

Article

Local Geometric Structure Feature for Dimensionality Reduction of Hyperspectral Imagery

Fulin Luo ^{1,2} , Hong Huang ^{2,*} , Yule Duan ², Jiamin Liu ² and Yinghua Liao ³

¹ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; luoflyn@163.com

² Key Laboratory of Optoelectronic Technology and Systems of the Education Ministry of China, Chongqing University, Chongqing 400044, China; duanyule@cqu.edu.cn (Y.D.); liujm@cqu.edu.cn (J.L.)

³ Department of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong 643000, China; liaoyinghua118@163.com

* Correspondence: hhuang@cqu.edu.cn; Tel.: +86-23-6510-6109

Academic Editors: Yanfei Zhong and Prasad S. Thenkabail

Received: 10 July 2017; Accepted: 29 July 2017; Published: 1 August 2017

Abstract: Marginal Fisher analysis (MFA) exploits the margin criterion to compact the intraclass data and separate the interclass data, and it is very useful to analyze the high-dimensional data. However, MFA just considers the structure relationships of neighbor points, and it cannot effectively represent the intrinsic structure of hyperspectral imagery (HSI) that possesses many homogenous areas. In this paper, we propose a new dimensionality reduction (DR) method, termed local geometric structure Fisher analysis (LGSFA), for HSI classification. Firstly, LGSFA uses the intraclass neighbor points of each point to compute its reconstruction point. Then, an intrinsic graph and a penalty graph are constructed to reveal the intraclass and interclass properties of hyperspectral data. Finally, the neighbor points and corresponding intraclass reconstruction points are used to enhance the intraclass-manifold compactness and the interclass-manifold separability. LGSFA can effectively reveal the intrinsic manifold structure and obtain the discriminating features of HSI data for classification. Experiments on the Salinas, Indian Pines, and Urban data sets show that the proposed LGSFA algorithm achieves the best classification results than other state-of-the-art methods.

Keywords: hyperspectral imagery; dimensionality reduction; manifold learning; local geometric structure; marginal Fisher analysis

1. Introduction

Hyperspectral imagery (HSI) is captured by remote sensors recording the reflectance values of electromagnetic wave, and each pixel in HSI is a spectral curve containing hundreds of bands from visible to near-infrared spectrum [1–4]. The HSI can offer much richer information, and it can discriminate the subtle differences in different land cover types [5,6]. HSI plays a significant role in the application of anomaly detection, agricultural production, disaster warning, and land cover classification [7–9]. However, the traditional classification methods commonly cause the Hughes phenomena because of the high dimensional characteristics in HSI [10–12]. Therefore, a huge challenge for HSI processing is to reduce the dimensionality of high-dimensional data with some valuable intrinsic information preserved.

Dimensionality reduction (DR) is commonly applied to reduce the number of bands in HSI and obtain some desired information [13–15]. A large number of methods have been designed for DR of HSI. Principal component analysis (PCA) has been widely used for high-dimensional data, and it applies orthogonal projection to maximize data variance [16]. To improve the noise robustness of PCA, the researchers proposed minimum noise fraction (MNF) with noise variance [17]. MNF maximizes

the signal-to-noise ratio to obtain the principal components, and it provides satisfactory results for HSI classification. However, PCA and MNF are unsupervised methods, and they restrain the discriminating power for HSI classification. To enhance the discriminating power, some supervised methods have been proposed. Linear discriminant analysis (LDA) is a traditional supervised method based on the mean vector and covariance matrix of classes, and it is defined by the maximization of between-class scatter and the minimization of within-class scatter [18]. However, LDA only involves $c - 1$ features (where c is the class number of data), which may not obtain sufficient features for hyperspectral classification. To address this problem, the researchers proposed maximum margin criterion (MMC) [19] and local Fisher discriminant analysis (LFDA) [20,21] for obtaining enough features, while these methods originate from the theory of statistics and neglect the geometry properties of hyperspectral data.

Recently, the intrinsic manifold structure has been discovered in HSI [22]. Many manifold learning methods have been applied to obtain the manifold properties from high-dimensional data [23–26]. Such methods include isometric mapping (Isomap) [27], Laplacian eigenmaps (LE) [28] and locally linear embedding (LLE) [29]. The Isomap method adopts the geodesic distances between data points to reduce the dimensionality of data. The LLE method preserves the local linear structure of data in a low-dimensional space. The LE method applies the Laplacian matrix to reveal the local neighbor information of data. However, these manifold learning algorithms cannot obtain explicit projection matrix that can map a new sample into the corresponding low-dimensional space. To overcome this problem, locality preserving projections (LPP) [30] and neighborhood preserving embedding (NPE) [31] were proposed to approximately linearize the LE and LLE algorithms, respectively. However, LPP and NPE are unsupervised DR methods that cannot perform good discriminating power in certain scenes.

To unify these methods, a graph embedding (GE) framework has been proposed to analyze the DR methods on the basis of statistics or geometry theory [32]. Many algorithms, such as PCA, LDA, LPP, ISOMAP, LLE and LE, can be redefined in this framework. The differences between these algorithms lie in the computation of the similarity matrix and the selection of the constraint matrix. With this framework, marginal Fisher analysis (MFA) is developed for DR. MFA designs an intrinsic graph to characterize the intraclass compactness and a penalty graph to characterize the interclass separability. The intrinsic graph represents the similarity of intraclass points from the same class, while the penalty graph illustrates the connected relationship of interclass points that belongs to different classes. MFA can reveal the intraclass and interclass manifold structures. However, it only considers the structure relationships of pairwise neighbor points, which may not effectively represent the intrinsic structure relationship of HSI with a larger number of homogenous areas [33,34]. Therefore, MFA may not obtain good discriminating power for HSI classification.

To address this problem, we propose a new DR method called local geometric structure Fisher analysis (LGSFA) in this paper. Firstly, it reconstructs each point with intraclass neighbor points. In constructing the intrinsic graph and the penalty graph, it compacts the intraclass neighbor points and corresponding reconstruction points, and it simultaneously separates the interclass neighbor points and corresponding reconstruction points. LGSFA can better represent the intrinsic manifold structures, and it also enhances the intraclass compactness and the interclass separability of HSI data. Experimental results on three real hyperspectral data sets show that the proposed LGSFA algorithm is more effective than other DR methods to extract the discrimination features for HSI classification.

The rest of this paper is organized as follows. Section 2 briefly reviews the theories of GE and MFA. Section 3 details our proposed method. Experimental results are presented in Section 4 to demonstrate the effectiveness of the proposed method. Finally, Section 5 provides some concluding remarks and suggestions for future work.

2. Related Works

Let us suppose a data set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{D \times n}$, where n and D are the number of samples and bands in HSI, respectively. The class label of \mathbf{x}_i is denoted as $l_i \in \{1, 2, \dots, c\}$, where c is the

number of classes. The low-dimensional data is represented as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in \mathbb{R}^{d \times n}$, where d is the embedding dimensionality. \mathbf{Y} is represented by $\mathbf{Y} = \mathbf{V}^T \mathbf{X}$ with projection matrix $\mathbf{V} \in \mathbb{R}^{D \times d}$.

2.1. Graph Embedding

The graph embedding (GE) framework is used to unify most popular DR algorithms [32]. In GE, an intrinsic graph is constructed to describe some of the desirable statistical or geometrical properties of data, while a penalty graph is utilized to represent some of the unwanted characteristics of data [35]. The intrinsic graph $\mathbf{G} = \{\mathbf{X}, \mathbf{W}\}$ and penalty graph $\mathbf{G}^P = \{\mathbf{X}, \mathbf{W}^P\}$ are two undirected weighted graphs with the weight matrices $\mathbf{W} \in \mathbb{R}^{n \times n}$ and $\mathbf{W}^P \in \mathbb{R}^{n \times n}$, where \mathbf{X} denotes the vertex set. Weight w_{ij} reveals the similarity characteristic of the edges between vertices i and j in \mathbf{G} , while weight w_{ij}^P refers to the dissimilarity structure between vertices i and j in \mathbf{G}^P .

The purpose of graph embedding is to project each vertex of the graph into a low-dimensional space that preserves the similarity between the vertex pairs. The objective function of the graph embedding framework is formulated as follows:

$$J(\mathbf{Y}) = \min_{\text{tr}(\mathbf{Y}^T \mathbf{H} \mathbf{Y})=h} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{y}_i - \mathbf{y}_j\|^2 w_{ij} = \mathbf{Y}^T \mathbf{L} \mathbf{Y} \tag{1}$$

where h is a constant, \mathbf{H} is a constraint matrix defined to find the non-trivial solution of (1), and \mathbf{L} is the Laplacian matrix of graph \mathbf{G} . Typically, \mathbf{H} is the Laplacian matrix of graph \mathbf{G}^P , that is, $\mathbf{H} = \mathbf{L}^P$. Laplacian matrices \mathbf{L} and \mathbf{L}^P can be reformulated as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \mathbf{D} = \text{diag}([\sum_{j=1}^n w_{1j}, \sum_{j=1}^n w_{2j}, \dots, \sum_{j=1}^n w_{nj}]), \mathbf{W} = [w_{ij}]_{i,j=1}^n \tag{2}$$

$$\mathbf{L}^P = \mathbf{D}^P - \mathbf{W}^P, \mathbf{D}^P = \text{diag}([\sum_{j=1}^n w_{1j}^P, \sum_{j=1}^n w_{2j}^P, \dots, \sum_{j=1}^n w_{nj}^P]), \mathbf{W}^P = [w_{ij}^P]_{i,j=1}^n \tag{3}$$

where $\text{diag}(\bullet)$ denotes that a vector is transformed as a diagonal matrix.

2.2. Marginal Fisher Analysis

MFA constructs an intrinsic graph and a penalty graph. The intrinsic graph connects each point with its neighbor points from the same class to characterize the intraclass compactness, while the penalty graph connects the marginal points from different classes to characterize the interclass separability.

In intrinsic graph \mathbf{G} , each data point \mathbf{x}_i is connected with the intraclass neighbor points that are from the same class. The similarity weight w_{ij} between \mathbf{x}_i and \mathbf{x}_j is defined as

$$w_{ij} = \begin{cases} 1, & \mathbf{x}_i \in N_1(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_1(\mathbf{x}_i) \text{ and } l_i = l_j \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $N_1(\mathbf{x}_i)$ denotes the k_1 intraclass neighbor points of \mathbf{x}_i . The k_1 intraclass neighbor points denotes the intraclass similarity relationship of data that should be preserved in low-dimensional embedding space.

For penalty graph \mathbf{G}^P , \mathbf{x}_i is connected with the interclass neighbor points that are from different classes. The penalty weight w_{ij}^P between \mathbf{x}_i and \mathbf{x}_j is set as

$$w_{ij}^P = \begin{cases} 1, & \mathbf{x}_i \in N_2(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_2(\mathbf{x}_i) \text{ and } l_i \neq l_j \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $N_2(x_i)$ denotes the k_2 interclass neighbor points of x_i . The k_2 interclass neighbor points represents the interclass similarity relationship of data that should be avoided in low-dimensional embedding space.

To enhance intraclass compactness and interclass separability, the optimal projection matrix \mathbf{V} can be obtained with the following optimization problem:

$$J(\mathbf{V}) = \min \frac{\sum_{i,j} \|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{x}_j\|^2 w_{ij}}{\sum_{i,j} \|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{x}_j\|^2 w_{ij}^p} = \frac{\mathbf{V}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{V}}{\mathbf{V}^T \mathbf{X} \mathbf{L}^p \mathbf{X}^T \mathbf{V}} \quad (6)$$

3. Local Geometric Structure Fisher Analysis

To effectively reveal the intrinsic manifold structure of hyperspectral data, a local geometric structure Fisher analysis (LGSFA) method was proposed based on MFA. This method computes the reconstruction point of each point with its intraclass neighbor points. Then, it uses the intraclass and interclass neighbor points to construct an intrinsic graph and a penalty graph, respectively. With the intrinsic graph, it utilizes the intraclass neighbor points and corresponding reconstruction points to compact the data points from the same class. With the penalty graph, it adopts the interclass neighbor points and corresponding reconstruction points to separate the data points from different classes. LGSFA further improves both the intraclass compactness and the interclass separability of hyperspectral data, and it can obtain better discriminating features to enhance the classification effect. The process of the LGSFA method is shown in Figure 1.

