Semisupervised Dimensionality Reduction With Pairwise Constraints for Hyperspectral Image Classification

Shiguo Chen and Daoqiang Zhang

***Abstract*—Dimensionality reduction is an important task in the analysis of hyperspectral image data. While traditional dimensionality reduction methods use class labels as prior information, this letter presents a general semisupervised dimensionality reduction framework for hyperspectral image classification based on new prior information, i.e., pairwise constraints which specify whether a pair of examples belongs to the same class or not. The proposed semisupervised dimensionality reduction framework contains two terms: 1) a discrimination term that assesses the separability between classes; and 2) a regularization term that characterizes some property of the original data set. Furthermore, a novel semisupervised dimensionality reduction method is derived from the framework based on sparse representation. Experimental results on two hyperspectral image data sets show that the proposed algorithms are remarkably effective in comparison to traditional dimensionality reduction methods.**

***Index Terms*—Dimensionality reduction, hyperspectral image classification, pairwise constraints, sparse representation.**

I. Introduction

HYPERSPECTRAL sensors have been developed and widely used for observing the Earth's surface by sampling a huge number of spectral bands, typically up to several hundreds. This unreasonably large number of spectral bands implies high dimensionality of hyperspectral image and causes several challenges to image classification. As a result, it is advantageous to reduce the number of original dimensions without sacrificing significant information, i.e., dimensionality reduction. Existing dimensionality reduction methods can be roughly categorized into supervised and unsupervised ones according to whether they use supervision information or not. Linear discriminant analysis [1] and nonparametric weighted feature extraction (NWFE) [2] are two widely used supervised dimensionality reduction methods for hyperspectral image data. In recent years, many extensions on those two methods have been proposed, such as modified Fisher's linear discriminant analysis [3], regularized linear discriminant analysis [4], cosine-based nonparametric feature extraction [5], and improved NWFE based on support vector machine and support vector domain description [6]. While the supervised dimensionality reduction methods use class labels to guide the process of dimensionality reduction, unsupervised dimensionality reduction methods do not use any supervision information. One of the most popular unsupervised dimensionality reduction methods is principal component analysis (PCA) [7], which maximizes data global variance by orthogonal projection. A refinement of PCA is independent component analysis (ICA) [8], which uses higher order statistics. In competition to PCA and ICA, Phillips *et al.* [9] and He and Mei [10] used singular value decomposition and random projection, respectively, to reduce the dimensions of hyperspectral image data. More recently, lower rank tensor approximation [11] and minimum change rate deviation [12] are proposed for hyperspectral image data by taking into account the spatial relation among neighboring image pixels.

In many practical applications, unlabeled training examples are readily available but labeled ones are fairly expensive to obtain, and therefore, semisupervised learning, which takes into account both labeled and unlabeled samples in the learning of the classifier, has attracted much attention recently [13]. In general, there are different forms of supervision information or prior knowledge, such as class labels, pairwise constraints, and others. In hyperspectral image data analysis, class labels have been widely used for classification and dimensionality reduction. However, to the best of our knowledge, the other forms of prior knowledge have rarely been investigated for dimensionality reduction, except that [14] uses the U.S. Geological Survey digital spectral libraries to estimate the parameter of the algorithm. In this letter, we focus on prior knowledge in the form of pairwise constraints, which can be defined as follows: Pairs of samples known as belonging to the same class are called must-link constraints and the ones belonging to different classes are called cannot-link constraints [15]. Pairwise constraints arise naturally in many real tasks. For example, in hyperspectral image data analysis, the true labels may not be known *a priori*, while it could be easier for a user to specify whether some pairs of instances belong to the same class or not.

In this letter, we first propose a general semisupervised dimensionality reduction framework for classification of hyperspectral image data based on pairwise constraints, which contains two terms: 1) a *discrimination term*; and 2) a *regularization term*. Then, the idea of sparse representation [16], [17] is introduced into our framework that will lead to novel semisupervised dimensionality reduction methods based on sparsity preserving (SSDRsp). Finally, the experimental results on two hyperspectral image data sets, i.e., Indian Pines 92AV3C

Manuscript received April 19, 2010; revised June 6, 2010 and July 26, 2010; accepted August 30, 2010. Date of publication October 25, 2010; date of current version February 25, 2011. This work was supported by the National Natural Science Foundation of China under Grant 60875030.

