Weighted Sparse Graph Based Dimensionality Reduction for Hyperspectral Images

Wei He[1], Hongyan Zhang[1]*, Liangpei Zhang[1], Wilfried Philips[2], Wenzhi Liao[2]

1. The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, P. R. China
2. Department of Telecommunications and Information, Ghent University, Ghent, Belgium

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. In this letter, we propose a weighted sparse graph based DR (WSGDR) method for HSIs. Instead of only exploring the locality structure (as in neighborhood preserving embedding) or the linearity structure (as in SGE) of the HSI data, the proposed method couples the locality and linearity properties of HSI data together in a unified framework for the DR of HSIs. The proposed method was tested on two widely used HSI data sets, and the results suggest that the locality and linearity are complementary properties for HSIs. In addition, the experimental results also confirm the superiority of the proposed WSGDR method over the other state-of-the-art DR methods.

Index terms—hyperspectral image, dimensionality reduction, weighted sparse coding, nearest neighbor graph, sparse graph embedding

I. INTRODUCTION

Hyperspectral 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 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]. 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 suffers from the well-known curse of dimensionality. In addition, high-dimensionality data processing also requires huge computational resources and storage capacities [3]. It is therefore an important preprocessing step to reduce the dimension of hyperspectral imagery.

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 [4], [5], supervised [6], [7], and semi-supervised approaches [8]. Recently, a general graph-embedding (GE) framework [9] was 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 [10]. In a graph, each element measures the similarity for 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 ε-radius ball [11] have been used to construct the graph. In [3], 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 has been exploited to produce a graph whose edges are intended to be sparse [12]. This sparse graph embedding (SGE) explores the linearity structure of the data, and has been widely used in HSI DR. Ly *et al*. [10] proposed block sparse graph based discriminant analysis (BSGDA), which learns a block sparse graph for supervised DR. In [13], collaborative representation among labeled samples was adopted to realize collaborative graph based discriminant analysis for HSIs. In [14], 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, they only explore the linearity structure of the HSI data. In this case, the sparse coding procedure may reconstruct a test pixel with training pixels which are far from the test pixel. Thus, the sparse graph tends to lose the local similarity of the training pixels. However, under certain assumptions, as pointed in [11] and [15], the locality is more essential than the sparsity. To overcome the drawback of sparse coding, we propose a more robust weighted sparse representation method which couples both the locality and linearity structure of the HSI data into a unified framework, to construct the weighted sparse graph for the DR of HSIs.

II. PROPOSED METHOD

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

and p is the number of classes in the data set. $\mathbf{X}^i = [\mathbf{x}_1,...,\mathbf{x}_{i-1},\mathbf{x}_{i+1},...,\mathbf{x}_N]$ denotes all the samples except the *i*-th pixel. $\mathbf{Y} = [\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_N] \in \mathbb{R}^{K\times N}$ represents the low-dimensional features extracted from $\mathbf{X}$, and the corresponding DR projection is **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 **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. Neighborhood Preserving Embedding (NPE)

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 [11], 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 $\mathbf{x}_i$.

ε neighborhood: put an edge between nodes *i* and *j* if $\left\|\mathbf{x}_j - \mathbf{x}_i\right\| \le \varepsilon$.

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:

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

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:

$$\begin{aligned}\mathbf{P}^* &= \arg\min_{\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}} \sum_{\mathbf{i}} \left\| \mathbf{P}^T\mathbf{x}_i - \sum\nolimits_j \mathbf{W}_{i,j}\mathbf{P}^T\mathbf{x}_j \right\|^2 \\ &= \arg\min_{\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}} \mathrm{tr}(\mathbf{P}^T\mathbf{X}\mathbf{L}_s\mathbf{X}^T\mathbf{P})\end{aligned} \tag{2}$$

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 sparse representation (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 [10], [16].

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 $\mathbf{x}_i$ itself, with the sparse representation coefficient vector $\boldsymbol{\alpha}^i$, which can be solved as follows:

$$\arg\min_{\alpha^i} \left\| \boldsymbol{\alpha}^i \right\|_1, \quad s.t.\ \mathbf{x}_i = \mathbf{X}^i\boldsymbol{\alpha}^i, i = 1,2,\ldots N \tag{3}$$

Subsequently, the weighted elements $\mathbf{W}_{i,j}$ can be denoted as:

$$\mathbf{W}_{i,j} = \begin{cases} \boldsymbol{\alpha}^i_j & if\ i > j \\ \boldsymbol{\alpha}^i_{j-1} & if\ i < j \\ 0 & if\ i = j \end{cases} \tag{4}$$

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 [10], 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 the samples within the same class only. As a result, the graph matrix $\mathbf{W}$ obtains the block structure, which is more accurate for the pairwise relations of the pixels.

