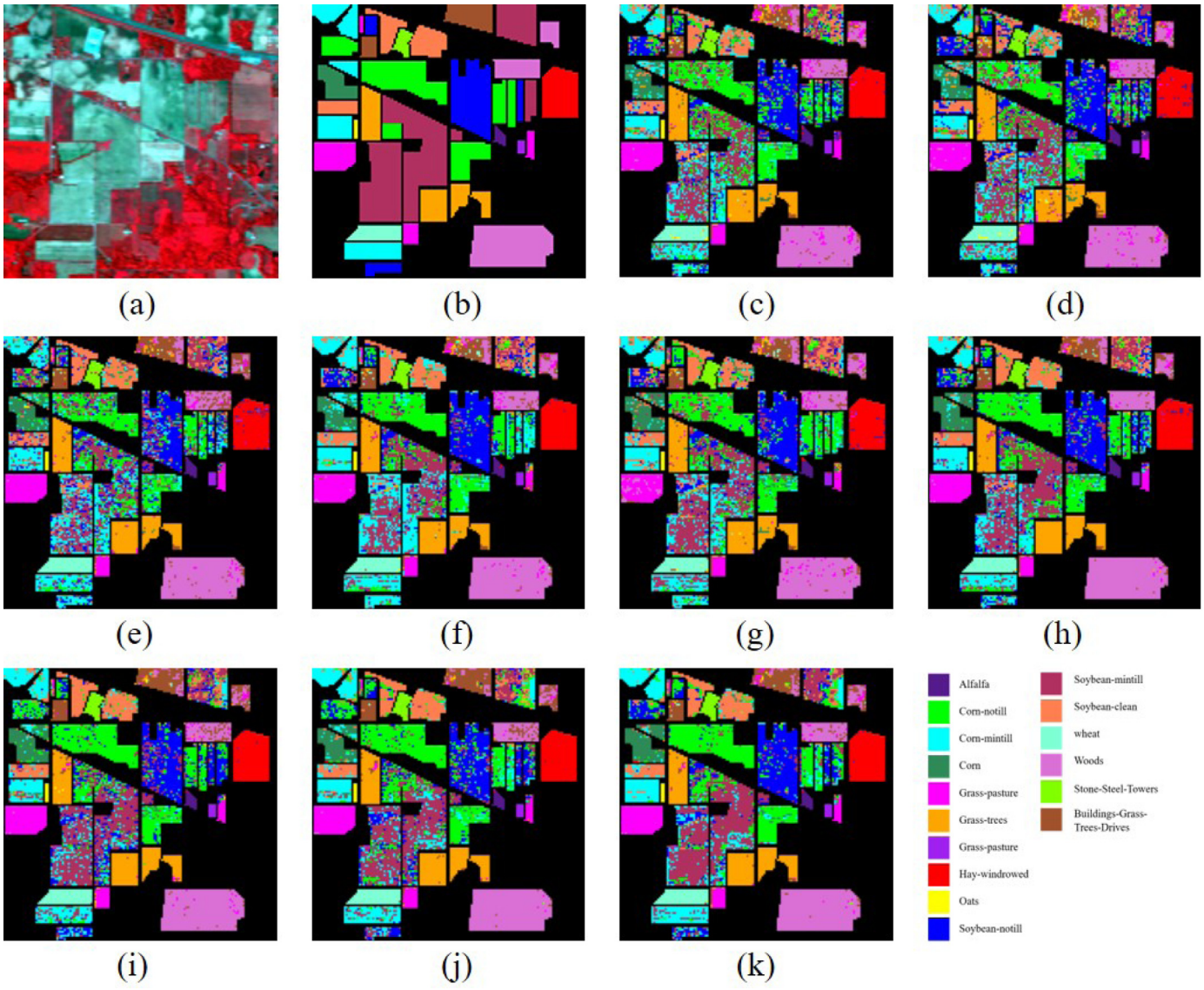

Fig. 7. SVM classification maps of the different methods with the Indian Pines data in dim=15. (a) False-color image of the Indian Pines data set (R: 57, G: 27, B: 17). (b) Ground-truth map. (c) Original (SVM). (d) PCA. (e) LFDA. (f) NWFE. (g) LDE. (h) RLDE. (i) LPNPE. (j) SPLDA. (k) SSDR.

the effect of parameters α and β in the two HSI experimental data sets. The parameters α and β were chosen from the range $[0, 0.1, \ldots, 0.9]$, respectively. The number of training samples per class was randomly chosen as 20 for both data sets.

Fig. 4 shows the OA values of the SSDR method under different values of parameters α and β. We use the SVM classification results of the transformed image by SSDR to validate the performance of the different parameters. It can be observed that, for these two experimental data sets, the results remain relatively stable when parameters α and β are varied. SSDR achieves the highest OA in the Indian Pines data set when the parameters are set as $(\alpha, \beta) = (0.5, 0.2)$. Similarly, the optimal parameters for the Pavia University data set are $\alpha = 0.2$ and $\beta = 0.1$. The proposed SSDR can achieve the best evaluation results when the contributions of the superpixel-guided graph and label-guided graph are at a certain proportion. Typically, the weight of the superpixel-guided graph is larger. This is mainly due to the fact that SPLDA takes the spatial information of the surrounding pixels into account, and can characterize data features more effectively than the traditional spectral-based DR methods.

4.3. Experimental results

4.3.1. Indian pines image

In order to fully testify the performance of the proposed methods, experiments with the different methods under different sample numbers and different dimensions of projected space were conducted. We randomly selected n (n = 5, 10, 15, 20, 25, 30) samples from each class as the training set, and the rest were used as the test set. Considering that the ground-truth sample number of two categories (grass-pasture-mowed and oats) in the Indian Pines data is less than 30, half of the total quantity was chosen for the training. The reduced dimensionality varied from 5 to 30.

Figs. 5 and 6 show the OAs of the KNN and SVM classifiers on the dimension-reduced images obtained by the different DR methods. The OAs of the KNN and SVM classifiers on the original Indian Pines data set with different sample numbers are used as the baselines. Compared to the other methods, SPLDA and SSDR achieve the best and the second-best classification results in almost all cases with different dimensions under different numbers of samples. The advantage is more obvious when the sample num-

Table 1
Classification accuracies of the different DR methods for the Indian Pines data set.

