

Research Article

Local and Global Geometric Structure Preserving and Application to Hyperspectral Image Classification

Huiwu Luo,¹ Yuan Yan Tang,¹ Chunli Li,¹ and Lina Yang^{1,2}

¹Department of Computer and Information Science, University of Macau, Avenida Padre Tomas Pereira, Taipa 1356, Macau

²Department of Mathematics and Computer Science, Guangxi Normal University of Nationalities, Chongzuo 532200, China

Correspondence should be addressed to Yuan Yan Tang; yytang@umac.mo

Received 15 December 2014; Revised 16 March 2015; Accepted 16 March 2015

Academic Editor: Hakim Naceur

Copyright © 2015 Huiwu Luo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Locality Preserving Projection (LPP) has shown great efficiency in feature extraction. LPP captures the locality by the K -nearest neighborhoods. However, recent progress has demonstrated the importance of global geometric structure in discriminant analysis. Thus, both the locality and global geometric structure are critical for dimension reduction. In this paper, a novel linear supervised dimensionality reduction algorithm, called *Locality and Global Geometric Structure Preserving* (LGGSP) projection, is proposed for dimension reduction. LGGSP encodes not only the local structure information into the optimal objective functions, but also the global structure information. To be specific, two adjacent matrices, that is, similarity matrix and variance matrix, are constructed to detect the local intrinsic structure. Besides, a margin matrix is defined to capture the global structure of different classes. Finally, the three matrices are integrated into the framework of graph embedding for optimal solution. The proposed scheme is illustrated using both simulated data points and the well-known Indian Pines hyperspectral data set, and the experimental results are promising.

1. Introduction

Hyperspectral image (HSI) processing, as a typical application of high dimensional data analysis, has witnessed great interest among worldwide researchers [1]. The acquisition of hyperspectral image is usually concerned with analysis, measurement, understanding, and interpretation from a given scenario at different airline distance by the satellite [2]. Different HSI data poses different level of challenge to the task of data analysis. However, a common issue of HSI data is the high dimensional feature space within relative small sample size [3], which is also known as the “Hughes phenomenon.” To increase efficiency, dimensionality of HSI data must be reduced before further processing. Dimension reduction plays a significant role in HSI community [4].

Popular dimension reduction literatures can be roughly categorized into two approaches. The first one is nonlinear approaches, for example, Isomap embedding (Isomap) [5], local tangent space alignment (LTSA) [6], Laplacian eigenmaps (LE) [7], local linear embedding (LLE) [8], and so forth. The other one is linear approaches, for example, principal component analysis (PCA), linear discriminant analysis

(LDA), random projection (RP) [9], Locality Preserving Projection (LPP) [10], and so forth. Furthermore, some of these methods are supervised approaches, and some of the others are unsupervised approaches.

Recently, some articles [11] pointed out that high dimensional data may rely on a submanifold that reflects the inherent geodesic structure. Under this circumstance, both PCA and LDA may fail to find the hidden manifold, whereas the nonlinear literatures, such as LLE and Isomap, have been proposed and developed to tackle this difficulty. However, the mapping of LLE and Isomap is implicit and there is no exact computational expression of new data points. That is, the projected data points of LLE and Isomap are defined on the training data points, and both methods can not directly embed a new data point in the projected space. Moreover, these methods are computationally expensive. This drawback makes these algorithms hard to be further developed and limits their massive application in various areas, especially in the hyperspectral image analysis community.

To address this issue, He and Niyogi [10] proposed the Locality Preserving Projection (LPP), which is a linear approximation of intrinsic manifold, to reduce the high

dimensional facial feature vectors into a low dimensional subspace. The neighborhood relationship in LPP is preserved in the projected submanifold. However, LPP is an unsupervised algorithm. The discriminant information is ignored. Wong and Zhao [11] proposed a supervised version LPP, where discriminant information of different classes is adopted to improve the classification performance. Vasuhi and Vaidehi [12] found that the basis of LPP in the projected space is not orthogonal. They applied an orthogonal basis to facial classification and found that the classified accuracy of orthogonal basis was better than conventional LPP. A common theme of many discriminant analysis based methods is this: by minimizing neighborhood distance from the same class, the locality based approaches utilize discriminant information to maximize the distance among data points from different classes, simultaneously preserving the intraclass compactness. The distance of adjacent data points represents the local geometrical structure of the same class, yet distance from different data points indicates the global geometrical structure of different classes. By doing so, the structure of data points in the projected space is expected to be similar to the original space.

Despite this, some articles, such as *Local and Global Structures Preserving Projection* (LGSP) [13] and *Joint Global and Local Structure Discriminant Analysis* (JGLDA) [14], reported that besides locality, the global structure is also important. The locality can be generally captured by a Laplacian matrix that comes from neighborhood relationship, that is, adjacent graph. Moreover, the global geometric structure can also be captured by a relationship matrix, for example, penalty matrix [15], K -farthest neighborhood adjacent matrix [16], or K -nearest neighborhood adjacent matrix [17]. However, these methods only capture the similarity structure of data points to learn the intrinsic geometric structure (local structure). They ignore the distribution of data points, and the structure of data points in the embedded space is destroyed. Consequently, it leads to incorrect description of data structure. In most instances, a single locality is insufficient for describing the intrinsic geometry of data points. Thus, it will be more discriminative if both local and global statistic properties are integrated to describe the geometry of data points.

Motivated by these factors, in this paper, we proposed a novel approach, that is, *Locality and Global Geometric Structure Preserving* (LGGSP) projection, that makes use of not only the local structure, but also the global structure of data points, to reduce the dimensionality of feature vectors. Specifically, we focus on the global distribution of data points, where the local structure is characterized by the similarity and the diversity of samples from the same class, respectively. Besides, the global structure is characterized by the margin of different samples. To achieve the goal of discovering both local and global structures hidden in data points, we first define three optimization functions. Then we solve them in the framework of graph embedding to make the LGGSP algorithm supervised. And finally, a linear transformation is found by utilizing the principle of discriminant analysis.

The rest of this paper is organized as follows. Section 2 provides a brief analysis of basic discriminant techniques. Proposed LGGSP is presented in Section 3. Results of

synthetic data sets and real hyperspectral image data are presented in Section 4. Finally, concluding remarks and discussion are drawn in Section 5.

2. Related Works

Before further discussion, some of the notations that will be used throughout this paper are listed in Notations section.

A brief review of discriminant analysis techniques, for example, Locality Preserving Projection (LPP) [10] and discriminant analysis [13, 14], are provided in this section. To facilitate the following discussion, we start with a supervised learning problem. Suppose that the n -dimensional data set $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$, $\mathbf{x}_i \in \mathbb{R}^n$ is distributed on a d -dimensional submanifold ($d < n$). And this data set \mathbf{X} belongs to C classes with class labels $\{l_i\}_{i=1}^N$, respectively. Let N_c be the samples number of class c ; then $N = \sum_{c=1}^C N_c$. We are expected to find a transformation $T : \mathbf{Y} = T^T \mathbf{X}$, $T \in \mathbb{R}^{n \times d}$ that projected the n -dimensional data points $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ to d -dimensional data points $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$ with the goal of preserving the data structure without losing any information needed. The notation \top represents the transpose of a matrix or a vector. Thereby, the problem at hand is how to evaluate the data model and formulate the objective transformation T .