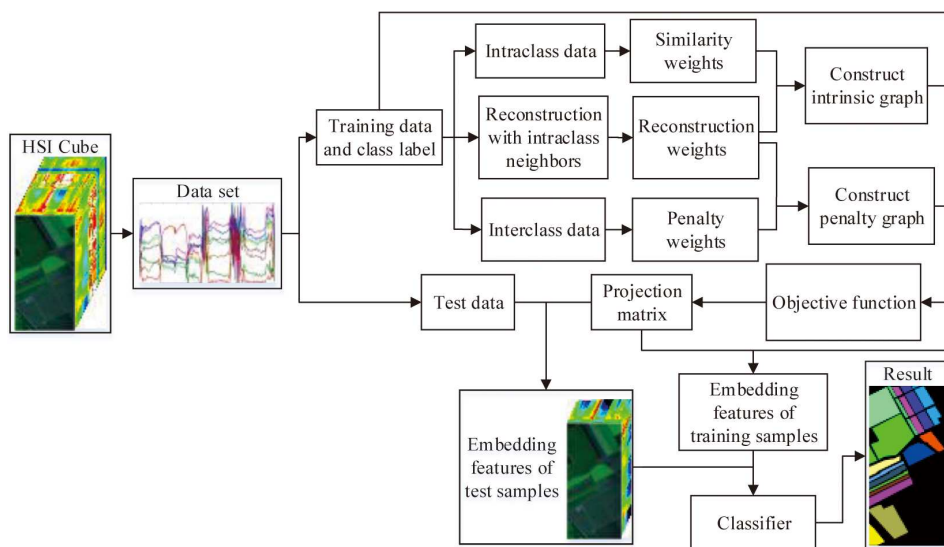


Figure 1. Process of the proposed LGSFA method.

Each point x_i can be reconstructed with its neighbor points from the same class. The reconstruction weight is computed by minimizing the sum of reconstruction errors

$$J(s_{ij}) = \min \sum_{i=1}^n \left\| \mathbf{x}_i - \sum_{j=1}^n s_{ij} \mathbf{x}_j \right\|^2 \quad (7)$$

where s_{ij} is the reconstruction weight between x_i and x_j . If x_j is the k_1 intraclass neighbor points of x_i , $s_{ij} \neq 0$ or $s_{ij} = 0$, and $\sum_{j=1}^n s_{ij} = 1$.

With some mathematical operations, (7) can be reduced as

$$\sum_{i=1}^n \left\| \mathbf{x}_i - \sum_{j=1}^n s_{ij} \mathbf{x}_j \right\|^2 = \sum_{i=1}^n \left\| \mathbf{x}_i \sum_{j=1}^n s_{ij} - \sum_{j=1}^n s_{ij} \mathbf{x}_j \right\|^2 = \sum_{i=1}^n \left\| [\mathbf{x}_i - \mathbf{x}_i^1, \mathbf{x}_i - \mathbf{x}_i^2, \dots, \mathbf{x}_i - \mathbf{x}_i^{k_1}] \mathbf{s}_i \right\|^2 = \sum_{i=1}^n \mathbf{s}_i^T \mathbf{z}_i \mathbf{s}_i \tag{8}$$

where $\mathbf{z}_i = [z_{mn}^i]_{k_1 \times k_1} = [\mathbf{x}_i - \mathbf{x}_i^1, \mathbf{x}_i - \mathbf{x}_i^2, \dots, \mathbf{x}_i - \mathbf{x}_i^{k_1}]^T [\mathbf{x}_i - \mathbf{x}_i^1, \mathbf{x}_i - \mathbf{x}_i^2, \dots, \mathbf{x}_i - \mathbf{x}_i^{k_1}]$, $\mathbf{x}_i^{k_1}$ is the i th intraclass neighbor point of \mathbf{x}_i .

Thus, (7) can be denoted as

$$J(s_{ij}) = \min \sum_{i=1}^n \mathbf{s}_i^T \mathbf{z}_i \mathbf{s}_i, \quad \text{s.t.} \quad \sum_{j=1}^n s_{ij} = 1 \tag{9}$$

According to the method of Lagrangian multipliers, the optimization solution is

$$s_{ij} = \begin{cases} \frac{\sum_{m=1}^{k_1} (z_{jm}^i)^{-1}}{\sum_{p=1}^{k_1} \sum_{q=1}^{k_1} (z_{pq}^i)^{-1}}, & \text{if } \mathbf{x}_j \in N_1(\mathbf{x}_i) \text{ and } l_i = l_j \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $z_{mn}^i = (\mathbf{x}_i - \mathbf{x}_m^i)^T (\mathbf{x}_i - \mathbf{x}_n^i)$. After obtaining $\mathbf{s}_i = [s_{i1}, s_{i2}, \dots, s_{in}]^T$, the reconstruction point of \mathbf{x}_i can be represented as \mathbf{Xs}_i .

To reveal the manifold structure of hyperspectral data, we construct an intrinsic graph to characterize the similarity properties of data from the same class and a penalty graph to stress the dissimilarity of data from different classes. The similarity weight between \mathbf{x}_i and \mathbf{x}_j of the intrinsic graph are defined as

$$w_{ij}^w = \begin{cases} \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2t_i^2}), & \text{if } \mathbf{x}_i \in N_1(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_1(\mathbf{x}_i) \text{ and } l_i = l_j \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where $t_i = \frac{1}{n} \sum_{j=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|$.

For the penalty graph, the weights are represented as

$$w_{ij}^b = \begin{cases} \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2t_i^2}), & \text{if } \mathbf{x}_i \in N_2(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_2(\mathbf{x}_i) \text{ and } l_i \neq l_j \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

To illustrate the graph construction of the LGSFA method, an example is shown in Figure 2. In the intrinsic graph, point \mathbf{x}_2 is the intraclass neighbor point of \mathbf{x}_1 , and point \mathbf{x}_3 is the reconstruction point of \mathbf{x}_2 obtained by its intraclass neighbor points that are in the curve of dot dash line. In the penalty graph, point \mathbf{x}_4 is the interclass neighbor point of \mathbf{x}_5 , and point \mathbf{x}_6 is the reconstruction point of \mathbf{x}_5 . The intrinsic graph and the penalty graph are constructed on the basis of intraclass neighbor points, interclass neighbor points and corresponding reconstruction points. In these graphs, we consider the structure relationships not only between each point and its neighbor points but also between that point and the reconstruction points of its neighbor points. This process can effectively represent the intrinsic structure of HSI, and it can improve the compactness of data from the same class and the separability of data from different classes.

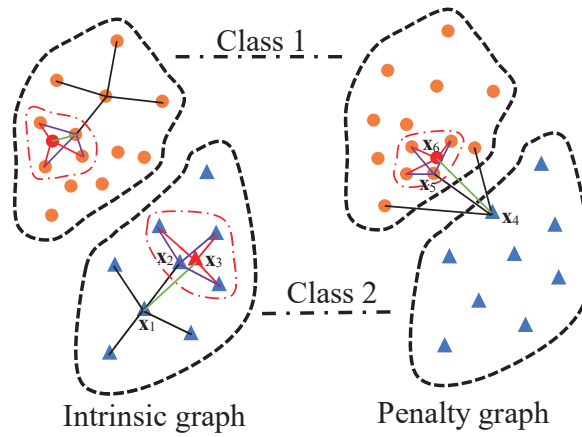


Figure 2. Graph construction of the LGSFA method.

To enhance the intraclass compactness, we apply the intraclass neighbor points and corresponding reconstruction points to construct an objective function in a low-dimensional embedding space

$$J_1(\mathbf{V}) = \min \sum_{i=1}^n \sum_{j=1}^n (\|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{x}_j\|^2 + \|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{X} \mathbf{s}_j\|^2) w_{ij}^w \quad (13)$$

With some mathematical operations, (13) can be reduced as

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{X} \mathbf{s}_j\|^2 w_{ij}^w \\ &= \frac{1}{2} \{ \mathbf{V}^T [\sum_{i=1}^n \sum_{j=1}^n (\mathbf{x}_i w_{ij}^w \mathbf{x}_i^T - \mathbf{x}_i w_{ij}^w \mathbf{s}_j^T \mathbf{X}^T + \mathbf{X} \mathbf{s}_j w_{ij}^w \mathbf{s}_j^T \mathbf{X}^T - \mathbf{X} \mathbf{s}_j w_{ij}^w \mathbf{x}_i^T)] \mathbf{V} \} \\ &= \frac{1}{2} [\mathbf{V}^T \mathbf{X} (\mathbf{D}^w - \mathbf{W}^w \mathbf{S}^T + \mathbf{S} \mathbf{W}^w \mathbf{S}^T - \mathbf{S}^T \mathbf{W}^w) \mathbf{X}^T \mathbf{V}] = \mathbf{V}^T \mathbf{X} \mathbf{B}_1^w \mathbf{X}^T \mathbf{V} \end{aligned} \quad (14)$$

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{x}_j\|^2 w_{ij}^w \\ &= \mathbf{V}^T [\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{x}_i w_{ij}^w \mathbf{x}_i^T - 2 \mathbf{x}_i w_{ij}^w \mathbf{x}_j^T + \mathbf{x}_j w_{ij}^w \mathbf{x}_j^T)] \mathbf{V} \\ &= \mathbf{V}^T \mathbf{X} (\mathbf{D}^w - \mathbf{W}^w) \mathbf{X}^T \mathbf{V} = \mathbf{V}^T \mathbf{X} \mathbf{B}_2^w \mathbf{X}^T \mathbf{V} \end{aligned} \quad (15)$$

where $\mathbf{B}_1^w = \frac{1}{2} (\mathbf{D}^w - \mathbf{W}^w \mathbf{S}^T + \mathbf{S} \mathbf{W}^w \mathbf{S}^T - \mathbf{S}^T \mathbf{W}^w)$, $\mathbf{B}_2^w = \mathbf{D}^w - \mathbf{W}^w$, $\mathbf{D}^w = \text{diag}([\sum_{j=1}^n w_{ij}^w]_{i=1}^n)$,

$\mathbf{W}^w = [w_{ij}^w]_{i,j=1}^n$, $\mathbf{S} = [s_{ij}]_{i,j=1}^n$.

According to (14) and (15), (13) can be represented as

$$\begin{aligned} J_1(\mathbf{V}) &= \min \sum_{i=1}^n \sum_{j=1}^n (\|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{x}_j\|^2 + \|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{X} \mathbf{s}_j\|^2) w_{ij}^w \\ &= \mathbf{V}^T \mathbf{X} (\mathbf{B}_1^w + \mathbf{B}_2^w) \mathbf{X}^T \mathbf{V} = \mathbf{V}^T \mathbf{X} \mathbf{M}^w \mathbf{X}^T \mathbf{V} \end{aligned} \quad (16)$$

where $\mathbf{M}^w = \mathbf{B}_1^w + \mathbf{B}_2^w$.

In addition, we also construct an objective function with interclass neighbor points and corresponding reconstruction points to improve the interclass separability as follows:

$$\begin{aligned} J_2(\mathbf{V}) &= \max \sum_{i=1}^n \sum_{j=1}^n (\|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{x}_j\|^2 + \|\mathbf{V}^T \mathbf{x}_i - \mathbf{V}^T \mathbf{X} \mathbf{s}_j\|^2) w_{ij}^b \\ &= \mathbf{V}^T \mathbf{X} \mathbf{M}^b \mathbf{X}^T \mathbf{V} \end{aligned} \quad (17)$$

where $\mathbf{M}^b = \mathbf{B}_1^b + \mathbf{B}_2^b$, $\mathbf{B}_1^b = \frac{1}{2}(\mathbf{D}^b - \mathbf{W}^b \mathbf{S}^T + \mathbf{S} \mathbf{W}^b \mathbf{S}^T - \mathbf{S}^T \mathbf{W}^b)$, $\mathbf{B}_2^b = \mathbf{D}^b - \mathbf{W}^b$, $\mathbf{D}^b = \text{diag}([\sum_{j=1}^n w_{ij}^b]_{i=1}^n)$, $\mathbf{W}^b = [w_{ij}^b]_{i,j=1}^n$.