Samples	Classifiers	Index	Original	PCA	LFDA	NWFE	LDE	RLDE	LPNPE	SPLDA	SSDR
5	KNN	OA	44.34 ± 2.61	43.16 ± 2.98	45.44 ± 3.55	53.83 ± 4.24	47.13 ± 3.58	53.91 ± 4.30	60.67 ± 3.15	61.33 ± 2.16	**61.60 ± 2.10**
		AA	56.87 ± 2.03	55.73 ± 2.04	57.24 ± 2.58	65.43 ± 2.47	59.47 ± 3.17	65.90 ± 3.23	74.71 ± 1.88	75.19 ± 0.91	**75.41 ± 1.08**
		Kappa	37.93 ± 2.56	36.66 ± 2.96	39.07 ± 3.63	48.34 ± 4.51	40.92 ± 3.80	48.44 ± 4.65	55.91 ± 3.48	56.57 ± 2.23	**56.86 ± 2.20**
	SVM	OA	48.07 ± 3.58	46.79 ± 3.27	45.44 ± 3.55	53.50 ± 4.75	46.22 ± 2.99	50.58 ± 8.11	60.89 ± 4.61	**61.37 ± 3.95**	60.60 ± 4.37
		AA	60.76 ± 2.69	58.91 ± 1.53	57.24 ± 2.59	65.69 ± 3.21	58.27 ± 2.79	65.65 ± 3.12	74.59 ± 2.04	**75.36 ± 1.98**	74.26 ± 1.54
		Kappa	37.93 ± 2.56	36.66 ± 2.96	39.07 ± 3.63	48.34 ± 4.51	40.92 ± 3.80	48.44 ± 4.65	55.91 ± 3.48	**56.57 ± 2.23**	56.86 ± 2.20
10	KNN	OA	49.07 ± 1.71	47.35 ± 1.95	40.55 ± 2.85	61.14 ± 1.39	52.09 ± 2.71	57.87 ± 2.49	64.77 ± 2.30	66.05 ± 2.09	**66.56 ± 1.98**
		AA	61.62 ± 1.69	59.61 ± 1.33	54.77 ± 1.74	72.46 ± 0.98	65.42 ± 1.26	71.38 ± 0.98	78.99 ± 1.25	79.82 ± 0.98	**79.94 ± 1.09**
		Kappa	43.20 ± 1.74	41.34 ± 1.94	34.36 ± 2.78	56.49 ± 1.49	46.60 ± 2.82	52.97 ± 2.62	60.58 ± 2.39	61.96 ± 2.17	**62.53 ± 2.17**
	SVM	OA	55.11 ± 2.70	52.93 ± 2.05	40.54 ± 2.83	61.39 ± 2.82	53.60 ± 3.64	57.34 ± 2.94	63.99 ± 4.12	65.84 ± 3.04	**67.13 ± 3.99**
		AA	67.71 ± 1.59	64.73 ± 1.86	54.77 ± 1.73	73.69 ± 1.49	66.19 ± 1.73	71.12 ± 1.41	78.24 ± 1.26	**79.49 ± 1.31**	79.25 ± 1.80
		Kappa	49.68 ± 2.86	47.25 ± 2.21	34.34 ± 2.76	56.69 ± 3.12	48.24 ± 3.87	52.44 ± 3.13	60.74 ± 4.19	62.53 ± 3.31	**62.71 ± 4.30**
15	KNN	OA	51.83 ± 1.76	50.01 ± 1.81	47.41 ± 1.88	64.83 ± 1.71	57.64 ± 2.08	64.84 ± 1.69	67.13 ± 1.85	68.66 ± 1.82	**69.73 ± 2.11**
		AA	65.08 ± 1.63	63.46 ± 1.73	61.53 ± 2.22	76.12 ± 0.83	69.88 ± 1.24	76.84 ± 0.67	80.46 ± 1.07	81.41 ± 0.89	**81.64 ± 0.93**
		Kappa	46.21 ± 1.90	44.22 ± 1.92	41.44 ± 1.93	60.53 ± 1.88	52.56 ± 2.19	60.52 ± 1.79	80.47 ± 1.93	64.83 ± 1.95	**65.99 ± 2.27**
	SVM	OA	58.80 ± 2.93	56.54 ± 2.99	46.88 ± 2.53	65.85 ± 3.89	60.05 ± 3.24	63.85 ± 2.09	68.07 ± 3.18	69.01 ± 2.99	**69.10 ± 2.69**
		AA	71.60 ± 1.48	68.32 ± 2.64	61.26 ± 2.24	77.84 ± 1.84	71.81 ± 1.92	76.83 ± 1.04	80.31 ± 1.36	**81.52 ± 1.19**	81.11 ± 1.29
		Kappa	53.79 ± 3.14	51.32 ± 3.31	40.98 ± 2.51	61.76 ± 3.72	55.23 ± 3.42	59.42 ± 2.21	64.22 ± 3.39	65.19 ± 3.18	**65.32 ± 2.89**
20	KNN	OA	54.62 ± 1.83	52.87 ± 1.34	56.92 ± 1.49	67.89 ± 1.91	59.19 ± 1.74	68.04 ± 2.30	68.43 ± 1.44	70.58 ± 1.28	**71.84 ± 1.46**
		AA	61.15 ± 1.04	59.36 ± 1.35	65.96 ± 1.22	72.32 ± 0.80	64.94 ± 0.79	73.70 ± 1.62	75.72 ± 0.75	76.98 ± 0.91	**77.43 ± 0.94**
		Kappa	49.18 ± 1.91	47.28 ± 1.44	51.82 ± 1.65	63.92 ± 2.04	54.27 ± 1.81	64.04 ± 2.49	64.56 ± 1.55	66.96 ± 1.38	**68.34 ± 1.57**
	SVM	OA	63.79 ± 1.84	60.68 ± 1.46	55.88 ± 1.66	71.45 ± 1.55	62.39 ± 1.34	66.78 ± 1.66	70.10 ± 1.80	71.06 ± 2.78	**74.22 ± 2.41**
		AA	69.39 ± 0.90	66.00 ± 0.89	65.18 ± 1.27	75.17 ± 1.43	67.79 ± 0.63	73.70 ± 1.22	76.69 ± 1.17	77.43 ± 1.73	**78.22 ± 1.47**
		Kappa	59.31 ± 1.96	55.81 ± 1.59	50.69 ± 1.86	67.82 ± 1.71	57.74 ± 1.45	62.73 ± 1.78	66.38 ± 2.00	68.42 ± 2.34	**70.96 ± 2.03**
25	KNN	OA	55.64 ± 1.26	53.68 ± 1.20	61.59 ± 0.72	69.71 ± 1.48	62.21 ± 1.11	70.69 ± 1.15	69.95 ± 1.61	72.42 ± 1.74	**73.81 ± 1.42**
		AA	61.96 ± 1.85	60.31 ± 1.65	69.65 ± 1.02	73.40 ± 1.39	67.64 ± 1.41	75.15 ± 1.18	76.49 ± 0.67	78.04 ± 0.58	**78.55 ± 0.70**
		Kappa	50.22 ± 1.39	48.11 ± 1.31	56.80 ± 0.87	65.84 ± 1.61	57.46 ± 1.29	66.93 ± 1.26	66.19 ± 1.70	68.95 ± 1.86	**70.45 ± 1.56**
	SVM	OA	66.03 ± 1.80	62.05 ± 1.70	60.19 ± 1.89	73.67 ± 1.97	65.89 ± 1.13	70.42 ± 0.71	71.32 ± 2.97	74.40 ± 1.83	**76.72 ± 2.00**
		AA	71.13 ± 0.77	67.39 ± 1.58	68.65 ± 1.13	76.85 ± 1.38	70.68 ± 1.49	75.54 ± 1.13	77.44 ± 1.39	79.60 ± 0.77	**79.61 ± 1.23**
		Kappa	61.75 ± 1.93	57.37 ± 1.93	55.32 ± 2.09	70.28 ± 2.13	61.61 ± 1.20	66.68 ± 0.77	67.69 ± 1.42	71.15 ± 2.00	**73.69 ± 2.15**
30	KNN	OA	55.90 ± 2.07	54.21 ± 2.10	63.33 ± 1.53	70.49 ± 2.21	62.45 ± 2.20	71.19 ± 1.70	70.06 ± 1.68	72.71 ± 1.69	**73.91 ± 1.55**
		AA	57.32 ± 1.14	55.73 ± 1.16	65.79 ± 0.96	68.74 ± 0.77	62.83 ± 1.48	70.22 ± 0.66	70.80 ± 0.67	72.70 ± 0.55	**72.84 ± 0.53**
		Kappa	50.65 ± 2.15	48.76 ± 2.21	58.83 ± 1.67	66.76 ± 2.36	57.86 ± 2.40	67.52 ± 1.84	66.33 ± 1.81	69.27 ± 1.81	**70.59 ± 1.67**
	SVM	OA	66.14 ± 2.55	62.96 ± 2.51	63.62 ± 1.44	73.88 ± 2.24	66.87 ± 1.93	71.01 ± 0.75	72.78 ± 2.78	74.61 ± 1.28	**76.88 ± 1.71**
		AA	66.31 ± 1.18	62.98 ± 1.55	65.90 ± 0.63	71.78 ± 1.06	66.43 ± 1.35	70.66 ± 0.76	72.48 ± 1.28	73.50 ± 0.82	**74.34 ± 0.93**
		Kappa	61.97 ± 2.69	58.36 ± 2.70	59.11 ± 1.53	70.56 ± 2.43	62.73 ± 2.09	67.37 ± 0.86	69.34 ± 2.41	71.31 ± 1.44	**73.92 ± 1.88**