C. Weighted Sparse Graph Based DR (WSGDR)

The L1-graph explores the linearity structure of the data, and has been widely used in the DR of HSIs. As described in [17], sparse coding results in fewer reconstruction errors thanks to the over-complete dictionary, and is robust to noise. Unfortunately, due to the mechanism of L1-minimization, the sparse coding coefficients can vary a lot, even for similar test samples. The sparse coding process may reconstruct a test pixel with training samples which are far from the test pixel, and thus bring an unstable reconstruction result. From another aspect, the locality information, as pointed out in [11] and [15], is more essential than sparsity under certain conditions, as locality must lead to sparsity, but not necessarily vice versa [17]. In addition, the locality information can ensure that similar pixels will have similar coefficients. In this letter, we propose an improved SR method named weighted sparse graph based DR (WSGDR), which incorporates a locality constraint into the sparse coding constraint to learn the local sparse representation of the test pixels.

Due to the superiority of BSGDA [10], we also describe a block version of weighted sparse graph based DR. Suppose $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_p] \in \mathbb{R}^{B \times N}$, 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:

$$\underset{\alpha_i^j}{arg\,min}\left\|\mathbf{d}^{i,j}\circ\boldsymbol{\alpha}_i^j\right\|_1,\ \ s.t.\ \mathbf{x}_i^j=\mathbf{X}_i^j\boldsymbol{\alpha}_i^j,\ i=1,2,\ldots p,\ j=1,2,\ldots N_i \tag{5}$$

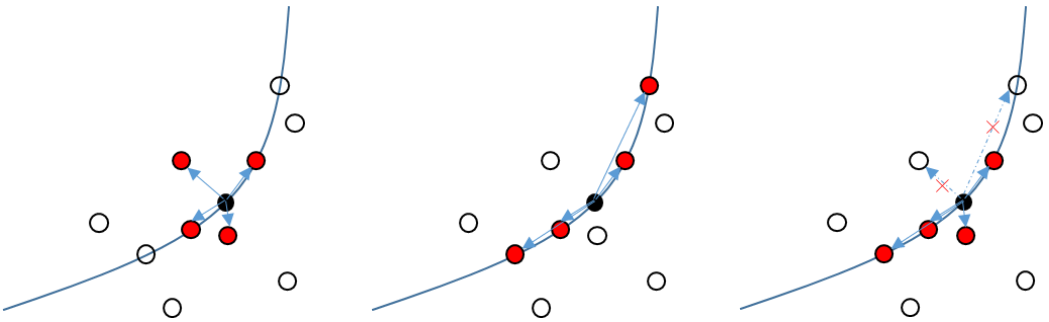
where $\mathbf{X}_i^j$ denotes training sample $\mathbf{X}_i$ expect the *j*-th column $\mathbf{x}_i^j$; the operator $\circ$ means element-wise multiplication; $\boldsymbol{\alpha}_i^j$ represents the SR coefficient vector of $\mathbf{x}_i^j$ with respect to the dictionary base $\mathbf{X}_i^j$; and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

$$\mathbf{d}^{i,j}=\left[\frac{\mathrm{dist}(\mathbf{x}_i^j,\mathbf{x}_i^1)}{\sigma}\cdots\frac{\mathrm{dist}(\mathbf{x}_i^j,\mathbf{x}_i^{N_i-1})}{\sigma}\right]^T \tag{6}$$

where $\mathrm{dist}(\mathbf{x}_i^j,\mathbf{x}_i^k)=\left\|\mathbf{x}_i^j-\mathbf{x}_i^k\right\|$ is the Euclidean distance between $\mathbf{x}_i^j$ and $\mathbf{x}_i^k$, and σ is used for adjusting the tolerance for the locality descriptor. A larger $\mathrm{dist}(\mathbf{x}_i^j,\mathbf{x}_i^k)$ 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 coefficient of the weighted sparse coding tends to integrate the locality and linearity characteristics together. For a test pixel, the weighted sparse graph computes the weight for a training pixel according to the distance or similarity relationship 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 complementary properties (locality and linearity) together to improve the robustness and representation accuracy of the test pixels.