To obtain a projection matrix, the optimization problems of (16) and (17) can be changed into another form as follows:

$$J(\mathbf{V}) = \max \frac{\mathbf{V}^T \mathbf{X} \mathbf{M}^b \mathbf{X}^T \mathbf{V}}{\mathbf{V}^T \mathbf{X} \mathbf{M}^w \mathbf{X}^T \mathbf{V}} \tag{18}$$

According to the method of Lagrangian multipliers, the optimization problem is transformed to solve a generalized eigenvalue problem, i.e.,

$$\mathbf{X} \mathbf{M}^b \mathbf{X}^T \mathbf{V} = \lambda \mathbf{X} \mathbf{M}^w \mathbf{X}^T \mathbf{V} \tag{19}$$

The optimization projection matrix $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d]$ can be obtained by the d minimum eigenvalues of (19) corresponding eigenvectors. Then, the low-dimensional features can be formulated by

$$\mathbf{Y} = \mathbf{V}^T \mathbf{X} \in \mathfrak{R}^{d \times n} \tag{20}$$

In summary, the proposed LGSFA method considers the neighbor points and corresponding reconstruction points to improve the intraclass compactness and the interclass separability of hyperspectral data. Therefore, it can effectively extract the discriminating feature for HSI classification. An example for the processing of LGSFA is shown in Figure 3.

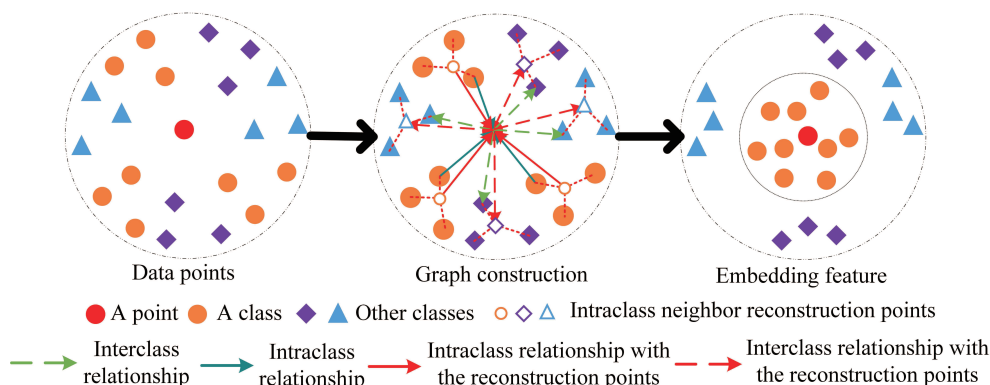


Figure 3. An example for the processing of LGSFA.

According to Figure 3, the proposed LGSFA method applies the intraclass, interclass, and intraclass neighbor reconstruction relationships to enhance the compactness of data from the same class and the separability of data from different classes. The detailed steps of the proposed LGSFA method are shown in Algorithm 1.

According to the process in Algorithm 1, we adopt big O notation to analyze the computational complexity of LGSFA. The number of intraclass neighbors and interclass neighbors is denoted as k_1 and k_2 , respectively. The reconstruction weight matrix \mathbf{S} is computed with the cost of $O(nk_1^3)$. The intraclass weight matrix \mathbf{W}^b and the interclass weight matrix \mathbf{W}^w take $O(nk_1)$ and $O(nk_2)$, respectively. The diagonal matrices \mathbf{D}^w and \mathbf{D}^b both cost $O(n)$. The intraclass manifold matrix \mathbf{M}_w and the interclass manifold matrix \mathbf{M}_b are both calculated with $O(n^3)$. The costs of $\mathbf{X} \mathbf{M}^b \mathbf{X}^T$ and $\mathbf{X} \mathbf{M}^w \mathbf{X}^T$ are both $O(Dn^2)$. It takes $O(D^3)$ to solve the generalized eigenvalue problem of (19). For $k_1 < n$ and $k_2 < n$, the total computational complexity of LGSFA is $O(n^3 + D^3 + Dn^2 + nk_1^3)$ that mainly depends on the number of bands, training samples, and neighbor points.

Algorithm 1 LGSFA

Input: data set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathfrak{R}^{D \times n}$ and corresponding class labels $\{l_1, l_2, \dots, l_n\}$, embedding

dimension d ($d < D$), the number of intraclass neighbors k_1 , the number of interclass neighbors k_2 .

1: Find k_1 intraclass neighborhood $N_1(\mathbf{x}_i)$ and k_2 interclass neighborhood $N_2(\mathbf{x}_i)$ of \mathbf{x}_i .

2: Compute the intraclass reconstruction weight of \mathbf{x}_i by

$$3: \min_{i=1}^n \|\mathbf{x}_i - \sum_{j=1}^n s_{ij} \mathbf{x}_j\|^2$$

4: where $s_{ij} \neq 0$ if $\mathbf{x}_j \in N_1(\mathbf{x}_i)$ and $\sum_{j=1}^n s_{ij} = 1$.

5: Calculate intraclass weight and interclass weight by

$$6: w_{ij}^w = \begin{cases} \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2t_i^2}), & \text{if } \mathbf{x}_i \in N_1(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_1(\mathbf{x}_i) \text{ and } l_i = l_j \\ 0, & \text{otherwise} \end{cases}$$

$$7: w_{ij}^b = \begin{cases} \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2t_i^2}), & \text{if } \mathbf{x}_i \in N_2(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_2(\mathbf{x}_i) \text{ and } l_i \neq l_j \\ 0, & \text{otherwise} \end{cases}$$

8: where $t_i = \frac{1}{n} \sum_{j=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|$.

9: Compute intraclass manifold matrix and interclass manifold matrix

$$10: \mathbf{M}^w = \mathbf{B}_1^w + \mathbf{B}_2^w, \mathbf{M}^b = \mathbf{B}_1^b + \mathbf{B}_2^b$$

11: where $\mathbf{B}_1^b = \frac{1}{2}(\mathbf{D}^b - \mathbf{W}^b \mathbf{S}^T + \mathbf{S} \mathbf{W}^b \mathbf{S}^T - \mathbf{S}^T \mathbf{W}^b)$, $\mathbf{B}_1^w = \frac{1}{2}(\mathbf{D}^w - \mathbf{W}^w \mathbf{S}^T + \mathbf{S} \mathbf{W}^w \mathbf{S}^T - \mathbf{S}^T \mathbf{W}^w)$,

$$\mathbf{B}_2^w = \mathbf{D}^w - \mathbf{W}^w, \mathbf{B}_2^b = \mathbf{D}^b - \mathbf{W}^b, \mathbf{D}^b = \text{diag}([\sum_{j=1}^n w_{ij}^b]_{i=1}^n), \mathbf{W}^b = [w_{ij}^b]_{i,j=1}^n, \mathbf{D}^w = \text{diag}([\sum_{j=1}^n w_{ij}^w]_{i=1}^n),$$

$$\mathbf{W}^w = [w_{ij}^w]_{i,j=1}^n, \mathbf{S} = [s_{ij}]_{i,j=1}^n.$$

12: Solve the generalized eigenvalue problem:

$$13: \mathbf{X} \mathbf{M}^b \mathbf{X}^T \mathbf{V} = \lambda \mathbf{X} \mathbf{M}^w \mathbf{X}^T \mathbf{V}$$

14: Obtain the projection matrix with the d smallest eigenvalues corresponding eigenvectors:

$$15: \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d] \in \mathfrak{R}^{D \times d}$$

Output: $\mathbf{Y} = \mathbf{V}^T \mathbf{X} \in \mathfrak{R}^{d \times n}$

4. Experimental Results and Discussion

We employed the Salinas, Indian Pines, and Urban HSI data sets to evaluate the proposed LGSFA method, and it was compared with some state-of-art DR algorithms.

4.1. Data Sets

Salinas data set: The HSI data set was collected by an airborne visible/infrared imaging spectrometer (AVIRIS) sensor over Salinas Valley, Southern California, in 1998. This data set has a geometric resolution of 3.7 m. The area possesses a spatial size of 512–217 pixels and 224 spectral bands from 400 nm to 2500 nm. Exactly 204 bands remained after the removal of bands 108–122, 154–167 and 224 as a result of dense water vapor and atmospheric effects. The data set contains sixteen land cover types. The scene in false color and its corresponding ground truth are shown in Figure 4.

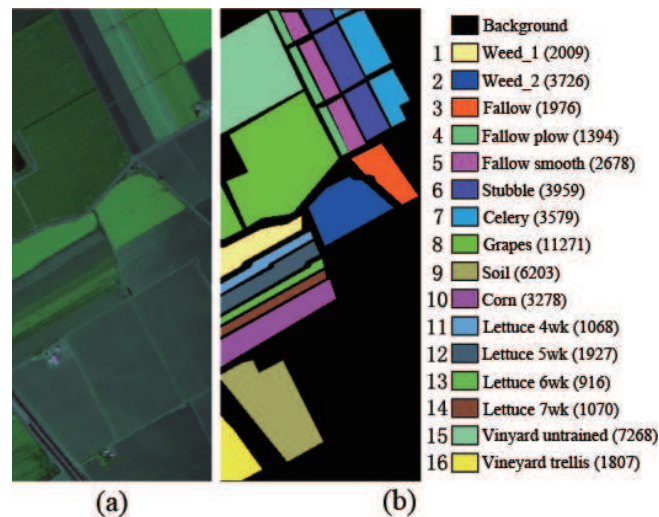


Figure 4. Salinas hyperspectral image. (a) HSI in false color; (b) Ground truth.

Indian Pines data set: This data set is a scene of the Northwest Indiana collected by the AVIRIS sensor in 1992. It consists of 145×145 pixels and 220 spectral bands within the range of 375–2500 nm. Several spectral bands, including bands 104–108, 150–163 and 220, with noise and water absorption phenomena were removed from the data set, leaving a total of 200 radiance channels to be used in the experiments. Sixteen ground truth classes of interest are considered in the data set. The scene in false color and its corresponding ground truth are shown in Figure 5.

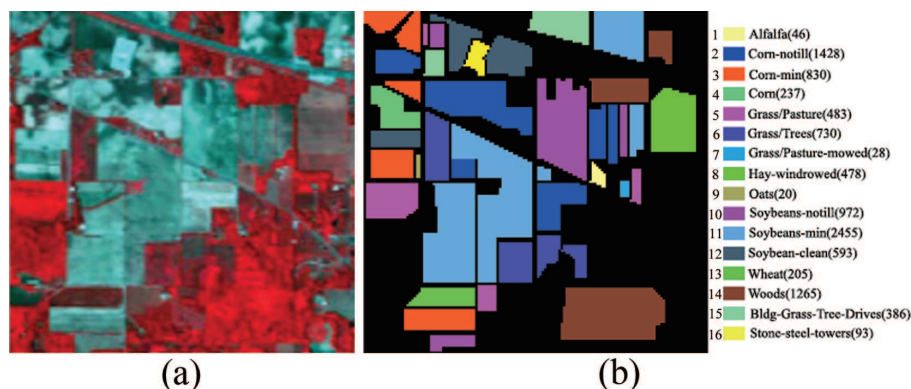


Figure 5. Indian Pines hyperspectral image. (a) HSI in false color; (b) Ground truth.

