

# Weighted Sparse Graph Based Dimensionality Reduction for Hyperspectral Images

Wei He, *Student Member, IEEE*, Hongyan Zhang, *Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, Wilfried Philips, *Senior Member, IEEE*, and Wenzhi Liao, *Member, IEEE*

**Abstract**—Dimensionality reduction (DR) is an important and helpful preprocessing step for hyperspectral image (HSI) classification. Recently, sparse graph embedding (SGE) has been widely used in the DR of HSIs. SGE explores the sparsity of the HSI data and can achieve good results. However, in most cases, locality is more important than sparsity when learning the features of the data. In this letter, we propose an extended SGE method: the weighted sparse graph based DR (WSGDR) method for HSIs. WSGDR explicitly encourages the sparse coding to be local and pays more attention to those training pixels that are more similar to the test pixel in representing the test pixel. Furthermore, WSGDR can offer data-adaptive neighborhoods, which results in the proposed method being more robust to noise. The proposed method was tested on two widely used HSI data sets, and the results suggest that WSGDR obtains sparser representation results. Furthermore, the experimental results also confirm the superiority of the proposed WSGDR method over the other state-of-the-art DR methods.

**Index Terms**—Dimensionality reduction (DR), hyperspectral image (HSI), nearest neighbor graph, sparse graph embedding (SGE), weighted sparse coding.

## I. INTRODUCTION

**H**YPERSPECTRAL images (HSIs) are acquired by high-spectral-resolution sensors and consist of hundreds of contiguous narrow spectral bands. With the wealth of available spectral information, hyperspectral imagery has become an invaluable tool for the detection, identification, and classification of materials and objects with complex compositions [1]. However, new challenges arise when dealing with extremely large hyperspectral data sets [2], [3]. When the ratio between the feature dimension (spectral bands) and the number of data samples (in vector-based pixels) is vastly different, high-dimensional data suffer from the well-known curse of dimensionality. In addition, high-dimensionality data processing also requires huge computational resources and storage capacities [4]. It is therefore an important preprocessing step to reduce the dimension of hyperspectral imagery.

Manuscript received November 14, 2015; revised January 5, 2016; accepted February 26, 2016. Date of publication March 18, 2016; date of current version April 20, 2016. This work was supported in part by the National Natural Science Foundation of China under Grants 61201342 and 41431175 and in part by the Fundamental Research Funds for Central Universities under Grant 2015904020202. (*Corresponding author: Hongyan Zhang.*)

W. He, H. Zhang, and L. Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, and the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zhanghongyan@whu.edu.cn).

W. Philips and W. Liao are with the Department of Telecommunications and Information, Ghent University, 9000 Ghent, Belgium.

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

Digital Object Identifier 10.1109/LGRS.2016.2536658

As a preprocessing step, dimensionality reduction (DR) tries to find a low-dimensional representation for high-dimensional data that may contain crucial information. To date, many DR methods have been proposed for hyperspectral imagery. These DR methods can be classified into unsupervised [5], [6], supervised [7], [8], and semisupervised approaches [9]. Recently, a general graph embedding (GE) framework [10] has been proposed to formulate most of the existing DR methods. In the GE framework, there are two main steps: graph construction and projection computing. A graph is a mathematical representation that describes the geometric structures of data nodes [11]. In a graph, each element measures the similarity of a pair of vertices. An appropriate graph provides a high level of DR and preserves the manifold structures of the data. Traditionally,  $k$ -nearest neighbor and  $\varepsilon$ -radius ball [12] have been used to construct the graph. In [4], a new method which integrates the spatial and spectral information of the HSI was proposed to learn a local discriminant graph. In recent years, sparse representation (SR) [13] has been exploited to produce a graph whose edges are intended to be sparse [14], [15]. This sparse graph embedding (SGE) explores the sparsity structure of the data and has been widely used in HSI DR. Ly *et al.* [11] proposed block sparse graph based discriminant analysis (BSGDA), which learns a block sparse graph for a supervised DR. In [16], collaborative representation among labeled samples was adopted to realize collaborative graph based discriminant analysis for HSIs. In [17], spatial information was integrated into the sparse graph learning process, and a spatial and spectral regularized local discriminant embedding (SSRLDE) method was proposed for the DR of HSIs.