2.1. Locality Preserving Projection. LDA aims to learn a global structure that separates samples efficiently. Nevertheless, for most real world applications, the local structure of neighborhood is also important. Locality Preserving Projection (LPP) is a graph based subspace learning algorithm, where the neighborhood structure will be preserved in the projected space. To achieve this goal, a weighted graph $\mathbb{G} = (\mathbb{V}, \mathbb{E}, W)$ is constructed, where \mathbb{V} represents the vertex set, \mathbb{E} denotes the edges of connected data points, and W is the similarity weight that characterizes the likelihood of pairwise data points.

For a new coming point \mathbf{x}_i , LPP defines a transformation in the mapping space; that is, $\mathbf{y}_i = T^T \mathbf{x}_i$. Then the criterion function of LPP becomes

$$\min_{\mathbf{Y}} \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 W_{i,j}, \quad (1)$$

where W is the similarity matrix of two data points. If two neighboring data points \mathbf{x}_i and \mathbf{x}_j are mapped far away, then $W_{i,j}$ incurs a heavy penalty. This property ensures that adjacent data points stay as close as possible in the embedded space.

By simple algebra formulation, it can be deduced from (1) that

$$\begin{aligned} & \frac{1}{2} \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 S_{i,j} \\ & = T^T \mathbf{X} \mathbf{L} \mathbf{X}^T T, \end{aligned} \quad (2)$$

where D is a diagonal matrix whose entries are column (or row) sums of S ; that is, $D_{i,i} = \sum_j S_{i,j}$. $L = D - S$ represents Laplacian matrix, which is the discrete approximation of

Laplace-Beltrami operator on compact Riemannian manifold [11]. Naturally, the matrix D provides a measure on the data points. The importance of \mathbf{y}_i is relevant to the value of $D_{i,i}$. To make a uniform measurement and remove the arbitrary scaling factor in the embedding, LPP imposes an additional constraint:

$$\begin{aligned} \mathbf{y}^\top D \mathbf{y} &= 1 \\ \Leftrightarrow T^\top X D X^\top T &= 1. \end{aligned} \quad (3)$$

This constraint is joined into the objective function. Finally the minimal problem is reduced to

$$\arg \min_{T^\top X D X^\top T=1} T^\top X L X^\top T. \quad (4)$$

The solution can be gained by solving a generalized eigenvector decomposition:

$$X L X^\top \boldsymbol{\psi} = \lambda X D X^\top \boldsymbol{\psi}. \quad (5)$$

Let $\{\lambda_i\}_{i=1}^d$ be the d -smallest eigenvalue of (5) with ascending order, that is, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$, and $\{\boldsymbol{\psi}_i\}_{i=1}^d$ the corresponding eigenvectors. Then the solution of (4) is given by

$$T_{\text{LPP}} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_d). \quad (6)$$

For a new testing instance $\boldsymbol{\mu}$, the new data points $\tilde{\boldsymbol{\mu}}$ in the embedding space are given by

$$\tilde{\boldsymbol{\mu}} = T_{\text{LPP}}^\top \boldsymbol{\mu}. \quad (7)$$

LPP can significantly find a projection that preserves the data structure. However, due to its unsupervised nature, data points that are close to boundaries may even be put closely in the projected space. In fact, these points may belong to different classes. Besides, LPP only makes use of nearest neighborhoods, and the global geometric structure is fully ignored in the calculation procedure. This drawback makes this algorithm apt to overfit the training samples. From the above analysis, we can see that LPP is sensitive to noise for those defective samples. For this reason, LPP congenitally has some deficiencies on learning ability and robusticity.

2.2. Laplacian Discriminant Analysis. As an extension of discriminate analysis, the efficiency of Laplacian linear discriminant analysis (LapLDA) has been proved by many studies [18]. The common behavior of these approaches is that an adjacent graph is employed to model the geometrical structure of the intrinsic manifold [19]. There are two popular approaches to conduct the adjacent matrix, of which the first approach is by adopting K -nearest neighborhood (the KNN approach), and the other one is by placing an edge on two data points within a controllable Euclidean distance ϵ (the ϵ -neighborhood approach). LapLDA depicts the locality by the following quadratic function:

$$\min \sum_{i,j=1}^N \left(T^\top (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^\top S_{i,j} T \right), \quad (8)$$

where $S_{i,j}$ denotes the ‘‘weights’’ of connected points with the following definition:

$$S_{i,j} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}\right), & \text{if } \mathbf{x}_i \in \text{KNN}(\mathbf{x}_j) \\ & \text{or } \mathbf{x}_j \in \text{KNN}(\mathbf{x}_i), \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

The notation $\text{KNN}(\mathbf{x}_*)$ in (9) denotes the neighbors of \mathbf{x}_* . By this definition, the smaller the distance between two connected neighborhoods, the bigger the ‘‘weight’’ they arise, and the closer the distance they should keep in the mapped space. Nevertheless, (8) also enforces data points with bigger distance to be closer in the low dimensional subspace, which may bring chaos to the structure between connected data pairs.

To cope with this issue, some researchers proposed a novel approach that integrates both global and local structure into the objective function [14]. In order to construct a reasonable locality adjacent matrix, the typical global structure of neighborhood data points can be presented by the following penalty matrix P :

$$P_{i,j} = \begin{cases} \exp\left(-\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{t}\right), & \text{if } \mathbf{x}_i \in \text{KFN}(\mathbf{x}_j) \\ & \text{or } \mathbf{x}_j \in \text{KFN}(\mathbf{x}_i), \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where $\text{KFN}(\mathbf{x}_*)$ denotes the K -farthest neighborhood of \mathbf{x}_* and $d(\mathbf{x}_i, \mathbf{x}_j)^2$ is the square Euclidean distance of two points \mathbf{x}_i and \mathbf{x}_j , respectively.

3. Proposed Methodology

The structure of HSI data is very complex; hence it is insufficient to represent HSI data using only global property or local property. To model the complex HSI data, a novel approach, which preserves both the local and global geometric structure of data samples, is proposed in this section. The new approach is called *Locality and Global Geometric Structure Preserving* (LGGSP) projection. Detailed motivation and formulation are given below.

3.1. Capturing the Local Structure of Intra-class Samples. Inspired by [11, 14], the local structure in LGGSP is described by two adjacent matrices, that is, the similarity matrix and the diversity matrix. To model the local structure, two adjacent graphs, that is, $G_d = \{\mathcal{X}_r, D\}$ and $G_s = \{\mathcal{X}_r, S\}$, are adopted to model the diversity and similarity over the whole training data samples from the same class, where the notation \mathcal{X}_r is the whole training samples, D is the diversity matrix, and S is the similarity matrix, respectively. G_d reflects the variance of nearby data points, and G_s characterizes the similarity among nearby data points.