ber is limited, which demonstrates that SPLDA is very effective for the small training sample problem. This is mainly because the proposed SPLDA method can make full use of the spatial consistency information generated by the SLIC segmentation. It can be observed that SPLDA also performs better than the LPNPE algorithm. The main reason for this is that SPLDA adopts a shape-adaptive neighborhood, which can accurately model the local spatial structure and ensure regional homogeneity. The performance of SSDR is better than that of SPLDA, due to the simultaneous preservation of the spatial and spectral similarity during the DR process.

Similarly, Fig. 6 reports the OAs of the different algorithms using the KNN classifier, which indicates that the SPLDA and SSDR algorithms outperform the other methods in each of the dimensions. It should be noted that PCA performs better than LFDA in some cases, due to the fact that LFDA performs poorly under small sample numbers.

The comparison results for the Indian Pines data set obtained with the SVM classifier in dim=15 are visually shown in Fig. 7, including the false-color image (a), the corresponding ground-truth map (b), and the classification images (c)–(k), respectively. It can be observed that the proposed SPLDA method performs better than the other compared spectral-based methods, in most land-over classes, and the SSDR algorithm produces more homogenous areas and smoother classification maps than the other methods, especially in the corn-notill, soybean-mintill, and hay-windrowed classes. To further illustrate the comparison results, the quantitative evaluation results for the OA (%) in dim=15 for all the methods are summarized in Table 1. The results include the average OA, AA, and kappa coefficient, with the associated standard deviation, over 10 runs for each method. The best results for each quality index are labeled in bold, and the second-best results are underlined for the KNN and SVM classifiers, respectively. From Table 1, it can be observed that the proposed SSDR method yields the best OA, kappa, and AA values in most cases, and the values of the SPLDA method are the second-highest among all the methods, which confirms the conclusion that SPLDA and SSDR outperform the other spectral-based DR methods and the spatial-based method which considers the spatial information based on a fixed neighborhood window. The spatial-spectral combined method (SSDR) performs better than the spatial-based method (SPLDA), which indicates that simultaneously combining the spectral information and spatial context provided by training samples is beneficial to DR.

4.3.2. Pavia university data set

This section describes the classification results of the different DR methods on the Pavia University data set. For this data set, we approximately set the number of superpixels to be 18000 and randomly selected n (n= 10, 20, 30, 40, 50) samples from each class as the training set. The remaining samples were used as the test set. The reduced dimensionality was varied from 5 to 30.

Figs. 8 and 9 show the OAs with regard to the different reduced dimensions and different sample sizes for the all methods with the SVM and KNN classifiers, respectively. From Fig. 8, it can be observed that, for the SVM classifier, the proposed SSDR method obtains better classification results than the other methods in almost all cases of different dimensions, and achieves the best classification result around dim 10. SPLDA performs better than the LPNPE method due to the flexibility of the neighborhood preserva-