The SGE-based methods can achieve state-of-the-art HSI DR performances. However, as pointed out in [18], the sparse coding used in SGE is helpful for learning only when the coding is local. That is to say, under certain assumptions, locality is more important than sparsity. Unfortunately, sparsity does not always guarantee locality [18]. To overcome the drawback of sparse coding in SGE, we propose a more robust DR method named weighted sparse graph based DR (WSGDR) for HSIs. WSGDR uses weighted sparse coding instead of sparse coding to explicitly encourage the sparse coding to be local. In WSGDR, the test pixel is more prone to establish relations with the pixels that are nearest to the test pixel.

## II. PROPOSED METHOD

First of all, we introduce the notations adopted throughout this letter. For hyperspectral data samples  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{B \times N}$ , we have the corresponding class labels  $\mathbf{Z} = [z_1, z_2, \dots, z_N]$ , where the class label of the  $m$ th pixel  $\mathbf{x}_m$  is  $z_m \in \{1, 2, \dots, p\}$  and  $p$  is the number of classes in the data

set.  $\mathbf{X}^i = [\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_N]$  denotes all of the samples except for the  $i$ th pixel.  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{K \times N}$  represents the low-dimensional features extracted from  $\mathbf{X}$ , and the corresponding DR projection is  $\mathbf{P}$ . We let  $\mathbf{G} = \{\mathbf{X}, \mathbf{W}\}$  be a graph, where  $\mathbf{X}$  is the vertex set and  $\mathbf{W} \in \mathbb{R}^{N \times N}$  is the similarity matrix for the vertex set  $\mathbf{X}$ . The aim of the DR technique is to find the matrix  $\mathbf{P}$  to project the data  $\mathbf{X}$  into the low-dimensional space  $\mathbf{Y}$  with  $\mathbf{Y} = \mathbf{P}^T \mathbf{X}$  while maintaining the similarity or affinity between vertices in the original graph  $\mathbf{W}$ .

#### A. NPE

Neighborhood preserving embedding (NPE) is an unsupervised DR method which aims to preserve the local neighborhood structure of the data. The first step of NPE, as introduced in [12], is to construct an adjacency graph. Typically, there are two ways to construct the adjacency graph.

- $k$ -nearest neighbors (KNNs): put a directed edge from node  $i$  to  $j$  if  $\mathbf{x}_j$  is among the KNNs of  $\mathbf{x}_i$ .
- $\varepsilon$  neighborhood: put an edge between nodes  $i$  and  $j$  if  $\|\mathbf{x}_j - \mathbf{x}_i\| \leq \varepsilon$ .

The adjacency graph provides the neighborhood information which can be used to compute the weights for each pixel in its neighborhood. We let  $\mathbf{W}$  denote the weight matrix, with  $\mathbf{W}_{i,j}$  being nonzero if pixel  $\mathbf{x}_j$  belongs to a neighbor of  $\mathbf{x}_i$  and 0 if it does not. The weights related to the neighbors can then be computed by minimizing the following objective function:

$$\min \sum_i \left\| \mathbf{x}_i - \sum_j \mathbf{W}_{i,j} \mathbf{x}_j \right\|^2 \quad \text{s.t.} \quad \sum_j \mathbf{W}_{i,j} = 1, \quad j = 1, 2, \dots, N. \quad (1)$$

From (1), we can find that the matrix  $\mathbf{W}$  denotes the weight matrix which summarizes the contribution of the  $j$ th neighboring pixel to the reconstruction of the  $i$ th pixel. After the optimal coefficients are obtained, the third step is to compute the projections. In this step, the DR is converted into solving the following objective function:

$$\begin{aligned} \mathbf{P}^* &= \arg \min_{\mathbf{P}^T \mathbf{X} \mathbf{X}^T \mathbf{P} = \mathbf{I}} \sum_i \left\| \mathbf{P}^T \mathbf{x}_i - \sum_j \mathbf{W}_{i,j} \mathbf{P}^T \mathbf{x}_j \right\|^2 \\ &= \arg \min_{\mathbf{P}^T \mathbf{X} \mathbf{X}^T \mathbf{P} = \mathbf{I}} \text{tr}(\mathbf{P}^T \mathbf{X} \mathbf{L}_s \mathbf{X}^T \mathbf{P}) \end{aligned} \quad (2)$$

where  $\mathbf{L}_s = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$  and  $\mathbf{I}$  is the identity matrix.