To make samples more separable, we define a sophisticated similarity matrix:

$$S_{i,j} = \begin{cases} p^2(l_i) \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_F^2}{\alpha}\right) \\ \cdot \left(1 + \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_F^2}{\alpha}\right)\right), \\ \text{if } \mathbf{x}_i \in \text{KNN}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \text{KNN}(\mathbf{x}_i), \\ l_i = l_j; \\ 0, \quad \text{otherwise,} \end{cases} \quad (11)$$

where $p(l_i)$ ($l_i \in \mathcal{C}$) is the class prior probabilities of the l_i th class, $\alpha > 0$ the slack parameter, and $\|\cdot\|_F^2$ the Frobenius norm.

Statistically, if two samples \mathbf{x}_i and \mathbf{x}_j are very close, that is, $\|\mathbf{x}_i - \mathbf{x}_j\|_F^2$ is small, then the distance between them is also small, and the similarity should be large enough in the embedding space. In contrast, if $\|\mathbf{x}_i - \mathbf{x}_j\|_F^2$ is large, which implies that they prefer to be dissimilar in distance, the corresponding similarity will be small. Note that, in (11), the class prior $p(l_i)$ is imposed to ensure that they have the same class prior probability in the embedded space.

On the other hand, to measure the distribution of nearby data points, diversity is introduced. Different from the similarity matrix, the numerical value of diversity between two connected samples with large distance will be large. On the other hand, diversity of two connected samples with small distance should be small. This property explicates the trivial diversity of two adjacent points from the same class. Thus, the diversity matrix D can be defined as follows:

$$D_{i,j} = \begin{cases} 1 \cdot \left[p^2(l_i) \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_F^2}{\alpha}\right) \right. \\ \cdot \left. \left(1 + \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_F^2}{\alpha}\right)\right) \right]^{-1}, \\ \text{if } \mathbf{x}_i \in \text{KNN}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \text{KNN}(\mathbf{x}_i), \\ l_i = l_j; \\ 0, \quad \text{otherwise,} \end{cases} \quad (12)$$

where the notations of t and α are the free tuning parameters.

Now consider the problem of mapping the original HSI data to a line so that the connected points from the same class can be preserved. Let $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$ be such mapped point from $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$. A reasonable criterion for selecting the ‘‘good’’

mapping should be the one that optimizes the following two objective functions:

$$\min \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 S_{i,j}, \quad (13)$$

$$\max \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 D_{i,j}. \quad (14)$$

Note that (13) incurs a heavy penalty on the within-class graph if two adjacent points \mathbf{x}_i and \mathbf{x}_j , which are close to each other, are mapped far apart, yet in fact they are from the same class. Similarly, (14) incurs a heavy penalty on the within-class graph if two neighboring points \mathbf{x}_i and \mathbf{x}_j are mapped close enough, that is, a single point, whereas they share the same label. Hence, minimizing (13) is to ensure that neighboring points which have the same label are also close in the embedding space. Simultaneously, maximizing (14) can prevent overfitting problem and the variation can also be preserved in the projected space. The limitation of (13) is that it may enforce connecting points with large distance to be very close to each other in the reduced space and lead to violations of topological structure preserving. By the constraint of (14), the situation may be alleviated.

By integrating (13) and (14) together, the structural topology can be approximately preserved in the embedding space. That is, connected data points with larger distance prefer to be larger. Simultaneously samples with small distance can be kept close enough in the embedded space. Thus, the local structure can be preserved under the objective functions of (13) and (14), respectively.

3.2. Capturing the Global Structure of Interclass Samples. To capture the global structure from different classes, an adjacent graph $G_m = \{\mathcal{X}, M\}$ is constructed over the whole training samples. The notation M denotes the weight matrix of graph G_m , and it is a variation (i.e., margin or distribution) of the connected samples from different classes on the entire training data set. Similar to local Fisher’s goal [20], we do not ‘‘weight’’ the value of different samples from different classes. The reason behind this is that, since we want to separate the samples from different classes at maximum, the affinity in the original feature space will be ignored in the embedded subspace. To encode the discriminant information into the variation matrix M , the elements of M can be defined as

$$M_{i,j} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in \text{KNN}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \text{KNN}(\mathbf{x}_i), \\ l_i \neq l_j, \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

Now consider the problem of mapping HSI data to a line so that connected data points in adjacent matrix G_m stay as far as possible. In order to encode the discriminant information, a reasonable mapping can be found by optimizing the following function:

$$\max \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 M_{i,j}. \quad (16)$$

Note that the objective function of (16) on the between-class graph G_m will incur a heavy penalty when two neighboring points \mathbf{x}_i and \mathbf{x}_j are mapped close enough, despairing the fact that the labels of two connected points \mathbf{x}_i and \mathbf{x}_j are actually different from each other. In this case, maximizing (16) will enforce the corresponding mapped points \mathbf{y}_i and \mathbf{y}_j to keep far apart. Thus, the global geometric structure of interclass samples could be well detected by (16).

3.3. Optimal Solution. Let \mathbf{x}_i and \mathbf{x}_j be the connected points in the original space, $T \in \mathbb{R}^{n \times d}$ the projected direction, \mathbf{y}_i and \mathbf{y}_j the embedded points; that is, $\mathbf{y}_i = T^\top \mathbf{x}_i$ and $\mathbf{y}_j = T^\top \mathbf{x}_j$, respectively. To solve the objective functions of (13), (14), and (16) in the Laplacian graph embedding framework, we substitute $\mathbf{y}_i^\top = T^\top \mathbf{x}_i$ into the three functions. For simple algebraic formulation, the three objective functions can be written as

$$\begin{aligned} \min \frac{1}{2} \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 S_{i,j} \\ = T^\top \mathbf{X}_r L^s \mathbf{X}_r^\top T. \end{aligned} \quad (17)$$

Likewise,

$$\begin{aligned} \max \frac{1}{2} \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 D_{i,j} \\ = T^\top \mathbf{X}_r L^d \mathbf{X}_r^\top T, \\ \max \frac{1}{2} \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 M_{i,j} \\ = T^\top \mathbf{X}_r L^m \mathbf{X}_r^\top T, \end{aligned} \quad (18)$$

where

$$L^s \equiv D^s - S, \quad L^d \equiv D^d - D, \quad L^m \equiv D^m - M. \quad (19)$$

The notations D^s , D^d , and D^m represent the N -dimensional diagonal matrices whose i th diagonal element is

$$D_{i,i}^s = \sum_j S_{i,j}, \quad D_{i,i}^d = \sum_j D_{i,j}, \quad D_{i,i}^m = \sum_j M_{i,j}, \quad (20)$$

respectively. The matrices of L^s , L^d , and L^m are the Laplacian matrices in graph embedding [15].