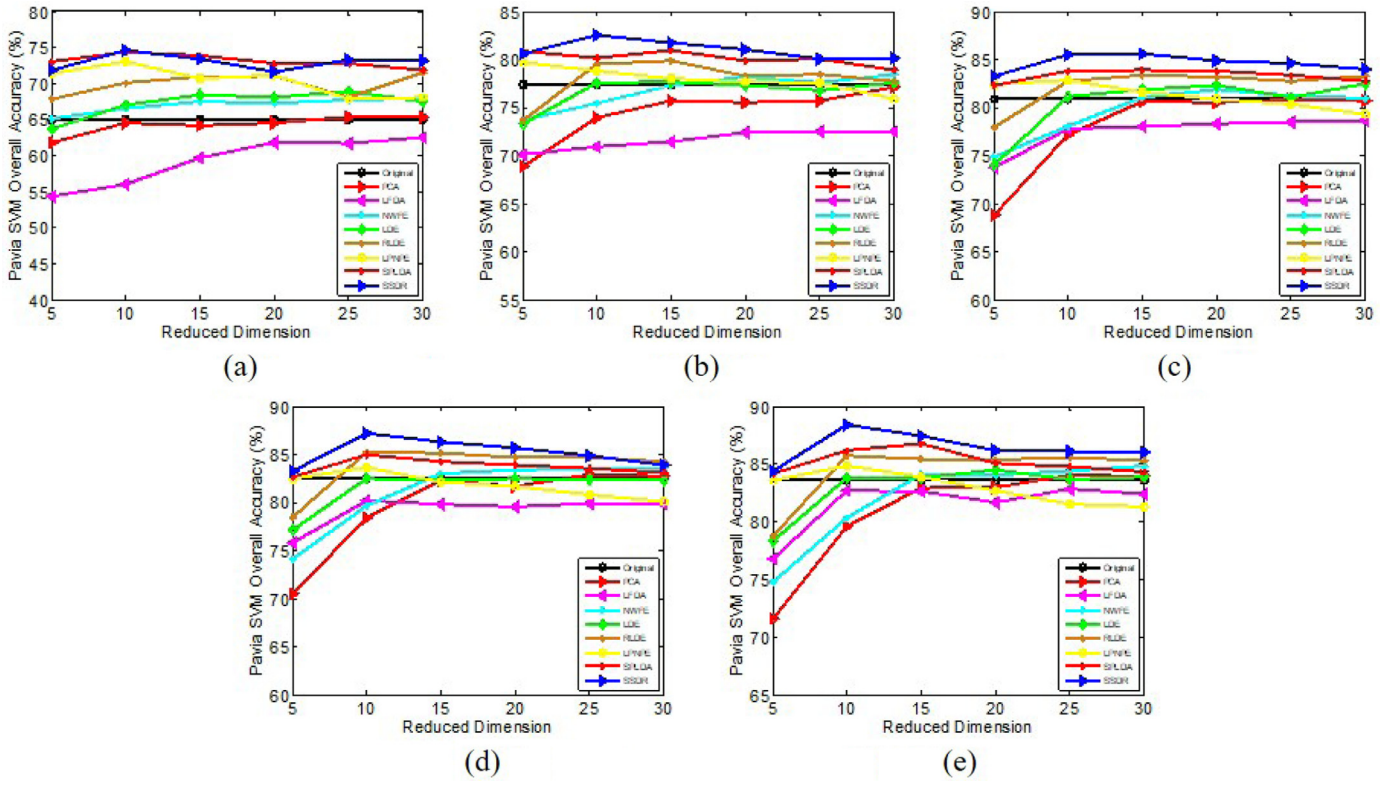

Fig. 8. OAs with respect to: (a)–(e) different sizes of training set (10, 20, 30, 40, 50) and different reduced dimensions (from 5 to 30) for the Pavia University data set, combined with the SVM classifier.

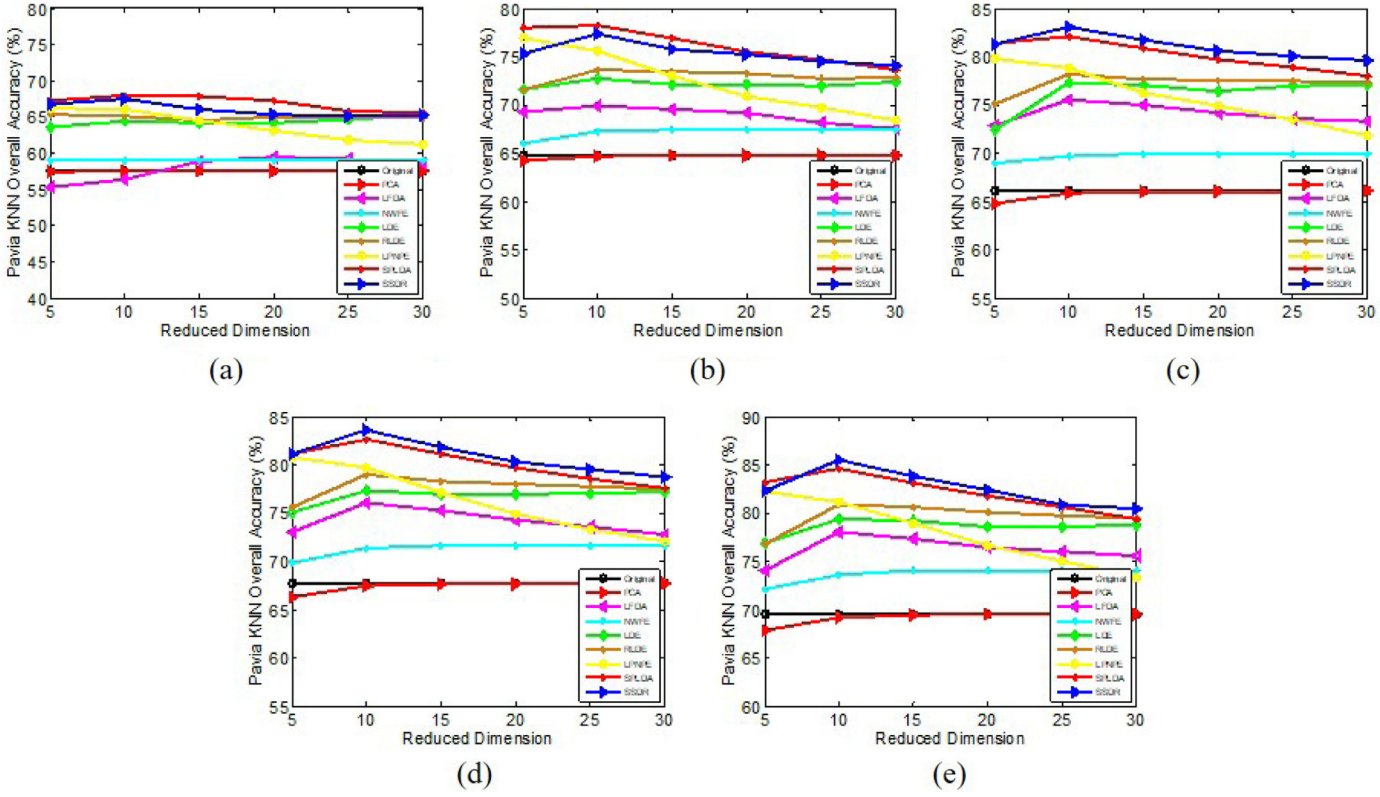

Fig. 9. OAs with respect to: (a)–(e) different sizes of training set (10, 20, 30, 40, 50) and different reduced dimensions (from 5 to 30) for the Pavia University data set, combined with the KNN classifier.