#### B. L1-Graph

The adjacency graph effectively characterizes the pairwise relations, while the relations between pixels can also be exactly estimated by SR. Therefore, it is natural to construct the adjacency graph by L1 optimization, since the L1 linear reconstruction error minimization can naturally lead to SR for pixels [11], [19].

Given a pixel  $\mathbf{x}_i \in \mathbf{X}$ , the SR model aims to represent  $\mathbf{x}_i$  using as few entries of  $\mathbf{X}$  as possible, except for  $\mathbf{x}_i$  itself, with the SR coefficient vector  $\boldsymbol{\alpha}^i$ , which can be solved as follows:

$$\arg \min_{\boldsymbol{\alpha}^i} \|\boldsymbol{\alpha}^i\|_1, \quad \text{s.t.} \quad \mathbf{x}_i = \mathbf{X}^i \boldsymbol{\alpha}^i, \quad i = 1, 2, \dots, N. \quad (3)$$

Subsequently, the similarity matrix element  $\mathbf{W}_{i,j}$  can be denoted as

$$\mathbf{W}_{i,j} = \begin{cases} \alpha_j^i & \text{if } i > j \\ \alpha_{j-1}^i & \text{if } i < j \\ 0 & \text{if } i = j. \end{cases} \quad (4)$$

After obtaining the L1-graph matrix  $\mathbf{W}$ , we can adopt the same procedure as (2) to implement the DR of high-dimensional data.

In [11], a supervised version of the L1-graph named BSGDA was proposed to implement the DR of HSI data. BSGDA assumes that the SR of a labeled sample is estimated using only the samples within the same class. As a result, the graph matrix  $\mathbf{W}$  obtains the block structure, which is more accurate for the pairwise relations of the pixels.

#### C. WSGDR

The L1-graph explores the sparsity structure of the data and has been widely used in the DR of HSIs. As described in [20], sparse coding results in fewer reconstruction errors owing to the overcomplete dictionary and is also robust to noise. Unfortunately, due to the overcompleteness of the dictionary, the sparse coding process may select quite different bases for similar pixels to favor sparsity, thus losing the correlations between similar pixels [20]. From another aspect, the locality information, as pointed out in [12] and [18], is more important than sparsity under certain conditions, as locality must lead to sparsity but not necessarily *vice versa*. In addition, the locality information can ensure that similar pixels will have similar bases to construct the test pixels. In this letter, we propose an improved SR method named WSGDR, which incorporates a locality constraint into the sparse coding constraint to learn the local SR of the test pixels.

Due to the superiority of BSGDA [11], we also describe a block version of WSGDR. Suppose  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p] \in \mathbb{R}^{B \times N}$ , where  $\mathbf{X}_i \in \mathbb{R}^{B \times N_i}$  are the samples belonging to the  $i$ th class,  $\sum_{i=1}^p N_i = N$ , and  $\mathbf{x}_i^j$  denotes the sample selected from the  $j$ th column of  $\mathbf{X}_i$ . The proposed weighted sparse graph solves the following weighted sparse coding problem:

$$\arg \min_{\boldsymbol{\alpha}_i^j} \left\| \mathbf{d}^{i,j} \circ \boldsymbol{\alpha}_i^j \right\|_1 \quad \text{s.t.} \quad \mathbf{x}_i^j = \mathbf{X}_i^j \boldsymbol{\alpha}_i^j, \quad i = 1, 2, \dots, p, \quad j = 1, 2, \dots, N_i \quad (5)$$

where  $\mathbf{X}_i^j$  denotes training sample  $\mathbf{X}_i$ , except for the  $j$ th column  $\mathbf{x}_i^j$ ; the operator  $\circ$  means elementwise multiplication;  $\boldsymbol{\alpha}_i^j$  represents the SR coefficient vector of  $\mathbf{x}_i^j$  with respect to the dictionary base  $\mathbf{X}_i^j$ ; and  $\mathbf{d}^{i,j}$  is the locality descriptor, which measures the similarity between  $\mathbf{x}_i^j$  and the other samples in  $\mathbf{X}_i^j$ . Specifically

$$\mathbf{d}^{i,j} = \left[ \frac{\text{dist}(\mathbf{x}_i^j, \mathbf{X}_i^{j,1})}{\sigma}, \dots, \frac{\text{dist}(\mathbf{x}_i^j, \mathbf{X}_i^{j,N_i-1})}{\sigma} \right]^T \quad (6)$$