Now let us join the three objective functions of (17) and (18) into one objective function, and the final optimal problem reduces to finding

$$T_{\text{LGGSP}} = \arg \max_T \frac{T^\top \tilde{S}_b T}{T^\top \tilde{S}_w T}, \quad (21)$$

where

$$\begin{aligned} \tilde{S}_b &= \alpha_1 S_b + \mathbf{X}_r (\alpha_2 L^d + (1 - \alpha_1 - \alpha_2) L^m) \mathbf{X}_r^\top \\ \tilde{S}_w &= \beta S_w + (1 - \beta) (\mathbf{X}_r L^s \mathbf{X}_r^\top). \end{aligned} \quad (22)$$

The notations α_1 , α_2 , and β in (22) represent the nonnegative constants that balance the ‘‘importance’’ on each criterion, where $0 \leq \alpha_1 \leq 1$, $0 \leq \alpha_2 \leq 1$, and $0 \leq \beta \leq 1$. In the whole experiments, we take the value of $\alpha_1 = 0.8$, $\alpha_2 = 0.1$, and $\beta = 0.5$. Moreover, the notations S_b and S_w represent the between-class scatter matrix and within-class scatter matrix, respectively.

Note that optimizing the problem of (21) will lead to a generalized eigenvalue decomposition problem:

$$\tilde{S}_b \boldsymbol{\Psi} = \lambda \tilde{S}_w \boldsymbol{\Psi}. \quad (23)$$

Let the column vector $\{\boldsymbol{\psi}_i\}_{i=1}^d$ be the solution of (23), where the column vectors are corresponding to the eigenvalues that are ordered by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$. Then the optimal projected direction of LGGSP is given by

$$\begin{aligned} \mathbf{x}_i \implies \mathbf{y}_i = T_{\text{LGGSP}}^\top \mathbf{x}_i \\ T_{\text{LGGSP}} = \{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_d\}, \end{aligned} \quad (24)$$

where $\mathbf{y}_i \in \mathbb{R}^d$ is a d -dimensional vector, $T \in \mathbb{R}^{n \times d}$ is a projected direction, and $\mathbf{x}_i \in \mathbb{R}^n$ is the original high dimensional point.

4. Experiments

4.1. Experiment on Synthetic Data Sets. To illustrate the effectiveness of proposed LGGSP algorithm, five synthetic data sets were investigated, that is, a toy example, ‘‘tulip’’ data set, ‘‘ripley’’ data set, a generated multimodal example, and a ‘‘two-moon’’ data set. Seven methods, that is, LDA [21], PCA [22], LPP [10], MFA [23], LGSP [13], JGLDA [14], and proposed LGGSP algorithm, were compared. There are 100 test samples for the toy example, 100 test samples for ‘‘tulip’’, 1000 test samples for ‘‘ripley’’, 200 test samples for the multimodal example, and 100 test samples for the ‘‘two-moon’’ data set. All algorithms were implemented in Matlab language and all computations are carried out on an Acer Aspire-5750G laptop with i7-2670QM processor (2.2 GHz) and Ubuntu 12.04.1 LTS (64-bit version) operating system.

Figure 1 shows the results of a simple case, that is, two classes, for the first 3 test data sets. Several conclusions can be extracted from these examples. First of all, LDA, MFA, JGLDA, and proposed LGGSP algorithms work quite well on a simple linear separable toy example. All algorithms produce comparable results on the ‘‘tulip’’ data set. For the ‘‘ripley’’ data set, only LPP, LDA, JGLDA, and proposed LGGSP find the optimal direction. The three examples indicate the robustness of proposed LGGSP algorithm.

Figure 2 shows the experimental results for the relatively complex examples. The extracted features of LGGSP in Figure 2(a) (the multimodal example) are nearly the most optimal. In particular, LDA and LGSP yield a very poor performance for this classification problem. In the ‘‘two-moon’’ examples (showed in Figure 2(b)), LDA, JGLDA, and the proposed LGGSP outperform the remaining methods. However, MFA, PCA, LPP, and LGSP fail to find the correct direction in this demonstration. JGLDA is the most

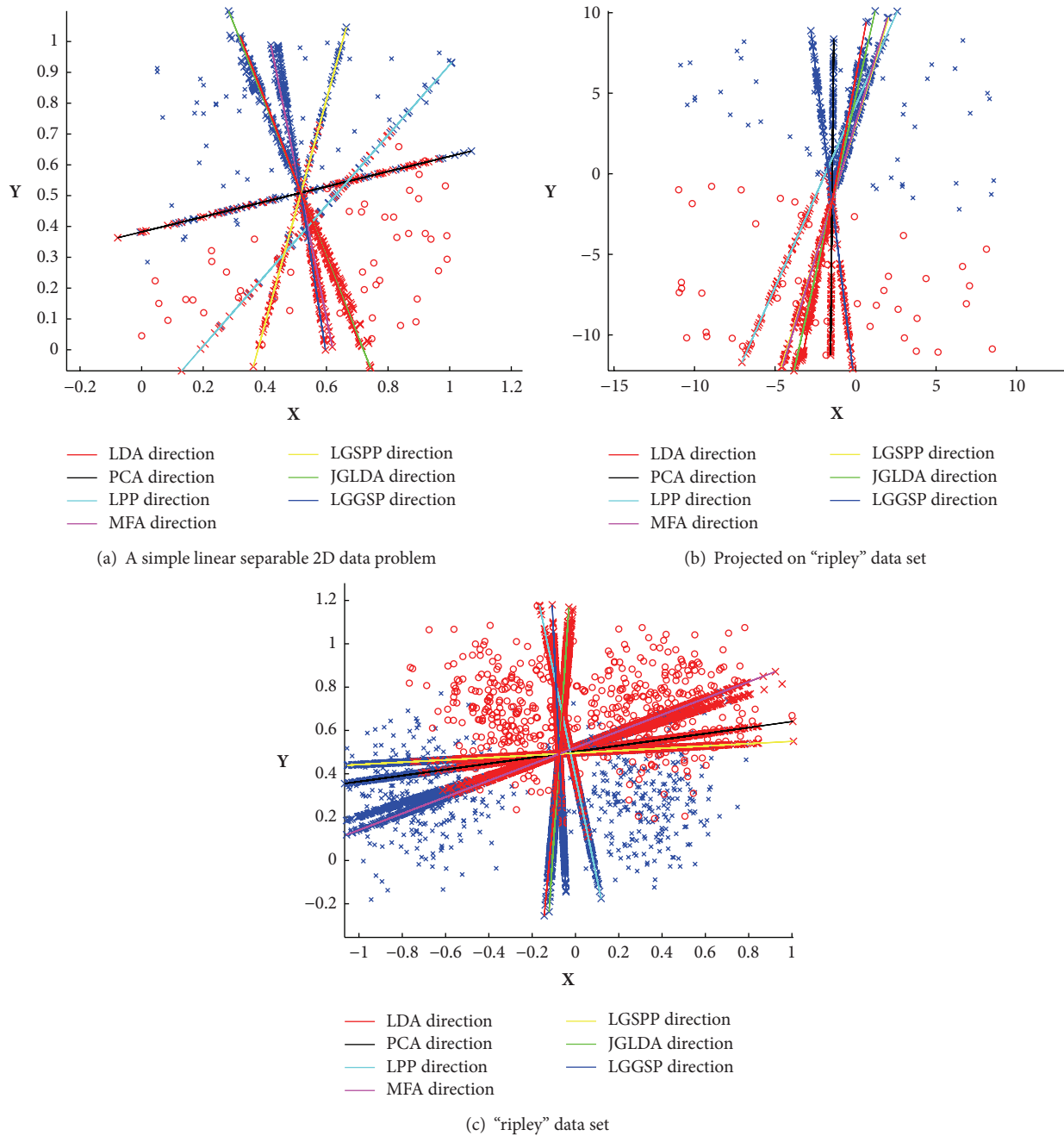


FIGURE 1: Simple examples of dimensionality reduction on generated 2D data sets produced by LDA, PCA, LPP, MFA, LGSP, JGLDA, and LGGSP, respectively. The samples from different classes are marked with different shapes and colors (blue cross × and red circle ○), respectively.

optimal method under this situation. LDA is slightly different from JGLDA and LGGSP but is still better than the other methods. The result produced by LGSP performs the worst in this data set. Note that, in this scenario, LGGSP can still produce comparable results with JGLDA, which reflects the robustness of proposed algorithm.