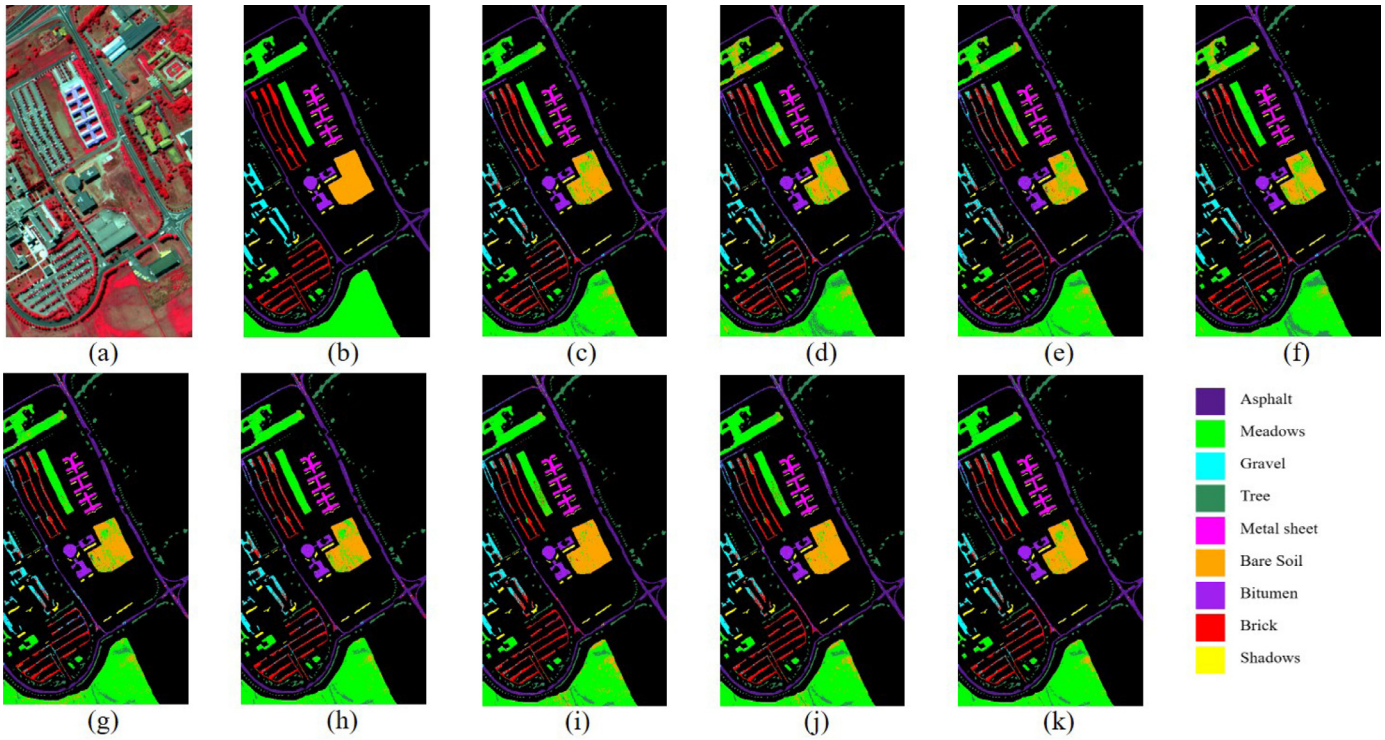

Fig. 10. SVM classification maps of the different methods with the Pavia University data in dim=10. (a) False-color image of the Pavia University data set (R: 102, G: 56, B: 31). (b) Ground-truth map. (c) Original (SVM). (d) PCA. (e) LFDA. (f) NWFE. (g) LDE. (h) RLDE. (i) LPNPE. (j) SPLDA. (K) SSDR.

Table 2
Classification accuracies of the different DR methods for the Pavia University data set.