where  $\text{dist}(\mathbf{x}_i^j, \mathbf{X}_i^{j,l}) = \|\mathbf{x}_i^j - \mathbf{X}_i^{j,l}\|$  is the Euclidean distance between  $\mathbf{x}_i^j$  and  $\mathbf{X}_i^{j,l}$  (denoting the  $l$ th atom of dictionary  $\mathbf{X}_i^j$ ) and  $\sigma$  is used for adjusting the tolerance for the locality descriptor. A larger  $\text{dist}(\mathbf{x}_i^j, \mathbf{x}_i^k)$  indicates a greater distance between  $\mathbf{x}_i^j$  and  $\mathbf{x}_i^k$ , and it can effectively characterize the similarity between the test sample and  $\mathbf{X}_i^j$ . As a result, the coding

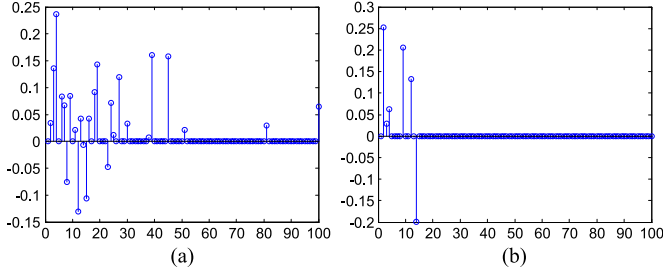


Fig. 1. Coefficients of a test pixel belonging to the first class of the Pavia University data set via different methods. (a) Coefficients of the L1-graph. (b) WSGDR.

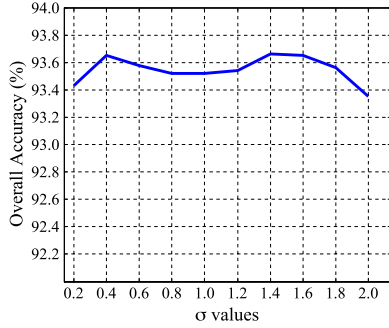


Fig. 2. OA values of the proposed method on the Pavia University data set with respect to parameter  $\sigma$ .

coefficient of the weighted sparse coding tends to integrate the locality and sparsity characteristics together. For a test pixel, the weighted sparse graph computes the weight for a training pixel according to the distance or similarity between the test pixel and the remaining training pixels. It then seeks the weighted representation of the test pixel with respect to the training pixels based on the L1-norm. The goal of WSGDR is that, given a test pixel, it pays more attention to those remaining training pixels that are more similar to the test pixel in representing the test pixel. In this case, WSGDR integrates the two properties (locality and sparsity) together to improve the robustness and representation accuracy of the test pixels.

It is clear that the WSGDR algorithm is an extension of the typical L1-graph algorithm. In fact, if the locality descriptor is set to an all-ones vector in (6), WSGDR degrades to the L1-graph method. The superiority of WSGDR is demonstrated in Fig. 1. This figure shows a test pixel belonging to the first class of the Pavia University data set (the dictionary contains 100 atoms, which are sorted by the increasing distance to the test pixel). Fig. 1(a) presents the coefficients of the L1-graph, and Fig. 1(b) gives the results of WSGDR. It can be observed that WSGDR can obtain sparser representation results, and in addition, it selects the pixels that are nearer to the test pixel to contribute to the SR. From another aspect, compared to NPE, WSGDR offers data-adaptive neighborhoods, i.e., the neighborhood data structures are not arbitrarily determined by a certain neighborhood with a fixed and predefined similarity measure.

After the sparse coding, we then construct the graph as follows. Suppose  $\mathbf{W}^i \in \mathbf{R}^{N_i \times N_i}$  is the graph matrix of the  $i$ th class sample; then, it can be denoted as

$$\mathbf{W}_{j,k}^i = \begin{cases} \alpha_{i,k}^j & \text{if } j > k \\ \alpha_{i,k-1}^j & \text{if } j < k \\ 0 & \text{if } j = k. \end{cases} \quad (7)$$