4.2. Experiments on Real HSI Data Set. In this subsection, we evaluate our proposed method with PCA [22], LapLDA [24], MFA [23], LPP [10], RP [9], LGSP [13], and JGLDA [14] on

a hyperspectral image, that is, the well-known Indian Pines scenario. In the following experiments, dimension reduction techniques are firstly adopted to reduce the dimensionality of input feature, following a classified procedure with a concrete classifier (e.g., KNN classifier or SVM classifier).

Indian Pine 1992 data set was gathered by National Aeronautics and Space Administration (NASA) airborne visible/infrared imaging spectrometer (AVIRIS) sensor over the northwestern Indian Pines test site in 1992, which consisted of 145×145 pixels and 224 spectral reflectance bands in the

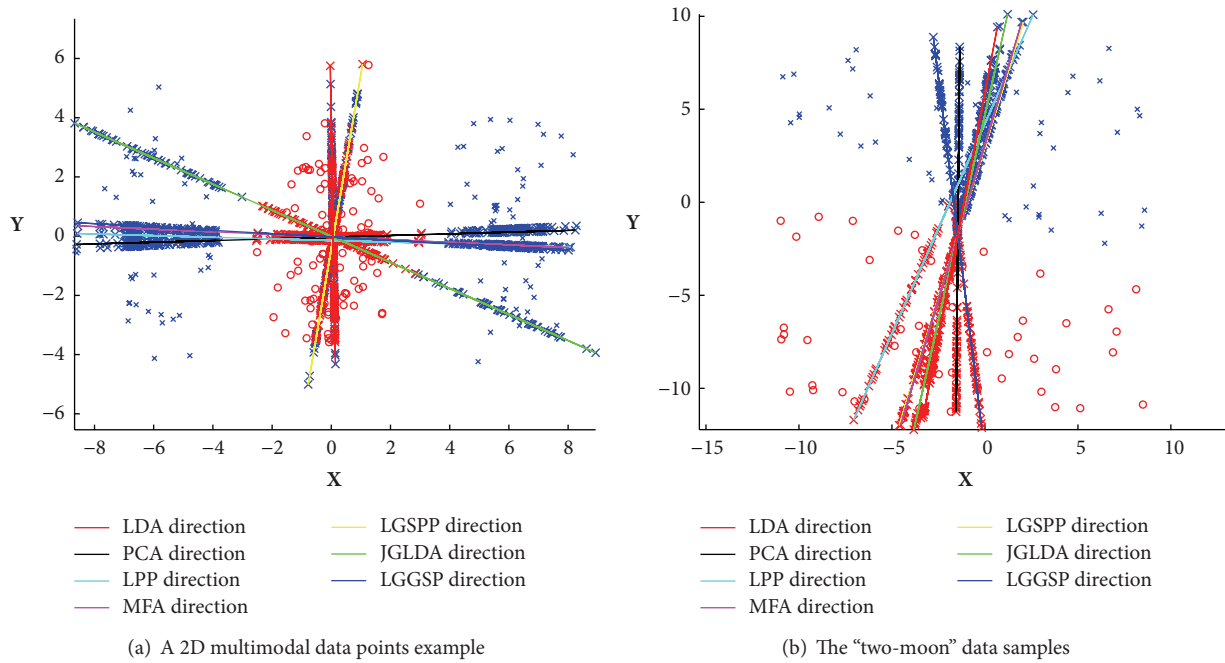


FIGURE 2: Complex examples of dimensionality reduction on generated 2D data sets produced by LDA, PCA, LPP, MFA, LGSP, JGLDA, and LGGSP, respectively. The samples from different classes are marked with different shapes and colors (blue cross \times and red circle \circ), respectively.

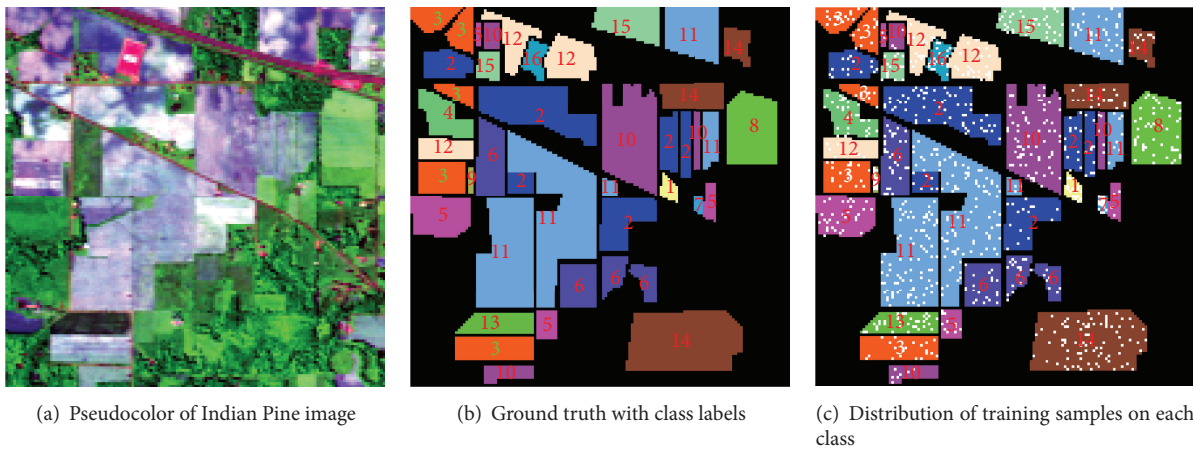


FIGURE 3: RGB composition, classification map, and training set for AVIRIS Indian Pines 1992 scenario.

wavelength range $0.4 \times 10^{-6} \sim 0.6 \times 10^{-6}$ meters, representing a vegetation-classification scenario. The water absorption bands ([104–108], [150–163], and 220) are removed before experiment. There are totally approximate 10,249 labeled pixels employed to train, test, and validate the efficacy of the proposed algorithm. The ground truth image contains 16 classes. Figure 3(a) shows the pseudocolor image of Indian Pines, along with the available labeled samples as shown in Figure 3(b). Since the purpose of this paper is to reduce the dimensionality of HSI data for classification, the performance will be measured by overall accuracy (OA), kappa coefficient (κ), and average accuracy (AA).