Urban data set: This data set was captured by the hyperspectral digital imagery collection experiment (HYDICE) sensor at the location of Copperas Cove, near Fort Hood, Texas, USA in October 1995. This data set is of size 307×307 pixels, and it is composed of 210 spectral channels with spectral resolution of 10 nm in the range from 400 to 2500 nm. After removing water absorption and low SNR bands, 162 bands were used for experiment. Six land cover types are considered in this data set. The scene in false color and its corresponding ground truth are shown in Figure 6.

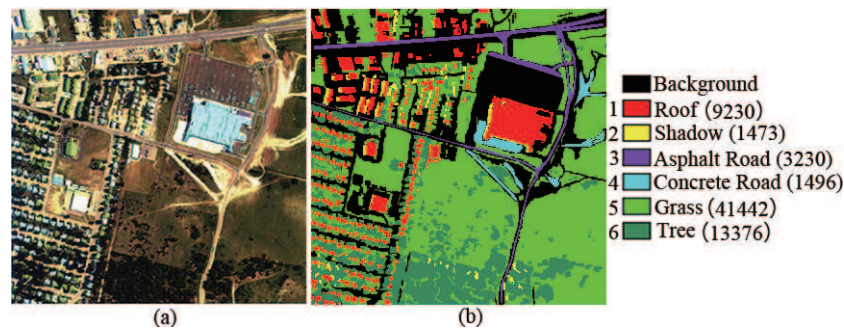


Figure 6. Urban hyperspectral image. (a) HSI in false color; (b) Ground truth.

4.2. Experimental Setup

In each experiment, the data set was randomly divided into training and test samples. A dimensionality reduction method was applied to learn a low-dimensional space with the training samples. Then, all test samples were mapped into a low-dimensional space. After that, we employed the nearest neighbor (NN) classifier, the spectral angle mapper (SAM) and the support vector machine based on composite kernels (SVMCK) [36] to classify test samples. For NN, it depends on the nearest Euclidean distance to discriminate the class of test samples. For SAM, the class of test samples is obtained by the smallest spectral angle. For SVMCK, it is an extensional SVM and simultaneously applies the spatial and spectral information of HSI to discriminate the class of test samples. Finally, the average classification accuracy (AA), the overall classification accuracy (OA), and the Kappa coefficient (KC) were adopted to evaluate the performance of each method. To robustly evaluate the results, the experiments were repeated 10 times in each condition, and we displayed the average classification accuracy with standard deviation (STD).

In the experiment, we compared the proposed LGSFA algorithm with the Baseline, PCA, NPE, LPP, sparse discriminant embedding (SDE) [13], LFDA, MMC and MFA methods, where the Baseline method represents that a classifier was directly used to classify the test samples without DR. To achieve optimal results for each method, we adopted cross-validation to obtain the optimal parameters of each method. For LPP, NPE and LFDA, the number of neighbor points was optimistically set to 9. For SDE, we set the error tolerances to 5. For MFA and LGSFA, we set the intraclass neighbor $k_1 = k$ and interclass neighbor $k_2 = \beta k$, where β is a positive integer. The values of k and β were set to 9 and 20, respectively. For SVMCK, we used a weighted summation kernel, which generated the best classification performance compared with other composite kernels [36]. The spatial information was represented by the mean of pixels in a small neighborhood, and the RBF kernel was used with the LibSVM Toolbox [37]. The penalty term C and the RBF kernel width δ were selected by a grid search with a given set $\{2^{-10}, 2^{-9}, \dots, 2^{10}\}$. The spatial window of size in the SVMCK classifier was set to 9×9 for Indian Pines and Salinas data sets, and 5×5 for Urban data set. The embedding dimension was 30 for all the DR methods. All the experiments were performed on a personal computer with i7-4790 central processing unit, 8-G memory, and 64-bit Windows 10 using MATLAB 2013b.

4.3. Two-Dimension Embedding

In this section, we use the Indian Pines data set to analyze the two-dimension embedding of the proposed LGSFA method. In the experiment, we only chose five land cover types from Indian Pines data set including *Corn-mintill*, *Grass-trees*, *Hay-windrowed*, *Wheat* and *Woods*, and they were denoted by 1, 2, 3, 4 and 5. We randomly chose 100 samples per class for training, and the remaining samples were used for two-dimension embedding. Figure 7 shows the data distribution after application of different DR methods.

As shown in Figure 7, the results of PCA, NPE and LPP produced the scattered distribution of points from the same class and the overlapped points from different classes. This phenomenon

may be caused by the unsupervised nature of the methods. The LFDA and MMC methods improved the compactness of points from the same class, while there still existed overlapping points between different classes. The reason is that LFDA and MMC originate from the theory of statistics that cannot effectively reveal the intrinsic manifold structure of data. SDE and MFA can reveal the intrinsic properties of data, but it may not effectively represent the manifold structure of hyperspectral data. This case resulted in some overlapping points between different classes. The proposed LGSFA method achieved better results than other DR methods, for the reason that LGSFA can effectively represent the manifold structure of hyperspectral data.

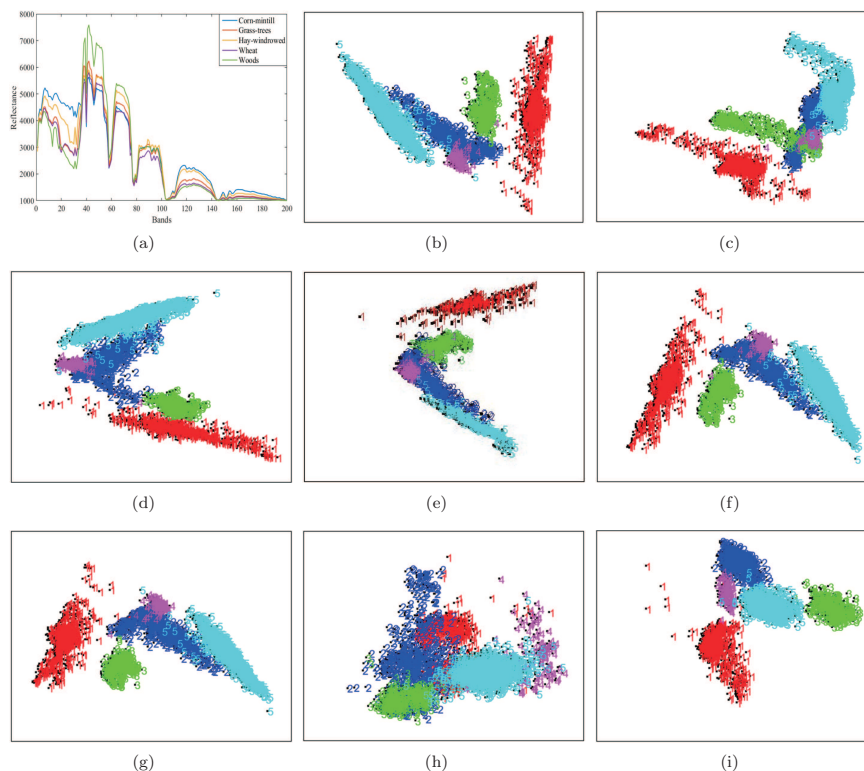


Figure 7. Two-dimension embedding of different DR methods on the Indian Pine data set. (a) Spectral signatures; (b) PCA; (c) NPE; (d) LPP; (e) SDE; (f) LFDA; (g) MMC; (h) MFA; (i) LGSFA.

4.4. Experiments on the Salinas Data Set

To explore the classification accuracy with different numbers of intraclass neighbor points and interclass neighbor points, we randomly selected 60 samples from each class for training and the remaining samples for testing. After DR, the NN classifier was used to discriminate the test samples. Parameters k and β were tuned with a set of $\{3, 5, 7, \dots, 25\}$ and a set of $\{5, 15, 20, \dots, 60\}$, respectively. We repeated the experiment 10 times in each condition. In Figure 8, a curved surface map was depicted to display the average OAs with respect to parameters k and β .

According to Figure 8, with the increase in k , the OAs first increased and then decreased, for a small or large number of intraclass neighbor points cannot effectively represent the intrinsic structure of HSI. When the value of k was lower than 15, the OAs improved and then maintained a stable value with an increasing parameter β . The OAs decreased quickly with a large value of β when the value of k exceeded 15. The reason is that too large values of k and β will result in the appearance of over-learning in the margins of interclass data. As a result, the parameters k and β possesses a small influence on the classification accuracy for the Salinas data set, and we selected k and β to 9 and 20 in the next experiments.

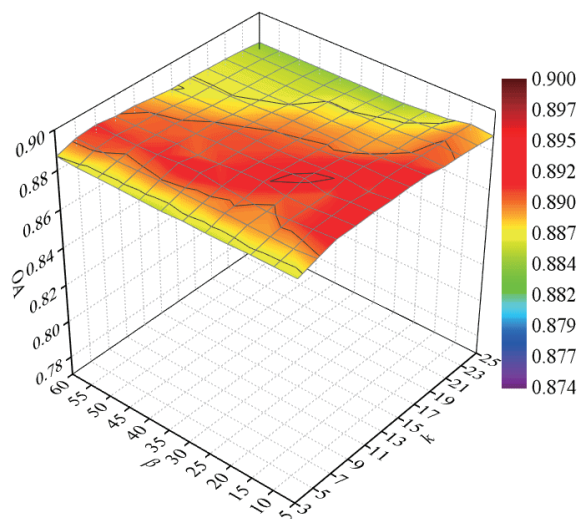


Figure 8. OAs with respect to different numbers of neighbors on the Salinas data set.

To analyze the influence of embedding dimension, we randomly selected 60 training samples for each class in Salinas data set and the NN classifier is used to discriminating the class of test samples. Figure 9 shows the average OAs under different dimensions with 10 times repeated experiment.

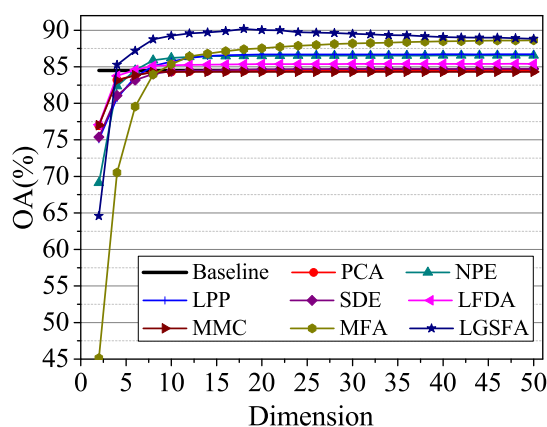


Figure 9. OAs with respect to different dimensions on the Salinas data set.

According to Figure 9, the classification accuracies improved with the increase of the embedding dimension and then reached a peak value. When the embedding dimension exceeded a certain value, the classification results of LGSFA began to decline, which resulted in the Hughes phenomena. NPE, LPP, LFDA, MFA, and LGSFA achieved better results than the baseline, thus indicating that the DR methods can reduce redundant information in HSI data. MFA and LGSFA generated higher OAs than other DR methods, the reason is that MFA and LGSFA can effectively reveal the intrinsic properties of hyperspectral data. In all the methods, LGSFA obtained the best classification accuracy, which indicates that LGSFA can better represent the intrinsic manifold of hyperspectral data that contains many homogenous areas.

To show the performance of LGSFA with different numbers of training samples, we randomly selected n_i samples from each class for training and the remaining samples were used for testing. We adopted NN, SAM and SVMCK for the classification of test samples and repeated the experiment 10 times in each condition. The average OAs with STD and the average KCs are given in Table 1.

Table 1. Classification results with different numbers of training samples on the Salinas data set (OA \pm std (%) (KC)).