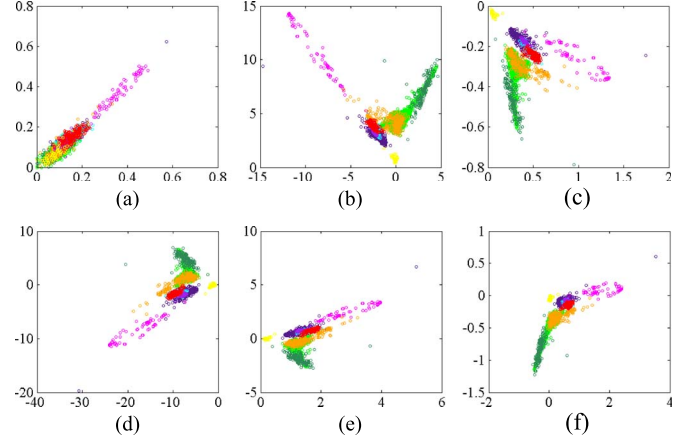


Fig. 3. Scatter plots of the first two bands for the different DR methods on the Pavia University data set. (a) Original. (b) LFDA [7]. (c) NPE [12]. (d) SGE [14]. (e) BSGDA [11]. (f) Proposed WSGDR.

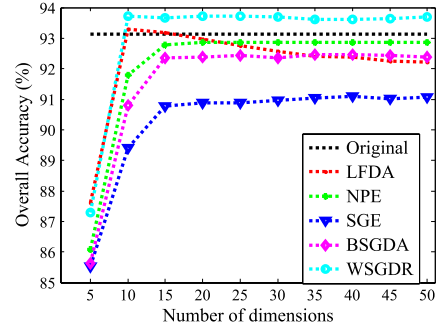


Fig. 4. Overall classification comparison of the different DR methods on the Pavia University data set with respect to different dimensions.

As a result, the graph matrix of all of the samples is  $\mathbf{W} = \text{diag}(\mathbf{W}^1, \mathbf{W}^2, \dots, \mathbf{W}^p)$ . We finally adopt the same procedure as (2) to assemble the transformation matrix. **Algorithm 1** illustrates the whole process of the proposed block version of WSGDR. In the algorithm, the SPAMS package [21],<sup>1</sup> is adopted to solve the weighted sparse coding problem (5).

---

**Algorithm 1** Block Version of Weighted Sparse Graph Based Dimensionality Reduction (WSGDR)

---

**Input:** Data set  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p] \in \mathbf{R}^{B \times N}$ , the desired reduced dimensionality  $K$ .

**Output:** Transformation matrix  $\mathbf{P}$

**for**  $i = 1$  to  $p$  **do**

**for**  $j = 1$  to  $N_i$  **do**

    Set  $\mathbf{d}^{i,j} \leftarrow 1 \times (N_i - 1)$  zero vector (locality constraint parameter)

    Compute  $\mathbf{d}^{i,j}$  via (6)

    Weighted sparse coding via (5)

**end for**

  Construct the  $i$ th class sample similarity matrix  $\mathbf{W}^i$  via (7)

**end for**

$\mathbf{W} = \text{diag}(\mathbf{W}^1, \mathbf{W}^2, \dots, \mathbf{W}^p)$ ,

<sup>1</sup> Available: <http://spams-devel.gforge.inria.fr/>.

TABLE I

OA (IN PERCENT), AA (IN PERCENT), INDIVIDUAL CLASS ACCURACY (IN PERCENT),  $\kappa$ , AND STANDARD DEVIATION OF TEN CONDUCTED MONTE CARLO RUNS OBTAINED BY THE DIFFERENT DR METHODS ON THE PAVIA UNIVERSITY DATA SET (THE REDUCED DIMENSION IS  $K = 15$ )