Samples	Classifiers	Index	Original	PCA	LFDA	NWFE	LDE	RLDE	LPNPE	SPLDA	SSDR
10	KNN	OA	57.56 ± 4.60	57.54 ± 4.59	56.37 ± 4.15	59.05 ± 4.31	64.41 ± 4.79	65.18 ± 4.03	66.07 ± 5.66	**68.01 ± 5.46**	67.52 ± 5.54
		AA	72.06 ± 2.28	72.02 ± 2.24	65.16 ± 2.92	75.12 ± 2.15	73.83 ± 2.06	77.05 ± 2.20	76.62 ± 2.55	78.29 ± 2.76	**79.11 ± 2.42**
		Kappa	48.85 ± 4.53	48.82 ± 4.54	46.59 ± 4.84	50.81 ± 4.35	56.12 ± 5.68	57.27 ± 5.69	58.62 ± 6.82	**60.83 ± 6.50**	60.26 ± 5.93
	SVM	OA	64.99 ± 3.04	64.51 ± 2.73	56.02 ± 4.38	66.55 ± 3.38	67.06 ± 3.46	70.09 ± 4.48	73.09 ± 3.77	74.38 ± 3.64	**74.67 ± 4.03**
		AA	75.67 ± 3.09	75.28 ± 2.81	64.72 ± 2.97	77.75 ± 1.89	73.90 ± 2.29	78.31 ± 2.11	78.89 ± 2.20	79.70 ± 1.49	**81.38 ± 1.60**
		Kappa	56.77 ± 3.43	56.17 ± 3.03	46.24 ± 4.69	58.72 ± 3.57	58.86 ± 3.51	62.73 ± 4.96	66.37 ± 4.12	67.83 ± 3.97	**68.14 ± 4.37**
20	KNN	OA	64.82 ± 2.54	64.67 ± 2.54	69.89 ± 2.49	67.23 ± 2.54	72.81 ± 2.28	73.61 ± 2.34	75.63 ± 1.98	78.23 ± 1.64	**78.28 ± 2.59**
		AA	75.89 ± 0.79	75.76 ± 0.79	78.76 ± 0.75	79.33 ± 1.00	81.42 ± 0.94	82.60 ± 1.13	81.62 ± 1.20	83.97 ± 0.93	**85.13 ± 1.08**
		Kappa	56.56 ± 2.66	56.39 ± 2.66	62.56 ± 2.68	59.49 ± 2.71	65.98 ± 2.62	66.93 ± 2.65	69.35 ± 2.28	72.46 ± 1.91	**72.56 ± 2.96**
	SVM	OA	77.43 ± 1.93	73.97 ± 2.27	70.96 ± 2.12	75.45 ± 2.97	77.62 ± 2.72	79.57 ± 2.75	78.79 ± 1.13	80.13 ± 2.26	**82.53 ± 2.07**
		AA	83.19 ± 1.42	81.38 ± 1.48	76.38 ± 1.37	83.45 ± 1.41	82.93 ± 1.69	84.92 ± 1.41	82.90 ± 1.27	84.07 ± 1.22	**86.62 ± 1.14**
		Kappa	71.15 ± 2.38	67.02 ± 2.55	63.33 ± 2.45	68.96 ± 2.40	71.49 ± 3.23	73.89 ± 2.66	73.05 ± 1.39	75.66 ± 2.68	**77.59 ± 2.53**
30	KNN	OA	66.11 ± 1.04	65.85 ± 1.05	75.55 ± 1.91	69.69 ± 1.24	77.30 ± 1.84	78.17 ± 1.54	78.88 ± 2.49	82.11 ± 2.06	**83.11 ± 1.93**
		AA	76.90 ± 0.90	76.71 ± 0.96	82.62 ± 1.12	80.97 ± 1.05	83.99 ± 0.94	84.99 ± 0.71	83.41 ± 1.09	85.98 ± 1.21	**87.29 ± 0.97**
		Kappa	58.03 ± 1.26	57.72 ± 1.26	69.11 ± 2.37	62.32 ± 1.43	71.25 ± 2.20	72.31 ± 2.17	73.15 ± 2.90	77.07 ± 2.32	**78.32 ± 2.01**
	SVM	OA	80.87 ± 2.19	77.16 ± 1.72	77.80 ± 2.18	78.09 ± 2.11	81.20 ± 2.30	82.81 ± 1.33	82.88 ± 1.51	83.82 ± 1.56	**85.54 ± 1.88**
		AA	85.25 ± 1.57	83.40 ± 0.98	82.13 ± 1.46	84.05 ± 0.92	85.41 ± 1.21	86.18 ± 1.46	85.32 ± 1.35	86.15 ± 1.60	**88.09 ± 1.21**
		Kappa	75.41 ± 2.69	70.88 ± 2.06	71.60 ± 2.56	71.99 ± 2.47	75.87 ± 2.78	77.75 ± 1.68	77.98 ± 1.87	79.11 ± 1.96	**81.26 ± 2.30**
40	KNN	OA	67.73 ± 0.79	67.50 ± 0.70	76.06 ± 1.40	71.40 ± 0.90	77.33 ± 1.26	79.05 ± 1.10	79.73 ± 1.63	82.63 ± 1.13	**83.60 ± 0.77**
		AA	78.50 ± 0.47	78.28 ± 0.46	84.16 ± 0.89	82.14 ± 0.31	84.94 ± 0.80	85.91 ± 0.60	84.46 ± 0.64	87.11 ± 0.57	**88.14 ± 0.52**
		Kappa	59.98 ± 0.77	59.69 ± 0.66	69.86 ± 2.37	64.34 ± 0.89	71.41 ± 1.37	73.42 ± 1.26	74.21 ± 1.93	77.78 ± 1.36	**78.97 ± 0.94**
	SVM	OA	82.53 ± 2.56	78.46 ± 2.23	80.19 ± 2.14	79.65 ± 2.59	82.47 ± 1.51	85.30 ± 2.01	83.67 ± 1.04	84.98 ± 1.91	**87.17 ± 1.64**
		AA	86.76 ± 1.05	84.61 ± 1.14	84.94 ± 0.36	85.77 ± 0.77	86.82 ± 0.62	88.51 ± 0.95	86.49 ± 0.76	87.56 ± 1.13	**89.68 ± 0.86**
		Kappa	77.54 ± 3.06	72.52 ± 2.65	74.69 ± 2.41	74.05 ± 2.89	77.48 ± 1.80	80.98 ± 1.51	79.00 ± 1.20	80.62 ± 2.29	**83.38 ± 2.01**
50	KNN	OA	69.54 ± 1.53	69.16 ± 1.48	78.03 ± 1.44	73.62 ± 1.68	79.39 ± 1.31	80.94 ± 0.91	81.17 ± 1.11	84.60 ± 0.73	**85.51 ± 1.06**
		AA	79.22 ± 1.04	78.89 ± 1.08	85.29 ± 0.71	83.04 ± 0.67	85.85 ± 0.48	86.72 ± 0.25	85.00 ± 0.70	87.99 ± 0.33	**89.08 ± 0.44**
		Kappa	61.88 ± 1.67	61.41 ± 1.64	72.13 ± 1.83	66.79 ± 1.82	73.79 ± 1.39	75.61 ± 1.00	75.90 ± 1.32	80.14 ± 0.87	**81.28 ± 1.27**
	SVM	OA	83.61 ± 2.42	79.60 ± 1.92	82.70 ± 1.25	80.38 ± 2.82	83.84 ± 1.03	85.76 ± 1.70	84.82 ± 1.50	86.23 ± 1.36	**88.40 ± 1.34**
		AA	87.83 ± 1.07	85.70 ± 0.90	86.36 ± 1.02	86.50 ± 1.33	87.79 ± 0.69	88.91 ± 0.93	87.60 ± 0.99	89.00 ± 0.69	**90.51 ± 0.55**
		Kappa	78.86 ± 2.93	73.91 ± 2.27	77.68 ± 1.44	74.91 ± 2.50	79.17 ± 1.24	81.54 ± 2.10	80.43 ± 1.83	82.20 ± 1.66	**84.91 ± 1.64**

tion. SPLDA also outperforms the other spectral-based DR methods, especially in the case of small numbers of training samples. From Fig. 9, it can be observed that, for the KNN classifier, the SSDR algorithm achieves the best OA value, up until the point where the sample number is larger than 30. This is mainly because the contribution of the label-guided graph is limited when the number of training samples is small.

In order to further show the performance of the proposed algorithms, Fig. 10 provides a visual depiction of the classification maps obtained by the different methods for the Pavia University

data set using the SVM classifier in dim=10. The false-color image and ground truth are also presented in Fig. 10. It can be observed that the proposed SPLDA and SSDR algorithms outperform the other methods. For the bare soil area, SPLDA performs better than SSDR, but in the meadows, SSDR performs better than SPLDA. Similarly, we implemented 10 independent runs with the different methods, and the average results of the OA, AA, and kappa values in dim=10 are listed in Table 2. The results of the SVM and KNN classifiers without DR are used as the baselines. SPLDA performs better than LPNPE, which shows the effectiveness of using SLIC superpixels in the proposed approach. The proposed SSDR algorithm achieves the best classification performance in terms of OA, AA, and kappa coefficient. The classification results indicate that SSDR not only explores the spectral similarity by the label-guided graph but it also makes full use of the spatial consistency property by the superpixel-guided graph, to enhance the classification performance.