The difference between neighborhood construction by NPE, the L1-graph, and the weighted sparse graph is shown in Fig. 1. NPE is prone to selecting the pixels nearest

to the reference pixel (black dot) to contribute to the representation, as presented in Fig. 1(a). However, for the L1-graph, it is prone to selecting pixels which underlie a low-dimensional subspace to conduct the representation, as presented in Fig. 1(b). In contrast, as shown in Fig. 1(c), the weighted sparse graph has more flexibility in constructing the correct neighborhood through the locality sparse coding scheme (5) to represent the reference pixel.



(a) (b) (c)

Fig. 1. Neighborhood construction by different methods: (a) NPE, (b) the L1-graph, and (c) the weighted sparse graph.

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

$$\mathbf{W}^{i}_{j,k} = \begin{cases} \boldsymbol{\alpha}^{j}_{i,k} & if \ \ j > k \\ \boldsymbol{\alpha}^{j}_{i,k-1} & if \ \ j < k \\ 0 & if \ \ j = k \end{cases} \tag{7}$$

As a result, the graph matrix of all the samples is $\mathbf{W} = diag(\mathbf{W}^{1}, \mathbf{W}^{2}, ..., \mathbf{W}^{p})$. 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 [18] is adopted to solve the weighted sparse coding problem (5).

Algorithm 1: Block version of WSGDR

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

Output: Transformation matrix **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

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

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 λ_k is the *k*-th minimum eigenvalue, and $\mathbf{p}_k$ is the corresponding eigenvector.

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

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[1] was adopted to implement the SVM classifier with a radial basis function (RBF) kernel, and the parameters were selected via cross-validation (CV). Several other DR methods, i.e., local Fisher discriminant analysis (LFDA) [6], NPE [11], SGE [12], and BSGDA [10], were also implemented for comparison. The codes for LFDA[2] and NPE[3] were downloaded online. SGE and BSGDA were implemented using the SPAMS tool[4],

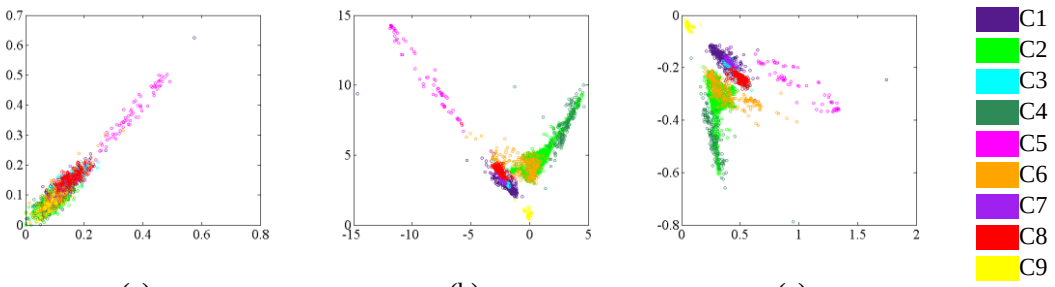
[1]http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[2]http://research.cs.buct.edu.cn/liwei/
[3]http://www.cad.zju.edu.cn/home/dengcai/
[4]http://spams-devel.gforge.inria.fr/

which was also used in the proposed WSGDR. We tested the proposed DR method based on the following two widely used HSIs. All the experiments involved independent Monte Carlo runs, and the average overall accuracy (OA), the average accuracy (AA), the kappa statistic (κ), and the standard deviation are reported.

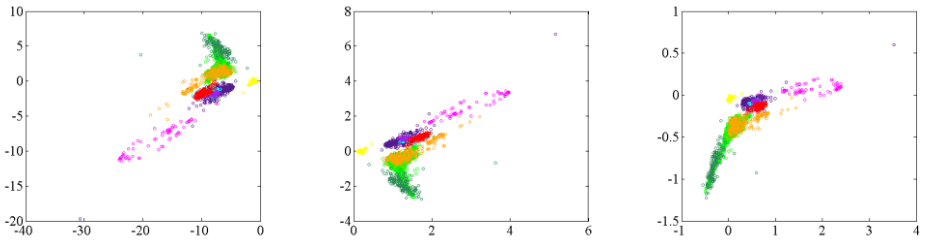
A. Experiments With the University of Pavia Data Set

We first tested the proposed method on the University of Pavia 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 μm. After removing 12 water absorption and noisy bands, 103 bands were used in the experiments. The data set is of 610 × 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 were used for testing.