Classifier	DR	$n_i = 20$	$n_i = 40$	$n_i = 60$	$n_i = 80$
NN	Baseline	82.4 \pm 0.9 (0.805)	84.1 \pm 0.7 (0.824)	84.7 \pm 0.6 (0.830)	85.1 \pm 0.4 (0.834)
	PCA	82.4 \pm 0.9 (0.805)	84.1 \pm 0.7 (0.824)	84.6 \pm 0.6 (0.830)	85.1 \pm 0.4 (0.834)
	NPE	83.3 \pm 0.9 (0.815)	85.8 \pm 0.6 (0.842)	86.2 \pm 0.5 (0.847)	87.1 \pm 0.4 (0.856)
	LPP	83.6 \pm 0.9 (0.818)	85.6 \pm 0.4 (0.840)	86.1 \pm 0.5 (0.845)	86.7 \pm 0.5 (0.852)
	SDE	82.5 \pm 1.0 (0.806)	83.8 \pm 0.8 (0.821)	84.9 \pm 0.6 (0.832)	85.1 \pm 0.3 (0.835)
	LFDA	83.5 \pm 1.0 (0.818)	84.8 \pm 0.6 (0.832)	85.3 \pm 0.6 (0.836)	85.7 \pm 0.4 (0.841)
	MMC	82.2 \pm 0.9 (0.803)	84.0 \pm 0.7 (0.822)	84.4 \pm 0.5 (0.827)	84.9 \pm 0.4 (0.832)
	MFA	86.6 \pm 1.0 (0.852)	87.9 \pm 0.7 (0.866)	88.2 \pm 0.8 (0.869)	88.4 \pm 0.7 (0.871)
LGSFA	87.3 \pm 1.4 (0.859)	88.8 \pm 0.4 (0.875)	89.4 \pm 0.6 (0.882)	89.8 \pm 0.5 (0.886)	
SAM	Baseline	82.9 \pm 1.0 (0.811)	84.0 \pm 0.9 (0.823)	84.3 \pm 0.7 (0.826)	85.5 \pm 0.5 (0.839)
	PCA	82.8 \pm 1.0 (0.809)	83.8 \pm 0.8 (0.821)	84.1 \pm 0.7 (0.824)	85.3 \pm 0.4 (0.837)
	NPE	84.2 \pm 1.1 (0.825)	85.7 \pm 1.0 (0.841)	86.1 \pm 0.9 (0.846)	87.3 \pm 0.5 (0.859)
	LPP	83.2 \pm 1.1 (0.814)	84.8 \pm 0.9 (0.832)	85.3 \pm 0.9 (0.837)	86.6 \pm 0.4 (0.851)
	SDE	83.4 \pm 1.0 (0.815)	84.0 \pm 1.1 (0.823)	85.2 \pm 0.5 (0.835)	85.2 \pm 0.4 (0.835)
	LFDA	83.7 \pm 1.0 (0.819)	84.4 \pm 0.9 (0.827)	84.5 \pm 0.7 (0.829)	85.6 \pm 0.5 (0.840)
	MMC	82.6 \pm 0.9 (0.807)	83.6 \pm 0.9 (0.818)	84.0 \pm 0.6 (0.823)	85.0 \pm 0.5 (0.833)
	MFA	85.6 \pm 1.1 (0.840)	86.7 \pm 0.7 (0.853)	87.0 \pm 0.6 (0.856)	87.4 \pm 0.9 (0.860)
LGSFA	88.2 \pm 0.6 (0.869)	89.3 \pm 1.0 (0.881)	89.4 \pm 0.8 (0.882)	90.0 \pm 0.4 (0.889)	
SVMCK	Baseline	89.2 \pm 1.2 (0.880)	92.3 \pm 0.9 (0.914)	94.1 \pm 0.5 (0.934)	94.6 \pm 0.4 (0.940)
	PCA	88.1 \pm 1.1 (0.868)	91.6 \pm 0.7 (0.906)	93.7 \pm 0.9 (0.930)	94.3 \pm 0.4 (0.937)
	NPE	87.0 \pm 2.3 (0.855)	91.2 \pm 1.0 (0.902)	93.4 \pm 0.9 (0.927)	95.1 \pm 0.3 (0.945)
	LPP	86.7 \pm 0.7 (0.853)	90.5 \pm 1.1 (0.894)	92.8 \pm 0.7 (0.920)	93.5 \pm 1.0 (0.927)
	SDE	86.2 \pm 2.0 (0.847)	90.4 \pm 1.2 (0.893)	92.6 \pm 0.7 (0.917)	93.4 \pm 0.6 (0.927)
	LFDA	89.6 \pm 1.4 (0.885)	92.1 \pm 1.2 (0.912)	94.3 \pm 0.7 (0.936)	94.8 \pm 0.7 (0.942)
	MMC	89.3 \pm 2.0 (0.881)	91.9 \pm 0.3 (0.910)	93.5 \pm 0.8 (0.928)	94.1 \pm 0.8 (0.934)
	MFA	92.3 \pm 1.5 (0.914)	94.9 \pm 0.9 (0.943)	95.9 \pm 0.3 (0.954)	96.1 \pm 0.4 (0.956)
LGSFA	94.2 \pm 1.2 (0.935)	95.8 \pm 0.8 (0.953)	96.8 \pm 0.5 (0.964)	97.1 \pm 0.4 (0.968)	

According to Table 1, the OAs and the KCs of each method improved as the increase of the number of training samples, because there is more priori information to represent the intrinsic properties of HSI. For different classifiers, each method with SVMCK possessed better classification accuracies than that with other classifiers, because SVMCK utilizes the spatial-spectral information that is beneficial to HSI classification. In all conditions, LGSFA achieved better results compared with MFA, and it also displayed the best accuracies than other DR methods. The reason is that LGSFA utilizes the neighbor points and corresponding intra-class reconstruction points to enhance the intra-class compactness and the inter-class separability. It can effectively represent the intrinsic structure of HSI and obtain better discriminating features for classification.

To explore the performance of LGSFA under different training conditions, we evaluated the classification accuracy of each class with a training set containing about 2% of samples per class. The remaining samples were used for testing. After the low-dimensional features were obtained with different DR methods, the SVMCK classifier was used to classify the test samples. The classification results are shown in Table 2 and corresponding classification maps are given in Figure 10.

As shown in Table 2, the proposed method obtained the strongest classification effect in most classes and achieved the best OA, AA, and KC. LGSFA is clearly effective in revealing the intrinsic manifold structure of HSI, and it extracts better discriminating features for classification. The Baseline method cost more running time than other method for classification, because the hyperspectral data contains a large number of spectral bands to increase the computational cost of classification. LGSFA took more running time to reduce the dimensionality of data for the graph construction. However, LGSFA reduced the total running time for classification compared with the Baseline method, and it improved the classification performances compared with other methods.

Table 2. Classification results of different DR methods with SVMCK on the Salinas data set.

Class	Samples		DR with SVMCK Classifier								
	Training	Test	Baseline	PCA	NPE	LPP	SDE	LFDA	MMC	MFA	LGSFA
1	40	1969	100	100	99.9	99.9	99.7	100	100	99.8	100
2	75	3651	99.5	99.9	100	99.9	99.5	99.8	99.9	100	100
3	40	1936	100	97.1	99.9	99.9	99.9	99.5	96.1	99.8	100
4	28	1366	99.9	99.9	99.8	99.7	99.7	99.9	99.9	99.9	100
5	54	2624	99.6	96.4	98.6	98.9	98.5	98.7	97.1	99.3	99.8
6	79	3880	100	99.9	99.9	100	100.0	100	99.9	100	100
7	72	3507	99.5	99.4	99.7	99.7	99.3	99.5	99.5	100	100
8	225	11,046	90.1	89.1	89.7	90.4	89.5	91.0	88.8	92.4	97.8
9	124	6079	99.7	99.5	100	99.1	100	100	99.1	100	100
10	66	3212	97.9	96.9	98.2	97.2	98.0	98.8	97.1	98.0	100
11	21	1047	97.8	96.0	99.4	99.0	96.6	99.5	97.1	99.1	100
12	39	1888	100	99.9	99.9	100	99.9	99.7	100	100	100
13	18	898	97.9	96.8	97.6	88.8	98.0	99.2	96.1	100	99.2
14	21	1049	99.3	98.7	99.7	97.6	96.6	97.0	98.7	98.7	98.8
15	145	7123	87.2	84.0	91.6	90.3	87.3	88.6	84.3	80.1	97.4
16	36	1771	98.9	99.1	98.7	97.7	97.5	98.9	98.8	99.5	100
	AA		98.0	97.0	98.3	97.4	97.5	98.1	97.0	97.9	99.6
	OA		95.8	94.8	96.4	96.0	95.5	96.3	94.7	95.5	99.2
	KC		0.954	0.942	0.960	0.955	0.950	0.958	0.941	0.950	0.991
	DR time (s)		-	0.059	0.220	0.119	2.228	0.045	0.243	0.179	2.185
	Classification time (s)		1194.0	360.0	353.8	349.5	400.5	359.9	357.2	356.5	353.7

According to Figure 10, LGSFA produced more homogenous areas than other DR methods, especially for the areas labeled as *Grapes*, *Corn*, *Lettuce 4wk*, *Lettuce 7wk* and *Vinyard untrained*.

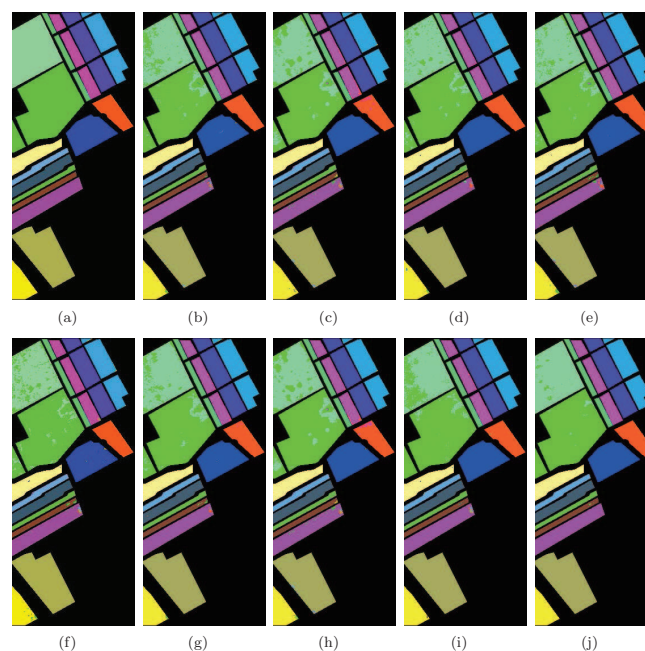


Figure 10. Classification maps of different methods with SVMCK on the Salinas data set. (a) Ground truth; (b) Baseline (95.8%, 0.954); (c) PCA (94.8%, 0.942); (d) NPE (96.4%, 0.960); (e) LPP (96.0%, 0.955); (f) SDE (95.5%, 0.950); (g) LFDA (96.3%, 0.958); (h) MMC (94.7%, 0.941); (i) MFA (95.5%, 0.950); (j) LGSFA (99.2%, 0.991). Note that OA and KC are given in parentheses.

4.5. Experiments on the Indian Pines Data Set

To analyze the performance of LGSFA with different land cover scenes, we utilized the AVIRIS Indian Pines data set for classification. In the experiment, we randomly selected n_i samples from each class for training, and the remaining samples were used for testing. For the classes that are very small, i.e., *Alfalfa*, *Grass/Pasture-mowed*, and *Oats*, the number of training samples was set to $n_i = N_i/2$ if $n_i \geq N_i/2$ where N_i is the number of the i -th class.

To explore the influence of parameters k and β , 60 samples were randomly selected from each class for training, and the remaining samples were used for testing. The NN classifier was adopted to classify the test samples. Figure 11 shows the OAs with respect to parameters k and β .