4.2.1. Numerical Results. There are totally 1029 samples that are chosen from the available labeled samples for training. In particular, 15 samples are chosen from each labeled class; then the missing samples are randomly chosen from the remaining unchosen samples. Table 1 summarizes the numerical statistics of training samples corresponding to each class. Finally, the remaining samples are used for testing (Figure 3(c)).

Two experiments were performed in this section. Note that the dimensionality d in LapLDA is a fixed value; that is, $d \equiv C$. For this reason, in the first experiment, we reduce the dimensionality of HSI data to 16. However, there are two

TABLE 1: Training set for Indian Pines.

ID	Class name
1	Alfalfa (46/16 , 34.78%)*
2	Corn-no-till (1428/121 , 8.47%)
3	Corn-min-till (830/72 , 8.67%)
4	Corn (237/31 , 13.08%)
5	Grass-pasture (483/48 , 9.94%)
6	Grass-trees (730/75 , 10.27%)
7	Grass-pasture-mowed (28/16 , 57.14%)
8	Hay-windrowed (478/48 , 10.04%)
9	Oats (20/15 , 75.00%)
10	Soybean-no-till (972/80 , 8.23%)
11	Soybean-min-till (2455/234 , 9.53%)
12	Soybean-clean (593/52 , 8.77%)
13	Wheat (205/34 , 16.59%)
14	Woods (1265/114 , 9.01%)
15	Buildings-grass-trees-drives (386/53 , 13.73%)
16	Stone-steel-towers (93/20 , 21.51%)
Total	10249/1029

*Numerical value in each row refers to number of total samples, number of training samples, and percentage of training samples in each class, respectively.

critical parameters (c and γ) in radial basis function (RBF) for SVM classifier. Parameter c controls the trade-off between the margin and the size of slack variables. We use five-fold cross-validation to find the best c and γ (suggested by [25]). For the other classifiers, the default parameters were employed. In the second experiment, we use all available data to generate the classified map to evaluate the performance of all methods on the whole HSI data.

Table 2 summarizes the classified performance of the first experiment. These results show that the nonlinear SVM with the RBF kernel outperforms the other ones in a general way. Moreover, it is notable that proposed LGGSP algorithm almost gains the best classified performance by different classifiers (except for the linear SVM classifier). MFA produces the worst classifier performance among all methods. JGLDA and LapLDA perform approximately the same, producing barely acceptable results. LGSP behaves a little better than JGLDA and is approximate to PCA. LPP and PCA are better than LapLDA and JGLDA but still worse than LGGSP. The MFA will be deduced in the next experiment, due to the seriously bad results.

The class accuracy for each class is listed in Table 3. The embedding data are then classified by 5NN classifier and RBF-SVM classifier. All these results show that the proposed LGGSP algorithm outperforms the other dimension reduction methods, providing almost the best classified performance in most cases.

TABLE 2: Classified performance for Indian Pine scene (%).

Item	Measurement		
	Overall accuracy	Kappa coefficient	Average accuracy
1NN			
PCA	67.37	62.72	71.02
MFA	28.57	19.25	26.51
LapLDA	48.29	40.96	50.68
RP	62.88	57.65	65.16
LPP	59.49	53.84	64.02
LGSP	61.88	56.49	65.49
JGLDA	55.20	48.73	54.57
LGGSP	75.80	72.41	81.77
5NN			
PCA	66.52	61.66	67.63
MFA	33.37	23.72	29.76
LapLDA	53.37	46.17	50.91
RP	64.34	58.99	65.21
LPP	60.98	55.13	62.35
LGSP	63.63	58.19	64.36
JGLDA	58.51	51.84	51.47
LGGSP	76.62	73.33	80.99
9NN			
PCA	66.32	61.26	67.27
MFA	36.14	25.98	29.39
LapLDA	55.66	48.32	49.08
RP	63.15	57.50	61.61
LPP	60.45	54.40	59.32
JGSP	62.98	57.21	63.44
JGLDA	59.63	52.70	49.52
LGGSP	75.75	72.28	79.31
Linear SVM			
PCA	57.13	48.55	52.00
MFA	36.93	20.20	16.92
LapLDA	50.89	40.82	28.52
RP	52.32	41.90	31.22
LPP	54.47	45.59	46.36
JGSP	54.87	45.35	41.65
JGLDA	47.88	35.26	33.52
LGGSP	63.08	55.95	48.33
Polynomial SVM			
PCA	62.40	55.74	60.29
MFA	41.78	30.94	29.77
LapLDA	51.97	43.37	37.87
RP	55.14	45.60	43.32
LPP	58.83	51.67	55.61
LGSP	60.05	52.66	57.36
JGLDA	58.25	50.49	51.65
LGGSP	72.64	68.14	72.57

TABLE 2: Continued.

Item	Overall accuracy	Measurement	
		Kappa coefficient	Average accuracy
RBF-SVM			
PCA	73.69	69.92	76.21
MFA	24.26	0.12	6.35
LapLDA	56.31	48.94	50.06
RP	71.90	67.76	72.54
LPP	68.45	63.90	71.64
LGSP	67.33	62.33	66.76
JGLDA	54.79	49.46	58.71
LGGSP	79.04	75.86	76.95

In the following experiment, we devote our attention to the visual inspection of the classified maps for all available samples. Due to the limited length of this paper, 5NN and RBF-SVM were selected as a demonstration, and classified pseudo images are shown for visual comparison. The best classification is selected to generate the classified images.

To achieve this purpose, the dimensionality is chosen as 15, while only one permitted embedding subspace for LapLDA is fixed to 16 due to the peculiar calculation. Figures 4 and 5 display the classified maps in pseudocolor images. It is clear that LapLDA and JGLDA both perform poorly, while PCA performs comparable result with proposed LGGSP in RBF-SVM classifier. However, the proposed LGGSP still works better than the other classifiers. For the 5NN classifier, it is easy to observe that the proposed LGGSP still performs equally acceptable results and better than LapLDA, JGLDA, and LPP. When conjuncted with 5NN classifier, LPP exhibits a distinguishing classification performance, which indicates the instability of LPP. Note that, in this scenario, LPP is very competitive with the state-of-the-art linear PCA. Moreover, performances of LGSP by two classifiers are almost the same.

The class accuracy is summarized in Table 4. From this table, it is found that different HSI data needs to achieve the highest results with different feature extraction methods and different classifiers. For example, when class 1 is under classifying (i.e., Alfalfa; see Table 1 for details), the highest class accuracy is based on the LGGSP feature extraction, plus a 5NN classifier. In contrast to class 1, the highest accuracy of class 3 (i.e., Corn-min-till) is based on PCA, followed by the RBF-SVM classifier. The reason is that the distribution of different classes may be different: for some classes, the structure may be simple; yet for others, the structure may be very complex. Many research papers [26] reported that there is no “best” classifier that works perfectly on “every data set.” Despite this, the proposed LGGSP does provide an optimal way to extract the “representative” features. The reduced data by the proposed LGGSP are more separable as

compared with those of PCA, LPP, LapLDA, LGSP, JGLDA, MFA, and RP. The purpose of this paper is that inclusion of geometric information in the form of similarity and deviation could indeed improve even more capability with eventually no additional cost under the framework of graph embedding [27].