Fig. 2 illustrates the scatter plots for the different DR methods considering the first two bands or features. As shown in Fig. 2(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.



(a) (b) (c)



(d) (e) (f)

Fig. 2. Scatter plots of the first two bands for the different DR methods on the University of Pavia data set. (a) Original. (b) LFDA [6]. (c) NPE [11]. (d) SGE [12]. (e) BSGDA [10]. (f) The proposed WSGDR.

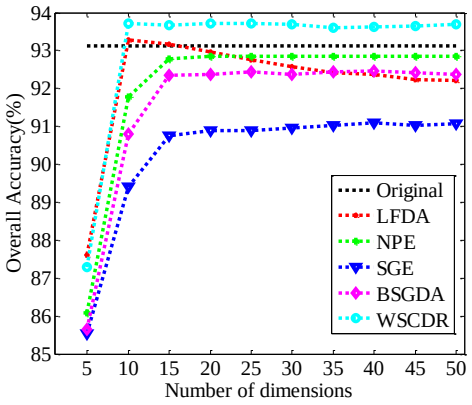


Fig. 3. Overall classification comparison of the different DR methods on the University of Pavia data set with respect to different dimensions.

TABLE I

OA (%), AA (%), Individual Class Accuracy (%), κ, and the Standard Deviation of 10 Conducted Monte Carlo Runs Obtained by the Different DR Methods on the University of Pavia 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	**93.95±0.75**	92.11±0.75	90.11±1.28	92.41±0.74	93.18±0.87
Meadows (C2)	932	17717	**98.29±0.26**	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	**77.37±3.06**
Tree (C4)	153	2911	91.51±1.52	**94.04±0.97**	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	**99.12±0.28**	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	**88.74±1.03**
Bitumen (C7)	67	1263	83.82±2.28	75.65±3.59	83.38±3.43	81.46±2.83	83.25±2.07	**84.64±2.18**
Self-blocking bricks (C8)	184	3498	**90.43±1.21**	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	**99.79±0.13**	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	**93.66±0.21**
AA	--	--	90.63±0.48	89.87±0.61	90.57±0.56	88.19±0.69	90.12±0.48	**91.41±0.39**
κ	--	--	91.10±0.28	91.15±0.33	90.69±0.35	88.15±0.78	90.14±0.48	**91.78±0.27**

Fig. 3 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. 3, the proposed WSGDR outperforms the other methods in almost all the cases. In particular, WSGDR outperforms BSGDA and NPE in all the dimension cases. That is to say, the neighborhood information and linearity property are both important and complementary for the analysis of HSI data. In addition, we also used a reduced dimensionality of K=15, and present the mean OA, AA, individual class accuracy, κ, and standard deviation of 10 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 κ values.

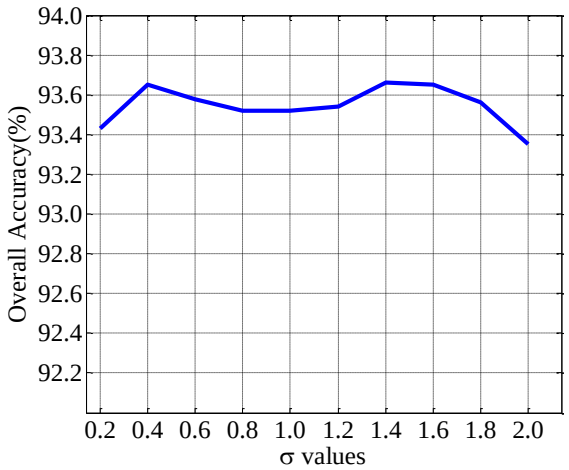


Fig. 4. OA values of the proposed method on the University of Pavia data set with respect to parameter σ .

For the proposed WSGDR, the selection of parameter σ affects the DR performance. Fig. 4 presents the OA values of WSGDR on the University of Pavia data set with respect to σ . From the figure, it can be observed that the result is relatively stable with regard to the value of σ . This inspired us to set σ =1.4 in all the

experiments.