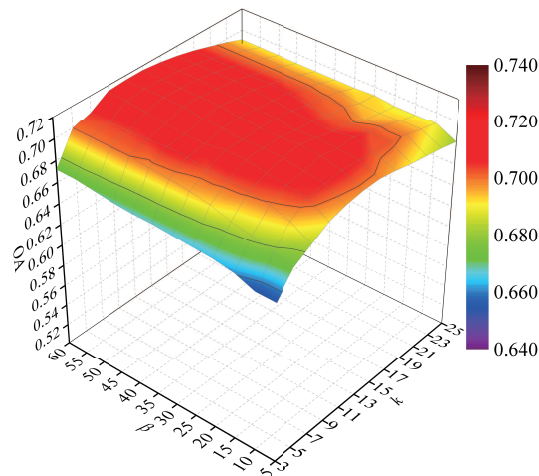


Figure 11. OAs with respect to different numbers of neighbors on the Indian Pines data set.

According to Figure 11, the OAs improved quickly and then declined with the increase of k , because a small value of k cannot obtain enough information to reveal the intraclass structure and a large value of k will cause over-fitting to represent the intrinsic properties of hyperspectral data. The increased β promoted the improvement of OAs, and then OAs reached to a stable peak value. To obtain optimal classification results, we set k and β to 9 and 20 in the experiments.

To analyze the classification accuracy under different embedding dimensions, 60 samples of each class were randomly selected as training set, and the NN classifier was used to classify the remaining samples. The results are shown in Figure 12. In the figure, LGSFA possessed the best classification accuracy, because it can better reveal the intrinsic manifold properties of hyperspectral data.

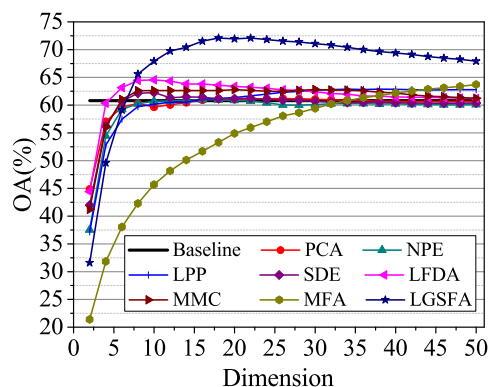


Figure 12. OAs with respect to different dimensions on the Indian Pines data set.

To compare the results of each DR with different classifiers, we randomly selected 20, 40, 60, 80 samples from each class for training, and the remaining samples were used for testing. Each experiment was repeated 10 times and Table 3 shows the average OAs with STD and the average KCs.

Table 3. Classification results with different numbers of training samples on the Indian Pines data set (OA \pm std (%) (KC)).

Classifier	DR	$n_i = 20$	$n_i = 40$	$n_i = 60$	$n_i = 80$
NN	Baseline	55.1 \pm 1.1 (0.498)	58.5 \pm 1.6 (0.535)	61.1 \pm 0.9 (0.562)	62.6 \pm 0.7 (0.578)
	PCA	55.0 \pm 1.2 (0.496)	58.6 \pm 1.7 (0.535)	61.4 \pm 1.1 (0.565)	62.6 \pm 0.9 (0.578)
	NPE	54.5 \pm 1.9 (0.491)	57.3 \pm 1.5 (0.521)	60.1 \pm 1.3 (0.552)	61.5 \pm 1.0 (0.566)
	LPP	55.9 \pm 1.4 (0.506)	60.3 \pm 1.6 (0.554)	62.8 \pm 1.1 (0.581)	64.4 \pm 0.4 (0.598)
	SDE	53.3 \pm 1.2 (0.478)	57.5 \pm 0.9 (0.523)	61.0 \pm 0.9 (0.561)	62.3 \pm 0.8 (0.575)
	LFDA	55.2 \pm 1.4 (0.499)	59.2 \pm 1.6 (0.542)	62.4 \pm 0.8 (0.576)	63.9 \pm 0.9 (0.593)
	MMC	54.8 \pm 1.4 (0.494)	58.7 \pm 1.6 (0.537)	61.9 \pm 1.0 (0.571)	63.4 \pm 0.5 (0.587)
	MFA	51.0 \pm 1.1 (0.454)	57.8 \pm 1.2 (0.528)	60.8 \pm 1.6 (0.559)	61.3 \pm 1.0 (0.564)
LGSFA	58.1 \pm 1.5 (0.530)	67.0 \pm 1.6 (0.628)	72.2 \pm 0.8 (0.686)	73.7 \pm 0.9 (0.702)	
SAM	Baseline	55.5 \pm 1.8 (0.502)	60.4 \pm 1.1 (0.555)	62.2 \pm 1.5 (0.574)	64.6 \pm 1.1 (0.600)
	PCA	55.8 \pm 1.8 (0.505)	60.9 \pm 1.2 (0.560)	62.3 \pm 1.4 (0.575)	64.6 \pm 1.2 (0.600)
	NPE	53.0 \pm 2.0 (0.474)	58.0 \pm 1.6 (0.528)	60.2 \pm 1.1 (0.551)	62.3 \pm 1.3 (0.575)
	LPP	55.8 \pm 1.8 (0.505)	61.0 \pm 1.5 (0.561)	62.8 \pm 1.2 (0.581)	65.0 \pm 1.1 (0.605)
	SDE	54.1 \pm 2.0 (0.487)	58.7 \pm 0.7 (0.537)	60.2 \pm 1.1 (0.552)	62.9 \pm 1.0 (0.581)
	LFDA	54.3 \pm 1.8 (0.489)	60.0 \pm 1.4 (0.550)	62.1 \pm 1.5 (0.573)	64.4 \pm 1.3 (0.598)
	MMC	58.6 \pm 2.0 (0.537)	63.6 \pm 1.0 (0.590)	65.7 \pm 1.2 (0.613)	67.3 \pm 1.1 (0.630)
	MFA	48.8 \pm 2.2 (0.432)	57.0 \pm 2.2 (0.519)	58.4 \pm 1.9 (0.533)	58.9 \pm 1.9 (0.538)
LGSFA	57.6 \pm 1.5 (0.525)	67.7 \pm 1.1 (0.635)	71.0 \pm 1.4 (0.672)	73.3 \pm 0.8 (0.697)	
SVMCK	Baseline	77.1 \pm 1.7 (0.743)	84.6 \pm 1.1 (0.825)	88.2 \pm 1.0 (0.865)	89.8 \pm 1.0 (0.883)
	PCA	73.1 \pm 2.5 (0.697)	81.8 \pm 1.5 (0.793)	85.4 \pm 1.0 (0.833)	87.9 \pm 1.5 (0.862)
	NPE	73.1 \pm 2.9 (0.697)	81.6 \pm 2.2 (0.791)	85.4 \pm 1.2 (0.834)	87.7 \pm 1.7 (0.859)
	LPP	70.7 \pm 1.6 (0.672)	81.1 \pm 2.1 (0.786)	84.6 \pm 1.2 (0.825)	87.4 \pm 1.4 (0.856)
	SDE	73.4 \pm 1.8 (0.700)	81.8 \pm 1.4 (0.794)	86.7 \pm 1.1 (0.849)	89.2 \pm 0.8 (0.876)
	LFDA	74.4 \pm 2.4 (0.712)	82.4 \pm 1.4 (0.800)	86.2 \pm 1.1 (0.843)	88.1 \pm 1.0 (0.863)
	MMC	74.7 \pm 2.3 (0.715)	81.3 \pm 1.3 (0.788)	84.4 \pm 2.7 (0.822)	84.7 \pm 2.4 (0.825)
	MFA	71.3 \pm 2.0 (0.678)	82.2 \pm 1.4 (0.798)	88.1 \pm 1.0 (0.865)	91.4 \pm 1.1 (0.901)
LGSFA	85.3 \pm 3.1 (0.833)	91.2 \pm 1.7 (0.900)	95.0 \pm 1.0 (0.942)	96.2 \pm 1.0 (0.957)	

As shown in Table 3, the classification accuracy of each DR method improved with the increasing number of training samples. The results of different DR methods with NN or SAM were unsatisfactory for the restricted discriminating power of NN or SAM. However, LGSFA with NN or SAM showed better results than other DR methods in most cases. The reason is that LGSFA effectively reveals the intrinsic manifold structure of hyperspectral data and obtains better discriminating features for classification. In addition, each DR method with SVMCK possessed better results than that with NN or SAM, and LGSFA with SVMCK achieved the best classification performances in all conditions.

To show the classification power of different methods for each class, 10% samples per class were selected for training, and other samples were used for testing. For very small classes, we take a minimum of ten training samples per class. The results of different methods with SVMCK are shown in Table 4.

Table 4. Classification results of different DR methods with SVMCK on the Indian Pines data set.

Class	Samples		DR with SVMCK Classifier								
	Training	Test	Baseline	PCA	NPE	LPP	SDE	LFDA	MMC	MFA	LGSFA
1	10	36	83.3	91.7	75.0	72.2	94.4	61.1	44.4	55.6	66.7
2	143	1285	87.9	87.2	86.2	85.3	85.7	84.4	85.2	91.3	96.5
3	83	747	93.3	89.6	85.4	85.0	87.1	91.6	84.5	94.8	96.3
4	24	213	93.9	79.3	93.0	90.1	85.0	73.7	84.5	74.2	97.7
5	48	435	95.2	95.9	95.6	95.9	97.2	96.3	96.8	95.9	97.5
6	73	657	99.7	100	100	99	99.7	100	97.0	99.1	100
7	10	18	94.4	94.4	100	100	94.4	94.4	88.9	100	100
8	48	430	99.8	97.4	99	100	99.5	99.1	99.8	100	100
9	10	10	100	100	100	100	100	100	100	90.0	100
10	97	875	85.7	86.6	86.5	83.3	82.1	86.3	80.2	86.1	96.1
11	246	2209	95.2	91.1	92.3	91.4	91.8	90.6	91.1	94.7	99.3
12	59	534	92.3	84.8	83.9	82.0	83.3	81.3	85.0	88.4	97.8
13	21	184	98.9	98.9	97.8	98.9	97.8	99.5	98.9	98.9	97.3
14	127	1138	97.8	98.1	98.6	98.0	97.7	98.2	96.9	97.7	99.9
15	39	347	96.3	96.5	97.4	96.8	94.2	94.2	88.8	93.7	100
16	10	83	97.6	96	100	89.2	90.4	88	92.8	98.8	92.8
	AA		94.5	93.0	93.2	91.7	92.5	89.9	88.4	91.2	96.1
	OA		93.9	91.8	92.0	91.0	91.1	90.9	89.9	93.5	98.1
	KC		0.930	0.907	0.908	0.897	0.899	0.896	0.885	0.926	0.978
	DR time (s)		-	0.012	0.163	0.051	2.056	0.012	0.528	0.086	1.827
	Classification time (s)		40.7	12.7	12.7	12.6	14.4	12.7	12.7	12.9	12.8

According to Table 4, LGSFA achieved the best classification accuracy for most classes, as well as the best AA, OA, and KC among all the methods, because it can effectively represent the hidden manifold structure of hyperspectral data. The corresponding classification maps are displayed in Figure 13, which clearly indicates that the proposed method produces a smoother classification map compared with other methods in many areas.