| Class                     | Samples |       | Methods           |                   |                   |            |                   |                   |
|---------------------------|---------|-------|-------------------|-------------------|-------------------|------------|-------------------|-------------------|
|                           | Train   | Test  | Origin            | LFDA              | NPE               | SGE        | BSGDA             | WSGDR             |
| Asphalt (C1)              | 332     | 6299  | 91.84±0.61        | <b>93.95±0.75</b> | 92.11±0.75        | 90.11±1.28 | 92.41±0.74        | 93.18±0.87        |
| Meadows (C2)              | 932     | 17717 | <b>98.29±0.26</b> | 98.00±0.34        | 97.47±0.41        | 96.95±0.32 | 97.29±0.43        | 97.88±0.24        |
| Gravel (C3)               | 105     | 1994  | 75.30±3.36        | 72.17±3.74        | 76.25±2.39        | 71.23±3.65 | 74.89±2.52        | <b>77.37±3.06</b> |
| Tree (C4)                 | 153     | 2911  | 91.51±1.52        | <b>94.04±0.97</b> | 91.98±1.89        | 89.86±2.01 | 92.29±1.79        | 92.61±1.62        |
| Painted metal sheets (C5) | 67      | 1278  | 99.05±0.27        | 98.79±0.99        | <b>99.12±0.28</b> | 98.78±0.33 | 99.02±0.36        | 98.90±1.24        |
| Bare Soil (C6)            | 251     | 4778  | 85.79±0.84        | 86.97±1.34        | 85.71±1.72        | 78.77±2.97 | 82.44±1.25        | <b>88.74±1.03</b> |
| Bitumen (C7)              | 67      | 1263  | 83.82±2.28        | 75.65±3.59        | 83.38±3.43        | 81.46±2.83 | 83.25±2.07        | <b>84.64±2.18</b> |
| Self-blocking bricks (C8) | 184     | 3498  | <b>90.43±1.21</b> | 89.65±2.02        | 89.34±0.86        | 87.03±1.39 | 89.66±1.65        | 89.85±1.90        |
| Shadows (C9)              | 47      | 900   | 99.64±0.30        | 99.60±0.23        | 99.76±0.27        | 99.52±0.34 | <b>99.79±0.13</b> | 99.48±0.60        |
| OA                        | --      | --    | 93.13±0.22        | 93.17±0.26        | 92.79±0.27        | 90.76±0.63 | 92.35±0.38        | <b>93.66±0.21</b> |
| AA                        | --      | --    | 90.63±0.48        | 89.87±0.61        | 90.57±0.56        | 88.19±0.69 | 90.12±0.48        | <b>91.41±0.39</b> |
| $\kappa$                  | --      | --    | 91.10±0.28        | 91.15±0.33        | 90.69±0.35        | 88.15±0.78 | 90.14±0.48        | <b>91.78±0.27</b> |

Solve the generalized eigenvalue problem:

$\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{p}_k = \lambda_k \mathbf{X}\mathbf{X}^T \mathbf{p}_k$ , where  $\lambda_k$  is the  $k$ th minimum eigenvalue, and  $\mathbf{p}_k$  is the corresponding eigenvector.

Construct the transformation matrix  $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_K) \in \mathbf{R}^{B \times K}$

### III. EXPERIMENTAL RESULTS AND ANALYSIS

The support vector machine (SVM) classifier was used to evaluate the performance of the different DR methods in the experiments. The LIBSVM toolkit<sup>2</sup> was adopted to implement the SVM classifier with a radial basis function kernel, and the parameters were selected via cross-validation. Several other DR methods, i.e., local Fisher discriminant analysis (LFDA) [7], NPE [12], SGE [14], and BSGDA [11] were also implemented for comparison. The codes for LFDA<sup>3</sup> and NPE<sup>4</sup> were downloaded online. SGE and BSGDA were implemented using the SPAMS tool,<sup>1</sup> which was also used in the proposed WSGDR. We tested the proposed DR method based on the following two widely used HSIs. All of the experiments involved independent Monte Carlo runs, and the average overall accuracy (OA), the average accuracy (AA), the kappa statistic ( $\kappa$ ), and the standard deviation are reported.

is relatively stable with regard to the value of  $\sigma$ . This inspired us to set  $\sigma = 1.4$  in all of the experiments.

Fig. 3 illustrates the scatter plots for the different DR methods considering the first two bands or features. As shown in Fig. 3(a), the different pixels are highly mixed in the first two bands of the original image. Fortunately, after DR by the different methods, the discrimination of the pixels (from different classes) related to the first two bands is greatly enhanced.

Fig. 4 presents the OA values of the different DR methods with respect to the reduced dimension numbers. In the figure, the result of SVM on the original Pavia University data set is used as a baseline. As depicted in Fig. 4, the proposed WSGDR outperforms the other methods in almost all of the cases. In particular, WSGDR outperforms BSGDA and NPE in all of the dimension cases. That is to say, the neighborhood information and sparsity property are both important for the analysis of HSI data. In addition, we also used a reduced dimensionality of  $K = 15$ , and we present the mean OA, AA, individual class accuracy,  $\kappa$ , and standard deviation of ten Monte Carlo runs obtained by the different DR methods in Table I. From the table, we can again see that WSGDR outperforms the other methods in terms of OA, AA, and  $\kappa$  values.