5. Discussion and Conclusion

A novel LGGSP method was proposed in this paper for dimensionality reduction and classification. LGGSP integrates both locality and global geometry structures, where the local structure is captured by the similarity matrix and variance matrix, respectively, while the global discriminant geometric structure is characterized by a weight matrix encoding with the K -nearest neighborhood relationship from different classes. By combining three objective functions into the objective functions, proposed LGGSP algorithm can be achieved by solving a common eigenvalue decomposition. Since this method is built on the theoretical basis of graph embedding, we also supply a theoretical analysis of Laplacian method.

The effectiveness and stability of LGGSP were demonstrated both on the synthetic data sets and real hyperspectral image data set. The experimental results show that the proposed LGGSP algorithm outperforms the other methods in most cases, which is acceptable in classified performance. Moreover, the proposed LGGSP significantly outperforms the other dimension reduction methods when using all the available samples for testing.

The proposed LGGSP can also be used for other applications as a preprocessing step for object recognition and high dimensional data visualization.

Notations

$\mathbf{x}_i \in \mathbb{R}^n, \mathbf{y}_i \in \mathbb{R}^d$:	n -dimensional and d -dimensional point
N, C :	Number of total points and number of classes, respectively
$\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$:	Set of raw points in high dimensional space
$\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$:	Set of projected points in low dimensional space
$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \in \mathbb{R}^{n \times N}$:	High dimensional points
$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) \in \mathbb{R}^{d \times N}$:	Projected low dimensional points
$\mathcal{C} = \{1, 2, \dots, C\}$:	Set of C classes
$\mathcal{L} = \{l_1, l_2, \dots, l_N\}, l_i \in \mathcal{C}$:	Labels of points that correspond to \mathcal{X}
N_c :	Amount of samples from the c th class ($c \in \mathcal{C}$)
$T \in \mathbb{R}^{n \times d}$:	Transformation matrix.

TABLE 3: Comparison of class accuracy for Indian Pine scene (embedding $d \equiv 16$, %).

Item		Dimensionality reduction methodologies							
Classifier	Class ID	PCA	MFA	LapLDA	RP	LPP	LGSP	JGLDA	LGGSP
5NN	1	64.86	50.00	45.95	70.59	44.44	66.67	36.36	86.11
	2	54.56	29.71	48.24	52.15	53.05	42.80	55.33	72.10
	3	51.07	34.39	38.62	33.42	36.14	46.08	29.80	62.79
	4	28.36	8.96	6.28	33.17	13.11	21.57	11.44	57.28
	5	73.95	23.39	63.62	74.00	74.59	76.73	55.80	84.93
	6	91.85	41.46	72.38	95.50	86.69	89.34	81.14	93.72
	7	81.82	54.55	81.82	88.89	100.00	90.00	54.55	100.00
	8	92.82	58.54	92.64	85.71	88.33	84.28	97.02	96.11
	9	100.00	0.00	60.00	80.00	75.00	75.00	0.00	100.00
	10	67.58	16.17	36.69	59.79	51.82	57.35	29.09	72.99
	11	67.96	37.33	60.59	69.19	62.82	70.94	65.86	70.77
	12	25.41	17.99	14.91	35.60	26.43	34.87	28.88	69.45
	13	91.11	12.50	69.61	87.43	91.76	91.38	87.86	97.30
	14	94.28	53.31	72.98	91.12	86.57	92.33	91.93	93.17
	15	17.25	6.36	19.63	12.76	28.79	7.06	11.95	54.14
	16	79.17	31.58	30.67	73.97	78.08	83.33	86.49	84.93
RBF-SVM	1	78.38	0.00	29.73	79.41	52.78	69.70	75.76	61.11
	2	67.82	0.00	49.54	58.53	58.93	49.62	56.24	74.24
	3	63.81	0.00	37.26	46.13	54.14	40.77	48.21	62.01
	4	49.25	0.00	3.86	50.50	52.43	35.29	25.87	62.14
	5	84.11	0.00	56.25	83.41	78.09	74.50	72.10	84.02
	6	93.16	0.00	67.95	91.30	85.80	91.74	80.99	90.88
	7	90.91	0.00	72.73	55.56	100.00	80.00	81.82	100.00
	8	96.30	1.55	94.54	88.89	93.36	84.97	89.68	96.80
	9	100.00	0.00	40.00	100.00	100.00	75.00	0.00	80.00
	10	57.45	0.00	29.78	61.16	57.86	62.74	44.09	67.54
	11	74.16	100.00	72.37	76.88	69.47	77.87	34.49	82.49
	12	57.12	0.00	15.79	58.72	41.59	38.75	41.13	80.36
	13	91.11	0.00	69.61	93.19	93.96	88.51	90.75	85.95
	14	92.92	0.00	78.03	93.25	88.75	89.71	76.37	93.34
	15	36.84	0.00	35.58	43.92	45.15	28.24	43.44	50.00
	16	86.11	0.00	48.00	79.45	73.97	80.77	78.38	60.27



FIGURE 4: RGB composition and classification map for AVIRIS Indian Pines. The overall accuracy/kappa coefficient/average accuracy are also included at the top of each map, respectively.

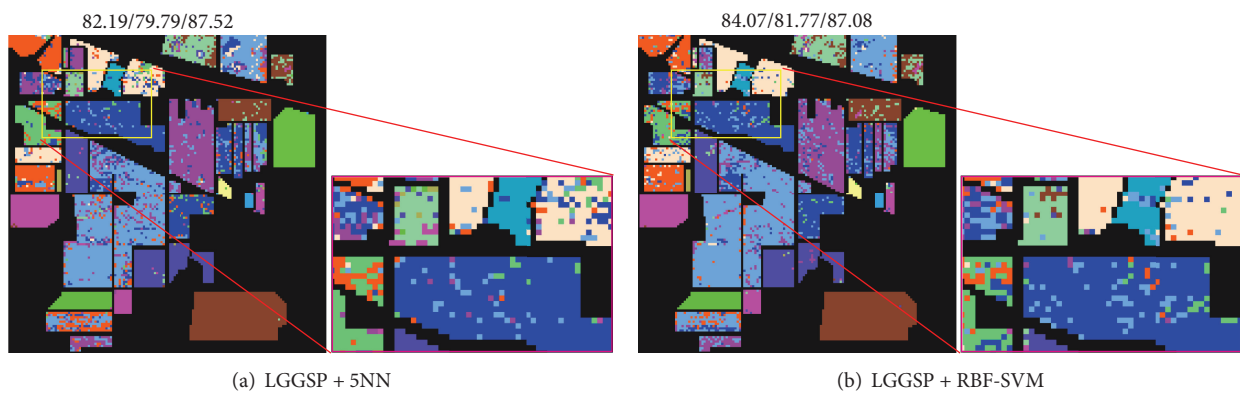


FIGURE 5: RGB composition and classification map for AVIRIS Indian Pines of proposed LGGSP.