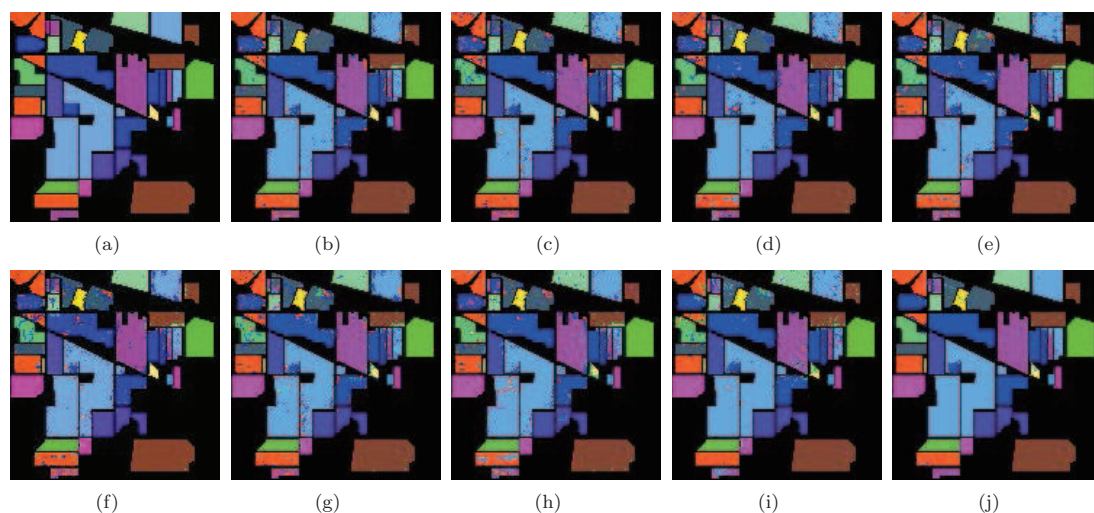


Figure 13. Classification maps of different methods with SVMCK on the Indian Pines data set. (a) Ground truth; (b) Baseline (93.9%, 0.930); (c) PCA (91.8%, 0.907); (d) NPE (92.0%, 0.908); (e) LPP (91.0%, 0.897); (f) SDE (91.1%, 0.899); (g) LFDA (90.9%, 0.896); (h) MMC (89.9%, 0.885); (i) MFA (93.5%, 0.926); (j) LGSFA (98.1%, 0.978). Note that OA and KC are given in parentheses.

4.6. Experiments on the Urban Data Set

To further explore the effectiveness of LGSFA with different sensors, we selected the Urban data set captured by HYDICE sensor. We randomly selected 60 samples as training set to analyze OAs with respect to parameters k and β . Figure 14 displays the classification results.

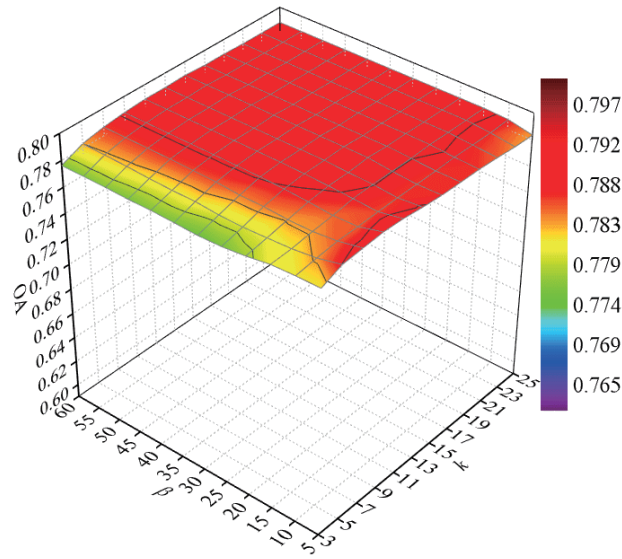


Figure 14. OAs with respect to different numbers of neighbors on the Urban data set.

According to Figure 14, with the increase of intra-class neighbor k , the classification accuracies improved and then reached a stable value. The main reason is that the larger k will provide more effective information to reveal the intrinsic properties. The OAs also increased when the value of β enlarged, because there is more information to represent the margin of different classes with a larger β . To obtain better results, we set k and β to 9 and 20 in the experiments.

To explore the relationship between the classification accuracy and the embedding dimension, we randomly selected 60 samples from each class for training, and the remaining samples were used for testing. Figure 15 shows the OAs with different dimensions using the NN classifier. As shown in Figure 15, the OAs improved with the increase of the embedding dimension and then kept a balanced value. Compared with other methods, the proposed LGSFA method provided the best classification results.

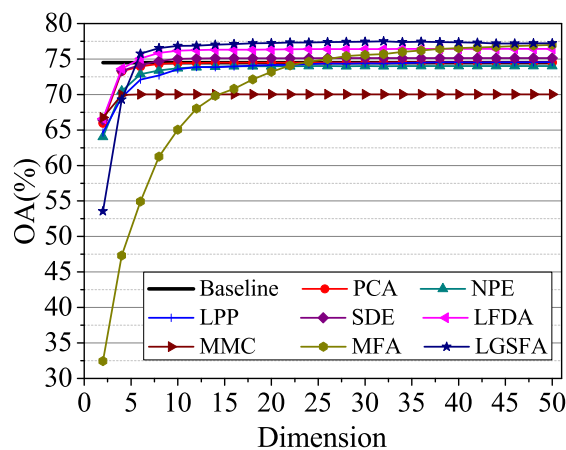


Figure 15. OAs with respect to different dimensions on the Urban data set.

To compare the proposed method with other methods under different numbers of training samples, we randomly selected 20, 40, 60, 80 samples as training set, and the remaining samples were applied for testing. After DR with different methods, we adopted the NN, SAM, and SVMCK classifiers to discriminate the class of test samples with 10 times repeat experiments. The average OAs with std and KC (in parentheses) are given in Table 5.

Table 5. Classification results with different numbers of training samples on the Urban data set (OA \pm std (%) (KC)).

Classifier	DR	$n_i = 20$	$n_i = 40$	$n_i = 60$	$n_i = 80$
NN	Baseline	72.3 \pm 3.5 (0.577)	72.6 \pm 2.1 (0.585)	75.1 \pm 1.0 (0.614)	76.0 \pm 1.3 (0.627)
	PCA	72.3 \pm 3.5 (0.577)	72.6 \pm 2.1 (0.585)	75.1 \pm 1.0 (0.615)	76.0 \pm 1.3 (0.627)
	NPE	72.2 \pm 3.5 (0.575)	71.7 \pm 2.1 (0.574)	74.8 \pm 1.0 (0.613)	75.7 \pm 1.4 (0.627)
	LPP	71.9 \pm 3.9 (0.571)	72.1 \pm 2.2 (0.579)	74.3 \pm 0.9 (0.604)	75.3 \pm 1.6 (0.617)
	SDE	72.6 \pm 3.6 (0.581)	72.9 \pm 2.2 (0.589)	75.3 \pm 1.1 (0.617)	76.1 \pm 1.3 (0.628)
	LFDA	75.3 \pm 2.7 (0.620)	75.6 \pm 1.6 (0.625)	76.9 \pm 1.0 (0.641)	77.5 \pm 1.3 (0.651)
	MMC	71.4 \pm 4.0 (0.565)	70.9 \pm 2.0 (0.562)	72.3 \pm 1.7 (0.576)	73.4 \pm 1.7 (0.590)
	MFA	74.6 \pm 3.3 (0.609)	74.1 \pm 2.5 (0.605)	76.2 \pm 1.0 (0.630)	77.1 \pm 1.1 (0.644)
	LGSFA	75.3 \pm 3.1 (0.622)	76.7 \pm 1.7 (0.641)	78.1 \pm 1.5 (0.658)	78.7 \pm 0.5 (0.669)
SAM	Baseline	70.3 \pm 2.8 (0.552)	72.0 \pm 3.4 (0.572)	72.6 \pm 2.1 (0.584)	73.5 \pm 1.3 (0.595)
	PCA	70.2 \pm 2.7 (0.550)	71.9 \pm 3.4 (0.570)	72.6 \pm 2.1 (0.583)	73.4 \pm 1.3 (0.594)
	NPE	72.0 \pm 2.6 (0.574)	74.7 \pm 3.6 (0.609)	75.4 \pm 2.4 (0.620)	75.8 \pm 1.4 (0.628)
	LPP	72.0 \pm 2.8 (0.576)	74.0 \pm 4.0 (0.600)	74.5 \pm 1.9 (0.610)	75.4 \pm 1.7 (0.623)
	SDE	70.3 \pm 2.7 (0.551)	71.9 \pm 3.4 (0.571)	72.6 \pm 2.1 (0.584)	73.5 \pm 1.3 (0.595)
	LFDA	74.6 \pm 2.0 (0.611)	74.9 \pm 3.3 (0.614)	75.5 \pm 1.5 (0.623)	75.8 \pm 1.2 (0.628)
	MMC	69.3 \pm 2.9 (0.537)	69.3 \pm 4.5 (0.535)	70.5 \pm 2.5 (0.552)	70.6 \pm 2.4 (0.554)
	MFA	72.4 \pm 2.1 (0.580)	73.6 \pm 3.6 (0.595)	74.3 \pm 1.8 (0.607)	75.0 \pm 1.3 (0.617)
	LGSFA	76.7 \pm 1.9 (0.641)	77.8 \pm 2.5 (0.655)	78.1 \pm 1.1 (0.659)	78.9 \pm 1.2 (0.671)
SVMCK	Baseline	75.5 \pm 3.1 (0.626)	78.2 \pm 0.9 (0.661)	79.8 \pm 0.9 (0.684)	81.4 \pm 1.4 (0.707)
	PCA	72.0 \pm 3.7 (0.581)	74.5 \pm 2.5 (0.610)	75.6 \pm 1.7 (0.627)	78.7 \pm 1.4 (0.666)
	NPE	73.1 \pm 3.0 (0.592)	77.3 \pm 1.7 (0.648)	78.0 \pm 1.5 (0.660)	80.4 \pm 1.8 (0.691)
	LPP	72.5 \pm 3.3 (0.586)	77.3 \pm 3.5 (0.652)	79.6 \pm 1.5 (0.683)	81.2 \pm 2.0 (0.703)
	SDE	73.0 \pm 3.5 (0.591)	75.6 \pm 3.2 (0.624)	77.8 \pm 3.0 (0.656)	80.2 \pm 2.7 (0.689)
	LFDA	75.2 \pm 2.6 (0.622)	77.9 \pm 1.6 (0.658)	79.1 \pm 1.9 (0.675)	80.5 \pm 2.1 (0.693)
	MMC	70.5 \pm 4.8 (0.558)	75.7 \pm 3.7 (0.628)	75.2 \pm 3.8 (0.621)	80.4 \pm 3.1 (0.691)
	MFA	72.6 \pm 2.8 (0.587)	76.7 \pm 1.7 (0.642)	78.3 \pm 1.8 (0.664)	80.2 \pm 1.4 (0.688)
	LGSFA	79.8 \pm 2.8 (0.688)	82.9 \pm 1.4 (0.731)	84.0 \pm 1.5 (0.748)	85.5 \pm 1.0 (0.767)

In Table 5, the results improved with the increase of the number of training samples in most conditions. Compared with other DR methods, LGSFA achieved the best OAs and KCs under different classifiers. Each DR method with SVMCK presented better classification results than other classifiers, and LGSFA with SVMCK possessed the best results in all cases. The experiment indicates that the proposed LGSFA method can effectively reveal the intrinsic manifold properties of hyperspectral data containing many homogenous areas.

To show the classification results of each class, we randomly selected 2% samples as training set, and the other samples were used for testing. The classification maps are shown in Figure 16.

As shown in Figure 16, the LGSFA method displayed more similar compared with the ground truth. The corresponding numerical results are shown in Table 6. In the table, the proposed method obtained better classification accuracy in most classes and achieved the best AA, OA, and KC among all the methods, which indicates that LGSFA is more beneficial to represent the hidden information of hyperspectral data.