#### B. Experiments With the AVIRIS Indian Pines Data Set

The second data set used in the experiments was acquired by the NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) instrument over the Indian Pines test site in Northwestern Indiana in 1992. The data set size is  $145 \times 145$  pixels and 220 bands. In our experiments, the noisy and water absorption bands were removed [22], leaving a total of 200 bands. A total of 10249 samples containing 16 classes are available, of which 10% was used as training samples and the rest was used for testing.

Table II presents the mean OA, AA, individual class accuracy,  $\kappa$ , and standard deviation of ten Monte Carlo runs obtained by the different DR methods on the Indian Pines data set. It can again be observed that the proposed WSGDR method outperforms the other methods. Notably, LFDA performs even worse than the baseline SVM on the original data. This is mainly because LFDA is sensitive to the number of nearest neighbors, and it fails in the case of a low number of training samples (e.g., alfalfa, grass-pasture-mowed, and oats).

#### A. Experiments With the Pavia University Data Set

We first tested the proposed method on the Pavia University data set, which was collected by the Reflective Optics System Imaging Spectrometer (ROSIS-03). This data set has 115 bands, with a spectral range of 0.43–0.86  $\mu\text{m}$ . After removing 12 water absorption and noisy bands, 103 bands were used in the experiments. The data set is of  $610 \times 340$  in size, and 42776 samples containing nine classes are available. In the experiments, 5% of the samples were used as training samples, and the rest was used for testing. For the proposed WSGDR, the selection of parameter  $\sigma$  affects the DR performance. Fig. 2 presents the OA values of WSGDR on the Pavia University data set with respect to  $\sigma$ . From the figure, it can be observed that the result

<sup>2</sup>Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

<sup>3</sup>Available: <http://research.cs.buct.edu.cn/liwei/>.

<sup>4</sup>Available: <http://www.cad.zju.edu.cn/home/dengcai/>.

TABLE II  
OA (IN PERCENT), AA (IN PERCENT), INDIVIDUAL CLASS ACCURACY (IN PERCENT),  $\kappa$ , AND STANDARD DEVIATION OF TEN CONDUCTED MONTE CARLO RUNS OBTAINED BY THE DIFFERENT DR METHODS ON THE INDIAN PINES DATA SET (THE REDUCED DIMENSION IS  $K = 20$ )