TABLE 4: Comparison of class accuracy for Indian Pine scene (embedding $d \equiv 15$, %).

Classifier	Item Class ID	Dimension reduction methodologies						
		PCA	LapLDA	RP	LPP	LGSP	JGLDA	LGGSP
5NN	1	83.33	76.09	75.00	82.61	51.35	69.57	91.30
	2	58.44	54.62	57.80	70.24	49.92	70.31	77.24
	3	53.60	38.80	38.85	54.46	36.63	42.89	69.64
	4	41.03	16.88	48.45	46.41	27.59	13.08	85.65
	5	86.92	71.43	82.85	85.30	79.50	74.12	94.20
	6	94.78	65.21	92.35	92.47	87.96	75.89	98.08
	7	92.31	75.00	81.82	82.14	72.73	64.29	96.43
	8	92.64	92.26	89.12	98.33	94.68	99.37	99.79
	9	80.00	50.00	80.00	100.00	80.00	10.00	100.00
	10	77.69	40.74	72.94	51.23	68.16	38.37	82.41
	11	69.49	54.83	63.72	64.73	70.45	65.50	73.16
	12	44.14	26.14	30.98	19.90	21.08	24.45	76.05
	13	96.70	63.90	91.06	99.51	86.49	95.61	99.51
	14	96.45	67.51	90.27	95.65	90.46	94.78	95.18
	15	15.00	21.50	26.81	36.53	24.55	31.09	67.10
	16	85.26	75.27	80.00	90.32	84.42	94.62	94.62
RBF-SVM	1	90.74	76.09	84.38	86.96	64.86	76.09	80.43
	2	81.31	63.45	74.17	77.31	49.77	57.42	82.84
	3	70.86	44.58	57.74	59.40	46.38	57.71	62.29
	4	56.84	12.66	67.53	81.01	59.11	39.24	77.22
	5	95.37	57.56	88.42	89.86	83.11	83.44	92.34
	6	96.12	64.79	92.35	92.60	89.90	77.53	96.85
	7	100.00	53.57	81.82	96.43	81.82	92.86	89.29
	8	94.48	95.82	93.20	99.58	91.44	94.14	99.37
	9	100.00	60.00	100.00	100.00	100.00	75.00	100.00
	10	71.07	34.36	75.23	53.70	64.94	60.70	66.56
	11	79.62	65.34	72.39	80.77	77.42	33.89	84.32
	12	81.92	32.04	51.99	50.93	42.16	60.54	69.31
	13	99.53	69.27	93.30	98.54	76.76	96.59	97.07
	14	94.13	81.50	88.92	95.65	92.23	77.55	94.47
	15	53.95	18.39	46.99	60.62	39.82	72.02	72.28
	16	94.74	77.42	93.33	92.47	83.12	94.62	70.97

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Research Grants of University of Macau, MYRG205(Y1-L4)-FST11-TYY, MYRG187(Y1-L3)-FST11-TYY, and RDG009/FST-TYY/2012, and the Science and Technology Development Fund (FDCT) of Macau, 100-2012-A3 and 026-2013-A. This research project was also supported by the National Natural Science Foundation of China, 61273244 and Guangxi Science and Technology Found, KY2015YB323.

References

- [1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.
- [2] A. Plaza, J. A. Benediktsson, J. W. Boardman et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, supplement 1, pp. S110–S122, 2009.
- [3] M. Hasanlou and F. Samadzadegan, "Comparative study of intrinsic dimensionality estimation and dimension reduction techniques on hyperspectral images using K-NN classifier," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 6, pp. 1046–1050, 2012.

- [4] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3044–3054, 2007.
- [5] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.
- [6] L. Ma, M. M. Crawford, and J.-W. Tian, "Generalised supervised local tangent space alignment for hyperspectral image classification," *Electronics Letters*, vol. 46, no. 7, pp. 497–498, 2010.
- [7] X. Chen, Y. Chen, and A. Hero, "Shrinkage fisher information embedding of high dimensional feature distributions," in *Proceedings of the Conference Record of the 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR '11)*, pp. 1877–1882, IEEE, Pacific Grove, Calif, USA, November 2011.
- [8] A. Mohan, G. Sapiro, and E. Bosch, "Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 206–210, 2007.
- [9] C. K. Chui and J. Wang, "Randomized anisotropic transform for nonlinear dimensionality reduction," *GEM—International Journal on Geomathematics*, vol. 1, no. 1, pp. 23–50, 2010.
- [10] X. He and P. Niyogi, *Locality Preserving Projections*, MIT Press, Cambridge, Mass, USA, 2004.
- [11] W. K. Wong and H. T. Zhao, "Supervised optimal locality preserving projection," *Pattern Recognition*, vol. 45, no. 1, pp. 186–197, 2012.
- [12] S. Vasuhi and V. Vaidehi, "Identification of human faces using orthogonal locality preserving projections," in *Proceedings of the International Conference on Signal Processing Systems (ICSPS '09)*, pp. 718–722, May 2009.
- [13] H. Cheng, K. A. Hua, and K. Vu, "Local and global structures preserving projection," in *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '07)*, vol. 2, pp. 362–365, IEEE Computer Society, Washington, DC, USA, October 2007.
- [14] Q. Gao, J. Liu, H. Zhang, X. Gao, and K. Li, "Joint global and local structure discriminant analysis," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 626–635, 2013.
- [15] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.
- [16] Y. Ren, G. Zhang, G. Yu, and X. Li, "Local and global structure preserving based feature selection," *Neurocomputing*, vol. 89, pp. 147–157, 2012.
- [17] X. He, D. Cai, and J. Han, "Learning a maximum margin subspace for image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 2, pp. 189–201, 2008.
- [18] Q. Gao, H. Zhang, and J. Liu, "Two-dimensional margin, similarity and variation embedding," *Neurocomputing*, vol. 86, pp. 179–183, 2012.
- [19] X. Yin and Q. Huang, "Integrating global and local structures in semi-supervised discriminant analysis," in *Proceedings of the 3rd International Symposium on Intelligent Information Technology Application (IITA '09)*, vol. 1, pp. 720–723, Nanchang, China, November 2009.
- [20] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.
- [21] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 862–873, 2009.
- [22] A. Agarwal, T. El-Ghazawi, H. El-Askary, and J. Le-Moigne, "Efficient hierarchical-PCA dimension reduction for hyperspectral imagery," in *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, pp. 353–356, Cairo, Egypt, December 2007.
- [23] C. Hou, F. Nie, C. Zhang, and Y. Wu, "Learning an orthogonal and smooth subspace for image classification," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 303–306, 2009.
- [24] J. Chen, J. Ye, and Q. Li, "Integrating global and local structures: a least squares framework for dimensionality reduction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.
- [25] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.
- [26] B.-C. Kuo, C.-H. Li, and J.-M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1139–1155, 2009.
- [27] H. Zhang, Z.-J. Zha, S. Yan, M. Wang, and T.-S. Chua, "Robust non-negative graph embedding: towards noisy data, unreliable graphs, and noisy labels," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2464–2471, June 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