5. Conclusions

In this paper, we have proposed a spatially regularized DR approach named SPLDA to preserve the spatial similarity during DR. Furthermore, we have also proposed SSDR to exploit the spatial and spectral information concurrently. Specifically, the HSI is first segmented into adaptive regions by the SLIC superpixel segmentation method, where each superpixel is considered to be homogeneous. A superpixel-guided graph is then constructed to capture the spatial similarity from the superpixels, and a label-guided graph is learned to explore the spectral similarity. Finally, we combine the superpixel-guided graph and the label-guided graph to explore the spatial and spectral similarity simultaneously. The proposed SPLDA and SSDR methods were tested on two HSI data sets, and achieved better classification results than the other widely used DR algorithms. Our future work will focus on the automatic selection of parameters during feature combination, to further improve the computational efficiency and classification accuracy.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China under grant 2018YFB0504500; and in part by the National Natural Science Foundation of China under grant 41571362 and grant 61871298.

References

[1] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Trans. Geosci. Remote Sens. 42 (8) (2004) 1778–1790.
[2] W. Li, G. Wu, F. Zhang, Q. Du, Hyperspectral image classification using deep pixel-pair features, IEEE Trans. Geosci. Remote Sens. 55 (2) (2017) 844–853.
[3] Y. Dong, B. Du, L. Zhang, L. Zhang, D. Tao, Lam3l: locally adaptive maximum margin metric learning for visual data classification, Neurocomputing 235 (2017) 1–9.
[4] H. Zhai, H. Zhang, L. Zhang, P. Li, Total variation regularized collaborative representation clustering with a locally adaptive dictionary for hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 57 (1) (2019) 166–180.
[5] R.M. Willett, M.F. Duarte, M.A. Davenport, R.G. Baraniuk, Sparsity and structure in hyperspectral imaging: sensing, reconstruction, and target detection, IEEE Signal Process. Mag. 31 (1) (2014) 116–126.
[6] H. Zhai, H. Zhang, L. Zhang, P. Li, Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery, ISPRS J. Photogramm. Remote Sens. 144 (2018) 235–253.
[7] W. He, H. Zhang, L. Zhang, Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (9) (2016) 4267–4279.
[8] B. Ayerdi, M. Graña, Hyperspectral image nonlinear unmixing and reconstruction by elm regression ensemble, Neurocomputing 174 (2016) 299–309.
[9] X. Lu, H. Wu, Y. Yuan, P. Yan, X. Li, Manifold regularized sparse NMF for hyperspectral unmixing, IEEE Trans. Geosci. Remote Sens. 51 (5) (2013) 2815–2826.
[10] W. Sun, C. Liu, J. Li, Y.M. Lai, W. Li, Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery, J. Appl. Remote Sens. 8 (1) (2014) 083641.
[11] A. Ertürk, M.-D. Iordache, A. Plaza, Sparse unmixing-based change detection for multitemporal hyperspectral images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (2) (2016) 708–719.
[12] Y. Yuan, H. Lv, X. Lu, Semi-supervised change detection method for multi-temporal hyperspectral images, Neurocomputing 148 (2015) 363–375.
[13] W. He, H. Zhang, H. Shen, L. Zhang, Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (3) (2018) 713–729.
[14] G. Camps-Valls, D. Tuia, L. Bruzzone, J.A. Benediktsson, Advances in hyperspectral image classification: earth monitoring with statistical learning methods, IEEE Signal Process. Mag. 31 (1) (2014) 45–54.
[15] T. Lillesand, R.W. Kiefer, J. Chipman, Remote Sensing and Image Interpretation, John Wiley & Sons, 2014.
[16] L. Zhang, Y. Zhong, B. Huang, J. Gong, P. Li, Dimensionality reduction based on clonal selection for hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 45 (12) (2007) 4172–4186.
[17] W. He, H. Zhang, L. Zhang, W. Philips, W. Liao, Weighted sparse graph based dimensionality reduction for hyperspectral images, IEEE Geosci. Remote Sens. Lett. 13 (5) (2016) 686–690.
[18] W. Li, S. Prasad, J.E. Fowler, L.M. Bruce, Locality-preserving dimensionality reduction and classification for hyperspectral image analysis, IEEE Trans. Geosci. Remote Sens. 50 (4) (2012) 1185–1198.
[19] A. Plaza, P. Martinez, J. Plaza, R. Perez, Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations, IEEE Trans. Geosci. Remote Sens. 43 (3) (2005) 466–479.
[20] S. Liu, Q. Du, X. Tong, A. Samat, H. Pan, X. Ma, Band selection-based dimensionality reduction for change detection in multi-temporal hyperspectral images, Remote Sens. 9 (10) (2017) 1008.
[21] F. Luo, B. Du, L. Zhang, L. Zhang, D. Tao, Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image, IEEE Trans. Cybern. 49 (7) (2019) 2406–2419.
[22] Y. Yuan, X. Zheng, X. Lu, Discovering diverse subset for unsupervised hyperspectral band selection, IEEE Trans. Image Process. 26 (1) (2017) 51–64.
[23] W. Sun, L. Zhang, B. Du, W. Li, Y.M. Lai, Band selection using improved sparse subspace clustering for hyperspectral imagery classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (6) (2015) 2784–2797.
[24] H. Zhai, H. Zhang, L. Zhang, P. Li, Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection, IEEE Trans. Geosci. Remote Sens. 57 (3) (2019) 1723–1740.
[25] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[26] H.-T. Chen, H.-W. Chang, T.-L. Liu, Local discriminant embedding and its variants, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2005, 2, IEEE, 2005, pp. 846–853.
[27] Y. Zhou, J. Peng, C.P. Chen, Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 53 (2) (2015) 1082–1095.
[28] N.H. Ly, Q. Du, J.E. Fowler, Sparse graph-based discriminant analysis for hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 52 (7) (2014) 3872–3884.
[29] X. He, P. Niyogi, Locality preserving projections, in: Proceedings of the Advances in Neural Information Processing Systems, 2004, pp. 153–160.
[30] K. Fukunaga, Introduction to Statistical Pattern Recognition, Elsevier, 2013.
[31] I. Jolliffe, Principal component analysis, in: International Encyclopedia of Statistical Science, Springer, 2011, pp. 1094–1096.
[32] J.B. Tenenbaum, V. De Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[33] M. Sugiyama, T. Idé, S. Nakajima, J. Sese, Semi-supervised local fisher discriminant analysis for dimensionality reduction, Mach. Learn. 78 (1–2) (2010) 35.
[34] D. Zhang, Z.-H. Zhou, S. Chen, Semi-supervised dimensionality reduction, in: Proceedings of the SIAM International Conference on Data Mining, SIAM, 2007, pp. 629–634.
[35] K. Tan, S. Zhou, Q. Du, Semisupervised discriminant analysis for hyperspectral imagery with block-sparse graph, IEEE Geosci. Remote Sens. Lett. 12 (8) (2015) 1765–1769.
[36] Y. Dong, B. Du, L. Zhang, L. Zhang, Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning., IEEE Trans. Geosci. Remote Sens. 55 (5) (2017) 2509–2524.
[37] F. Feng, W. Li, Q. Du, B. Zhang, Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity, Remote Sens. 9 (4) (2017) 323.
[38] L. Fang, S. Li, W. Duan, J. Ren, J.A. Benediktsson, Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels, IEEE Trans. Geosci. Remote Sens. 53 (12) (2015) 6663–6674.
[39] P. Ghamisi, J.A. Benediktsson, M.O. Ulfarsson, Spectral-spatial classification of hyperspectral images based on hidden markov random fields., IEEE Trans. Geosci. Remote Sens. 52 (5) (2014) 2565–2574.