| Class                       | Samples |      | Methods           |                   |                   |                   |                   |                   |
|-----------------------------|---------|------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                             | Train   | Test | Origin            | LFDA              | NPE               | SGE               | BSGDA             | WSGDA             |
| Alfalfa                     | 5       | 41   | 80.49±9.55        | 58.54±14.3        | 80.00±11.7        | 73.66±12.3        | 78.29±7.40        | <b>83.17±7.58</b> |
| Corn-notill                 | 143     | 1285 | 83.14±2.22        | 83.32±1.84        | 78.44±2.87        | 79.49±2.36        | 82.88±2.11        | <b>83.56±2.02</b> |
| Corn-mintill                | 83      | 747  | 76.63±3.76        | 69.77±2.87        | 68.53±4.15        | 74.39±2.30        | 72.13±4.58        | <b>78.27±2.77</b> |
| Corn                        | 24      | 213  | 73.43±9.86        | 65.63±10.1        | 70.14±11.4        | 62.21±9.21        | 74.37±8.55        | <b>80.89±4.94</b> |
| Grass-pasture               | 48      | 435  | 91.33±1.97        | 90.90±3.27        | 90.94±2.12        | 90.39±2.61        | 92.28±1.99        | <b>93.08±1.94</b> |
| Grass-trees                 | 73      | 657  | 96.26±3.11        | 96.88±2.92        | 94.89±3.18        | 95.24±3.13        | <b>97.40±2.64</b> | 96.54±2.56        |
| Grass-pasture-mowed         | 3       | 25   | 87.20±3.68        | 49.2±22.79        | <b>88.80±4.13</b> | <b>88.80±4.13</b> | 87.60±4.40        | 88.00±3.77        |
| Hay-windrowed               | 48      | 430  | 98.88±0.98        | <b>99.53±0.49</b> | 98.74±1.32        | 98.37±1.53        | 99.16±0.53        | 99.53±0.55        |
| Oats                        | 2       | 18   | 57.22±21.7        | 29.44±27.7        | 48.33±20.6        | 50.56±24.7        | 61.11±25.9        | <b>68.89±19.8</b> |
| Soybean-notill              | 97      | 875  | <b>78.71±3.92</b> | 69.19±3.19        | 73.81±3.05        | 77.50±5.05        | 75.89±2.84        | 77.95±2.91        |
| Soybean-mintill             | 246     | 2209 | 87.22±1.40        | 86.22±2.28        | 83.57±1.63        | <b>87.52±1.17</b> | 86.47±2.07        | 84.69±1.46        |
| Soybean-clean               | 59      | 534  | 81.82±4.31        | 83.18±3.65        | 77.08±5.14        | 76.67±3.54        | 82.60±3.33        | <b>86.95±3.34</b> |
| Wheat                       | 21      | 184  | 97.93±1.25        | 96.68±2.75        | 96.96±2.13        | 96.09±2.84        | 97.88±1.24        | <b>98.75±0.85</b> |
| Woods                       | 127     | 1138 | 95.15±2.39        | 95.34±1.26        | 96.24±1.81        | 95.98±1.73        | <b>96.50±1.16</b> | 94.59±1.49        |
| Buildings-Grass-Trees-Drive | 39      | 347  | 59.94±5.80        | 67.29±3.22        | 58.33±5.42        | 53.29±6.48        | 63.23±5.44        | <b>68.41±5.53</b> |
| Stone-Steel-Towers          | 9       | 84   | 87.62±4.02        | 59.52±12.6        | 87.02±5.25        | <b>87.86±4.59</b> | 85.95±4.34        | 84.05±4.53        |
| OA                          | --      | --   | 85.82±0.99        | 83.86±0.90        | 82.73±0.45        | 84.16±1.08        | 85.45±0.65        | <b>86.20±0.79</b> |
| AA                          | --      | --   | 83.31±1.69        | 75.04±4.80        | 80.74±1.01        | 80.5±1.44         | 83.36±1.54        | <b>85.46±1.15</b> |
| $\kappa$                    | --      | --   | 84.36±1.06        | 82.25±0.97        | 81.08±0.48        | 82.57±1.16        | 83.97±0.70        | <b>84.81±0.85</b> |

#### IV. CONCLUSION

In this letter, we have proposed a novel WSGDR method for the DR of HSI data. The proposed method learns a weighted sparse graph which computes the weight for a training pixel according to the distance or similarity between the test pixel and the remaining training pixels. It then represents the test pixel by exploiting the weighted training pixels based on the L1-norm. WSGDR integrates both the locality and sparsity structure of the training pixels. The proposed method was compared with other DR methods on two HSI data sets. The experimental results confirm the superiority of the proposed WSGDR, with better performances and higher classification accuracies.

#### REFERENCES

- [1] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.
- [2] J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction in remote sensing using hyperspectral imaging [applications corner]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 149–154, Jul. 2014.
- [3] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
- [4] Y. Zhou *et al.*, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.
- [5] C.-I. Chang and Q. Du, "Interference and noise-adjusted principal components analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2387–2396, Sep. 1999.
- [6] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
- [7] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.
- [8] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.
- [9] K. Tan, S. Zhou, and Q. Du, "Semisupervised discriminant analysis for hyperspectral imagery with block-sparse graph," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1765–1769, Aug. 2015.
- [10] S. Yan *et al.*, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
- [11] N. H. Ly, D. Qian, and J. E. Fowler, "Sparse graph-based discriminant analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3872–3884, Jun. 2014.
- [12] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE 10th ICCV*, 2005, pp. 1208–1213.
- [13] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.
- [14] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with-graph for image analysis," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 858–866, Apr. 2010.
- [15] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, doi: 10.1109/TGRS.2016.2524557, to be published.
- [16] N. H. Ly, Q. Du, and J. E. Fowler, "Collaborative graph-based discriminant analysis for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2688–2696, Jun. 2014.
- [17] Z. Xue, P. Du, J. Li, and H. Su, "Simultaneous sparse graph embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 1–20, Nov. 2015.
- [18] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2223–2231.
- [19] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recog.*, vol. 43, pp. 331–341, 2010.
- [20] J. Wang *et al.*, "Locality-constrained linear coding for image classification," in *Proc. IEEE CVPR*, 2010, pp. 3360–3367.
- [21] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.
- [22] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.