B. Experiments With the AVIRIS Indian Pines Data Set

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

TABLE II

OA (%), AA (%), Individual Class Accuracy (%), κ, and the Standard Deviation of 10 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±14.31	80.00±11.71	73.66±12.32	78.29±7.40	**83.17±7.58**
Corn-notill	143	1285	83.14±2.22	83.32±1.84	78.44±2.87	79.49±2.36	82.88±2.11	**83.56±2.02**
Corn-mintill	83	747	76.63±3.76	69.77±2.87	68.53±4.15	74.39±2.30	72.13±4.58	**78.27±2.77**
Corn	24	213	73.43±9.86	65.63±10.08	70.14±11.42	62.21±9.21	74.37±8.55	**80.89±4.94**
Grass-pasture	48	435	91.33±1.97	90.90±3.27	90.94±2.12	90.39±2.61	92.28±1.99	**93.08±1.94**
Grass-trees	73	657	96.26±3.11	96.88±2.92	94.89±3.18	95.24±3.13	**97.40±2.64**	96.54±2.56
Grass-pasture-mowed	3	25	87.20±3.68	49.2±22.79	**88.80±4.13**	**88.80±4.13**	87.60±4.40	88.00±3.77
Hay-windrowed	48	430	98.88±0.98	**99.53±0.49**	98.74±1.32	98.37±1.53	99.16±0.53	99.53±0.55
Oats	2	18	57.22±21.76	29.44±27.72	48.33±20.63	50.56±24.77	61.11±25.93	**68.89±19.81**
Soybean-notill	97	875	**78.71±3.92**	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	**87.52±1.17**	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	**86.95±3.34**
Wheat	21	184	97.93±1.25	96.68±2.75	96.96±2.13	96.09±2.84	97.88±1.24	**98.75±0.85**
Woods	127	1138	95.15±2.39	95.34±1.26	96.24±1.81	95.98±1.73	**96.50±1.16**	94.59±1.49
Buildings-Grass-Trees-Drives	39	347	59.94±5.80	67.29±3.22	58.33±5.42	53.29±6.48	63.23±5.44	**68.41±5.53**
Stone-Steel-Towers	9	84	87.62±4.02	59.52±12.57	87.02±5.25	**87.86±4.59**	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	**86.20±0.79**
AA	--	--	83.31±1.69	75.04±4.80	80.74±1.01	80.5±1.44	83.36±1.54	**85.46±1.15**
κ	--	--	84.36±1.06	82.25±0.97	81.08±0.48	82.57±1.16	83.97±0.70	**84.81±0.85**

Table I presents the mean OA, AA, individual class accuracy, κ, and the standard

deviation of 10 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).

IV. CONCLUSION

In this paper, we have proposed a novel weighted sparse graph based dimensionality reduction (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 relationship 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 linearity structure of the training pixels, and 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, pp. 3050-3061, Jan. 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, pp. 149-154, Jul. 2014.

[3] Y. Zhou, J. Peng, and C.P. Chen, "Dimension Reduction Using Spatial and Spectral Regularized Local Discriminant Embedding for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.* , vol. 53, pp. 1082-1095, Feb. 2015.

[4] C.-I. Chang, and Q. Du, "Interference and noise-adjusted principal components analysis," *IEEE Trans. Geosci. Remote Sens.* , vol. 37, pp. 2387-2396, Sep. 1999.

[5] 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, pp. 65-74, Jan. 1988.

[6] 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, pp. 1185-1198, Apr. 2012.

[7] 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, pp. 242-256, Jan. 2013.

[8] K. Tan, S. Zhou, and Q. Du, "Semisupervised Discriminant Analysis for Hyperspectral Imagery With Block-Sparse Graph," *IEEE Geosci. Remote Sens. Lett.* , vol. 12, pp. 1765-1769, Aug. 2015.

[9] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.* , vol. 29, pp. 40-51, Jan. 2007.

[10] N.H. Ly, D. Qian, and J.E. Fowler, "Sparse Graph-Based Discriminant Analysis for Hyperspectral Imagery," *IEEE Trans. Geosci. Remote Sens.* , vol. 52, pp. 3872-3884, Jun. 2014.

[11] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2005, pp. 1208-1213.

[12] B. Cheng, J. Yang, S. Yan, Y. Fu, and T.S. Huang, "Learning with-graph for image analysis," *IEEE Trans. Image Process.* , vol. 19, pp. 858-866, Apr. 2010.

[13] 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, pp. 2688-2696, Jun. 2014.

[14] Z. Xue, P. Du, J. Li, and H. Su, "Simultaneous Sparse Graph Embedding for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.* , vol. PP, pp. 1-20, Nov. 2015.

[15] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Advances in neural information processing systems*, 2009, pp. 2223-2231.

[16] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognit.* , vol. 43, pp. 331-341, 2010.

[17] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained Linear Coding for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 3360-3367.

[18] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research*, vol. 11, pp. 19-60, 2010.