[40] S. Li, T. Lu, L. Fang, X. Jia, J.A. Benediktsson, Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 54 (12) (2016) 7416–7430.
[41] J.A. Benediktsson, J.A. Palmason, J.R. Sveinsson, Classification of hyperspectral data from urban areas based on extended morphological profiles, IEEE Trans. Geosci. Remote Sens. 43 (3) (2005) 480–491.
[42] J. Li, P.R. Marpu, A. Plaza, J.M. Bioucas-Dias, J.A. Benediktsson, Generalized composite kernel framework for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 51 (9) (2013) 4816–4829.
[43] Y. Zhou, J. Peng, C.L.P. Chen, Extreme learning machine with composite kernels for hyperspectral image classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (6) (2015) 2351–2360, doi:10.1109/JSTARS.2014.2359965.
[44] L. Fang, S. Li, X. Kang, J.A. Benediktsson, Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model, IEEE Trans. Geosci. Remote Sens. 53 (8) (2015) 4186–4201.
[45] S. Jiay, B. Dengy, X. Jiaz, Superpixel-level sparse representation-based classification for hyperspectral imagery, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 2016, pp. 3302–3305.
[46] Y. Fang, H. Li, Y. Ma, K. Liang, Y. Hu, S. Zhang, H. Wang, Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding, IEEE Geosci. Remote Sens. Lett. 11 (10) (2014) 1712–1716.
[47] H. Pu, Z. Chen, B. Wang, G.-M. Jiang, A novel spatial–spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 52 (11) (2014) 7008–7022.
[48] L. Ma, X. Zhang, X. Yu, D. Luo, Spatial regularized local manifold learning for classification of hyperspectral images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (2) (2016) 609–624.
[49] Z. Feng, S. Yang, S. Wang, L. Jiao, Discriminative spectral–spatial margin-based semisupervised dimensionality reduction of hyperspectral data, IEEE Geosci. Remote Sens. Lett. 12 (2) (2015) 224–228.
[50] H. Huang, G. Shi, H. He, Y. Duan, F. Luo, Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning, IEEE Trans. Cybern. (2019) 1–13, doi:10.1109/TCYB.2019.2905793.
[51] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, et al., Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[52] H. Xu, H. Zhang, W. He, L. Zhang, Superpixel based dimension reduction for hyperspectral imagery, in: Proceedings of the IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 2575–2578, doi:10.1109/IGARSS.2018.8519070.
[53] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51.
[54] F.R. Chung, F.C. Graham, Spectral Graph Theory, American Mathematical Society, 1997.
[55] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, B. Schölkopf, Learning with local and global consistency, in: Proceedings of the Advances in Neural Information Processing Systems, 2004, pp. 321–328.
[56] M. Sugiyama, Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis, J. Mach. Learn. Res. 8 (May) (2007) 1027–1061.
[57] B.-C. Kuo, D.A. Landgrebe, Nonparametric weighted feature extraction for classification, IEEE Trans. Geosci. Remote Sens. 42 (5) (2004) 1096–1105.

Huilin Xu received the B.S. degree in College of Geological Engineering and Geomatics from Changan University, Xian, China, in 2016, and the M.S. degree in Surveying, Mapping and Remote Sensing (LIESMARS) from Wuhan University, Wuhan, China, in 2018, where she is currently pursuing the Ph.D. degree. Her research interests include hyperspectral image processing and deep learning.

Hongyan Zhang (M'13-SM'16) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2005 and 2010, respectively.

He has been currently a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University since 2016. He has authored/co-authored more than 70 research papers. His research interests include image reconstruction for quality improvement, hyperspectral image processing, sparse representation and low rank methods for sensing image imagery.

Dr. Zhang currently serves as an Associate Editor of the Computers & Geosciences Journal. He is a Reviewer of more than 30 international academic journals, including IEEE Transaction on Geoscience and remote sensing, IEEE Transaction on Image processing, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Geoscience and remote sensing Letters and so on.

Wei He (S'14-M'17) received the B.S. degree in School of Mathematics and statistics and the Ph.D degree in Surveying, Mapping and Remote Sensing (LIESMARS) from Wuhan University, Wuhan, China, in 2012 and 2017, respectively.

He is currently a researcher with the Geoinformatics unit, RIKEN Center for Advanced Intelligence Project, Japan. His research interests include image quality improvement, remote sensing image processing and low rank representation.

Liangpei Zhang (M'06-SM'08) received the B.S. degree in physics from Hunan Normal University, ChangSha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in Photogrammetry and Remote Sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 500 research papers and five books. He is the holder of 15 patents. He has won the 2010 the best paper Boeing award and 2013 best paper ERDAS award of American Society of Photogrammetry and Remote Sensing (ASPRS), respectively His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He is a Fellow of the Institution of Engineering and Technology, an Executive Member (Board of Governors) of the China National Committee of the International GeosphereCBiosphere Programme, an Executive Member of the China Society of Image and Graphics,etc. He regularly serves as a Cochair of the series International Society for Optics and Photonics (SPIE) Conferences on Multispectral Image Processing and Pattern Recognition, the Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor of the International Journal of Ambient Computing and Intelligence, the International Journal of Image and Graphics, the International Journal of Digital Multimedia Broadcasting, the Journal of Geo-spatial Information Science, the Journal of Remote Sensing, and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING; he is a Guest Editor of the Journal of Applied Remote Sensing and the Journal of Sensors. He was the General Chair for the Fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) and a Guest Editor of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS). He was the recipient of the Best Reviewer Awards from the IEEE GRSS for his service to the IEEE JSTARS in 2012 and the IEEE Geoscience and Remote Sensing Letters in 2014. He was also the recipient of the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing. His research teams were the recipient of the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium Student Paper Contest in recent years.