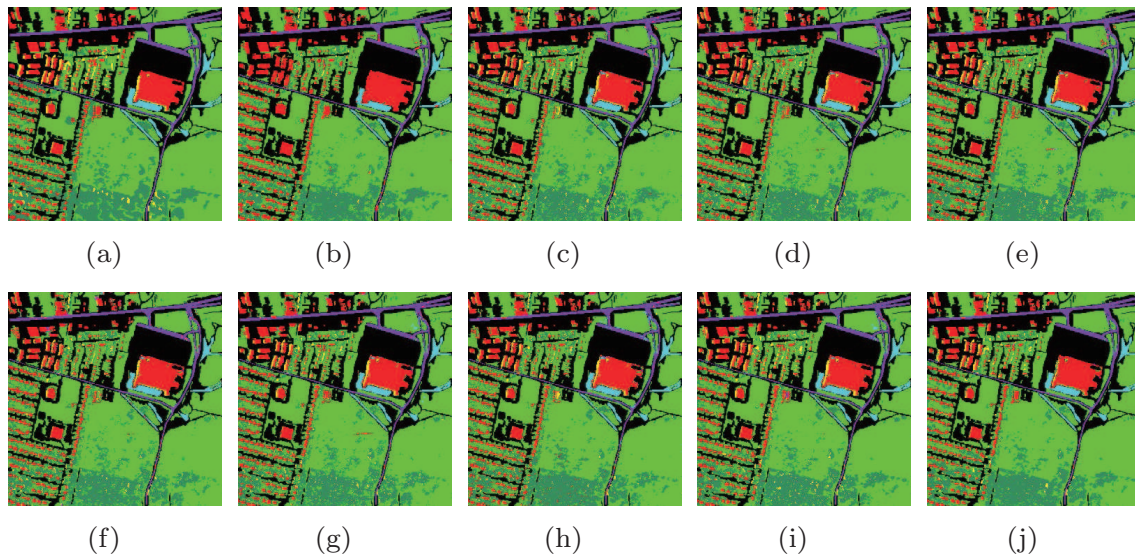


Figure 16. Classification maps of different methods with SVMCK on the Urban data set. (a) Ground truth; (b) Baseline (86.0%, 0.765); (c) PCA (85.1%, 0.750); (d) NPE (86.5%, 0.773); (e) LPP (87.0%, 0.781); (f) SDE (85.7%, 0.758); (g) LFDA (85.6%, 0.759); (h) MMC (86.1%, 0.765); (i) MFA (87.5%, 0.788); (j) LGSFA (88.8%, 0.809). Note that OA and KC are given in parentheses.

Table 6. Classification results of different DR methods with SVMCK on the Urban data set.

Class	Samples		DR with SVMCK Classifier								
	Training	Test	Baseline	PCA	NPE	LPP	SDE	LFDA	MMC	MFA	LGSFA
1	185	9045	90.6	86.3	87.1	87.2	85.3	87.9	84.1	87.3	88.6
2	29	1444	14.1	54.1	53.0	51.3	51.2	55.9	46.6	60.4	53.6
3	65	3165	94.7	95.6	96.3	95.7	94.5	95.4	91.9	94.8	97.2
4	30	1466	90.0	93.0	92.2	91.5	88.3	93.4	92.2	91.5	95.0
5	829	40,613	90.7	88.9	90.8	91.4	90.4	89.5	91.6	92.3	93.5
6	268	13,108	73.7	72.2	73.5	74.8	72.7	72.3	72.8	73.4	75.5
	AA		75.6	81.7	82.2	82.0	80.4	82.4	79.9	83.3	83.9
	OA		86.0	85.1	86.5	87.0	85.7	85.6	86.1	87.5	88.8
	KC		0.765	0.750	0.773	0.781	0.758	0.759	0.765	0.788	0.809
	DR time (s)		-	0.040	0.308	0.085	3.390	3.390	0.449	2.185	3.837
	Classification time (s)		1637.7	438.0	421.3	445.5	438.5	434.4	438.1	435.8	430.4

4.7. Discussion

The following interesting points are revealed in the experiments on the Salinas, Indian Pines, and Urban HSI data sets.

- The proposed LGSFA method consistently outperforms Baseline, PCA, NPE, LPP, SDE, LFDA, MMC and MFA in most conditions on three real HSI data sets. The reason for this appearance is that LGSFA utilizes neighbor points and corresponding intraclass reconstruction points to construct the intrinsic and penalty graphs, while MFA just uses the neighbor points to construct the intrinsic and penalty graphs. That is to say, our proposed method can effectively compact the intraclass data and separate the interclass data, and it can capture more intrinsic information hidden in HSI data sets than other methods.
- It is clear that LGSFA produces a smoother classification map and achieves better accuracy compared with other methods in most classes. LGSFA effectively reveal the intrinsic manifold structures of hyperspectral data. Thus, this method obtains good discriminating features

and improves the classification performance of the NN, SAM and SVMCK classifiers for hyperspectral data.

- In the experiments, it is noticeable that the SVMCK classifier always performs better than NN and SAM. The reason is that SVMCK applies the spatial and spectral information while NN and SAM only use the spectral information for HSI classification.
- With the running time of different DR methods, the computational complexity of LGSFA depends on the number of bands, training samples, and neighbor points. The proposed method costs more time than other DR algorithms. The reason is that LGSFA needs much running time to construct the intrinsic graph and the penalty graph.
- In the experiments of two-dimension embedding, LGSFA achieves better distribution for data points than other DR methods. The results show that LGSFA can improve the intra-manifold compactness and the inter-manifold separability to enhance the diversity of data from different classes.

5. Conclusions

In this paper, we proposed a new method called LGSFA for DR of HSI. The proposed LGSFA method constructs an intrinsic graph and a penalty graph using the intraclass neighbor points, the interclass neighbor points and their corresponding reconstruction points. It utilizes neighbor points and corresponding intraclass reconstruction points to improve the intraclass compactness and the interclass separability. As a result, LGSFA effectively reveals the intrinsic manifold structure of hyperspectral data to extract effective discriminating features for classification. Experiments on the Salinas, Indian Pines, and Urban data sets show that the proposed method achieves better results than other state-of-the-art methods in most conditions and LGSFA with SVMCK obtains the best classification accuracy. Our future work will focus on how to improve the computational efficiency and the misclassified pixels.

Acknowledgments: The authors would like to thank David A Landgrebe at Purdue University for providing the AVIRIS Indian Pines data set, and Lee F Johnson and J Anthony Gualtieri for providing the AVIRIS Salinas data set. The authors would like to thank the anonymous reviewers and associate editor for their valuable comments and suggestions to improve the quality of the paper. This work was supported in part by the National Science Foundation of China under Grant 41371338, the Visiting Scholar Foundation of Key Laboratory of Optoelectronic Technology and Systems (Chongqing University), Ministry of Education, and China Postdoctoral Program for Innovative Talent under Grant BX201700182.

Author Contributions: Fulin Luo was primarily responsible for mathematical modeling and experimental design. Hong Huang contributed to the experimental analysis and revised the paper. Yule Duan completed the comparison with other methods. Jiamin Liu provided important suggestions for improving the paper. Yinghua Liao provided some suggestions in the revised manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liang, H.M.; Li, Q. Hyperspectral imagery classification using sparse representations of convolutional neural network features. *Remote Sens.* **2016**, *8*, 919.
2. He, W.; Zhang, H.Y.; Zhang, L.P.; Philips, W.; Liao, W.Z. Weighted sparse graph based dimensionality reduction for hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 686–690.
3. Zhong, Y.F.; Wang, X.Y.; Zhao, L.; Feng, R.Y.; Zhang, L.P.; Xu, Y.Y. Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery. *J. Photogramm. Remote Sens.* **2016**, *119*, 49–63.
4. Zhou, Y.C.; Peng, J.T.; Chen, C.L.P. Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1082–1095.
5. Feng, F.B.; Li, W.; Du, Q.; Zhang, B. Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity. *Remote Sens.* **2017**, *9*, 323.

6. Sun, W.W.; Halevy, A.; Benedetto, J.J.; Czaja, W.; Liu, C.; Wu, H.B.; Shi, B.Q.; Li, W.Y. Ulisomap based nonlinear dimensionality reduction for hyperspectral imagery classification. *J. Photogramm. Remote Sens.* **2014**, *89*, 25–36.
7. Rathore, M.M.U.; Paul, A.; Ahmad, A.; Chen, B.W.; Huang, B.; Ji, W. Real-time big data analytical architecture for remote sensing application. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4610–4621.
8. Tong, Q.X.; Xue, Y.Q.; Zhang, L.F. progress in hyperspectral remote sensing science and technology in China over the past three decades. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 70–91.
9. Huang, H.; Luo, F.L.; Liu, J.M.; Yang, Y.Q. Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding. *J. Photogramm. Remote Sens.* **2015**, *106*, 42–54.
10. Zhang, L.P.; Zhong, Y.F.; Huang, B.; Gong, J.Y.; Li, P.X. Dimensionality reduction based on clonal selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 4172–4186.
11. Cheng, G.L.; Zhu, F.Y.; Xiang, S.M.; Wang, Y.; Pan, C.H. Semisupervised hyperspectral image classification via discriminant analysis and robust regression. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 595–608.
12. Shi, Q.; Zhang, L.P.; Du, B. Semisupervised discriminative locally enhanced alignment for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4800–4815.
13. Huang, H.; Yang, M. Dimensionality reduction of hyperspectral images with sparse discriminant embedding. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5160–5169.
14. Yang, S.Y.; Jin, P.L.; Li, B.; Yang, L.X.; Xu, W.H.; Jiao, L.C. Semisupervised dual-geometric subspace projection for dimensionality reduction of hyperspectral image data. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3587–3593.
15. Zhang, L.F.; Zhang, L.P.; Tao, D.C.; Huang, X.; Du, B. Compression of hyperspectral remote sensing images by tensor approach. *Neurocomputing* **2015**, *147*, 358–363.
16. Cheng, X.M.; Chen, Y.R.; Tao, Y.; Wang, C.Y.; Kim, M.S.; Lefcourt, A.M. A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. *Trans. ASAE* **2004**, *47*, 1313–1320.
17. Guan, L.X.; Xie, W.X.; Pei, J.H. Segmented minimum noise fraction transformation for efficient feature extraction of hyperspectral images. *Pattern Recognit.* **2015**, *48*, 3216–3226.
18. Cai, D.; He, X.; Han, J. Semi-supervised discriminant analysis. In Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–7.
19. Li, H.F.; Jiang, T.; Zhang, K.S. Efficient and robust feature extraction by maximum margin criterion. *IEEE Trans. Neural Netw.* **2006**, *17*, 157–165.
20. Sugiyama, M. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.
21. Shao, Z.; Zhang, L. Sparse dimensionality reduction of hyperspectral image based on semi-supervised local Fisher discriminant analysis. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *31*, 122–129.
22. Bachmann, C.M.; Ainsworth, T.L.; Fusina, R.A. Exploiting manifold geometry in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 441–454.
23. Yang, H.L.; Crawford, M.M. Domain adaptation with preservation of manifold geometry for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 543–555.
24. Ma, L.; Zhang, X.F.; Yu, X.; Luo, D.P. Spatial regularized local manifold learning for classification of hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 609–624.
25. Tang, Y.Y.; Yuan, H.L.; Li, L.Q. Manifold-Based Sparse Representation for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7606–7618.
26. Zhang, L.F.; Zhang, Q.; Zhang, L.P.; Tao, D.C.; Huang, X.; Du, B. Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding. *Pattern Recognit.* **2015**, *48*, 3102–3112.
27. Tenenbaum, J.B.; de Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323.
28. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396.
29. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326.

30. Wang, Z.Y.; He, B.B. Locality perserving projections algorithm for hyperspectral image dimensionality reduction. In Proceedings of the 19th International Conference on Geoinformatics, Shanghai, China, 24–26 June 2011; pp. 1–4.
31. He, X.F.; Cai, D.; Yan, S.C.; Zhang, H.J. Neighborhood preserving embedding. In Proceedings of the 10th International Conference on Computer Vision, Beijing, China, 17–21 October 2005; pp. 1208–1213.
32. Yan, S.C.; Xu, D.; Zhang, B.Y.; Zhang, H.J.; Yang, Q.; Lin, S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 40–51.
33. Chen, Y.S.; Zhao, X.; Jia, X.P. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2381–2392.
34. Feng, Z.X.; Yang, S.Y.; Wang, S.G.; Jiao, L.C. Discriminative spectral-spatial margin-based semisupervised dimensionality reduction of hyperspectral data. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 224–228.
35. Luo, F.L.; Huang, H.; Ma, Z.Z.; Liu, J.M. Semisupervised sparse manifold discriminative analysis for feature extraction of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6197–6211.
36. Camps-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Vila-Frances, J.; Calpe-Maravilla, J. Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 93–97.
37. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. 2001. Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (accessed on 18 May 2017).



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).