The authors are with the Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: sgchen@nuaa.edu.cn; dqzhang@nuaa.edu.cn).

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.2010.2076407

data set and Washington DC Mall data set [18], validate the effectiveness of the proposed methods.

II. Proposed Methodology

Let us consider the problem of semisupervised dimensionality reduction. Given a set of data samples $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, together with some must-link constraints set $(\mathbf{M})$ and cannot-link constraints set $(\mathbf{C})$, then the problem is to find a set of projective vectors $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d\}$, such that the embedding low-dimensional representations $\mathbf{y}_i = \mathbf{w}_i^{\mathrm{T}}\mathbf{x}_i$ can preserve the structure of the original data set as well as the constraints in $\mathbf{M}$ and $\mathbf{C}$ [15]. The original data samples are all D-dimensional vectors, and the low-dimensional representations are d-dimensional $(D > d)$ vectors.

A. Semisupervised Dimensionality Reduction Framework

The semisupervised dimensionality reduction framework contains two terms: a *discrimination term* and a *regularization term*. The global objective function is made up of the two terms.

1) *Discrimination term*—This term is based on pairwise constraints. Here, the intuition is to let the distance between instances involved by the cannot-link constraints in $\mathbf{C}$ as large as possible, while the distances between instances involved by the must-link constraints in $\mathbf{M}$ as small as possible in the embedding low-dimensional space. As a result, we define the objective function of *discrimination term* as follows:

$$\boldsymbol{J_D}(\mathbf{w}) = \frac{\mathbf{1}}{\boldsymbol{2n_C}} \sum_{(\boldsymbol{i},\boldsymbol{j})\in\mathbf{C}} \left(\mathbf{w}^{\mathrm{T}}\mathbf{x_i} - \mathbf{w}^{\mathrm{T}}\mathbf{x_j}\right)^{\mathbf{2}} - \frac{\alpha}{\boldsymbol{2n_M}} \sum_{(\boldsymbol{i},\boldsymbol{j})\in\boldsymbol{M}} \left(\mathbf{w}^{\mathrm{T}}\mathbf{x_i} - \mathbf{w}^{\mathrm{T}}\mathbf{x_j}\right)^{\mathbf{2}}. \quad (1)$$

Here, $\mathbf{x}_i$ and $\mathbf{x}_j$ are the pairs of samples belonging to must-link constraints or cannot-link constraints, $\boldsymbol{n}_C$ is the number of cannot-link constraints, and similarly, $\boldsymbol{n}_M$ is the number of must-link constraints. The scaling parameter α is used here to balance the contribution of the must-link constraints. There exists a concise form for (1)

$$\boldsymbol{J_D}(\mathbf{w}) = \frac{\mathbf{1}}{\mathbf{2}} \sum_{i,j} \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \mathbf{w}^{\mathrm{T}}\mathbf{x}_j\right)^{\mathbf{2}} \mathbf{A}_{ij} \quad (2)$$

where the coefficient matrix $\mathbf{A}$ is defined as

$$\mathbf{A}_{ij} = \begin{cases} \frac{\mathbf{1}}{\boldsymbol{n}_C} & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{C} \\ -\frac{\alpha}{\boldsymbol{n}_M} & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{M} \\ \mathbf{0} & \text{otherwise.} \end{cases} \quad (3)$$

If we denote $\boldsymbol{\Delta}$ as a diagonal matrix whose nonzero elements are column sums of $\mathbf{A}$, i.e.,$\boldsymbol{\Delta_{ii}} = \sum_{\boldsymbol{j}} \mathbf{A_{ij}}$, thus (2) can be simplified as

$$\boldsymbol{J_D}(\mathbf{W}) = \mathbf{w}^{\mathrm{T}}\mathbf{X}(\Delta - \mathbf{A})\mathbf{X}^{\mathrm{T}}\mathbf{w}. \quad (4)$$

2) *Regularization term*—This term is to characterize some property of the original data set, including abundant unlabeled data. The motivation of using this term is to use the unlabeled data to enhance the performance when constraints are few. Since some property of abundant unlabeled data is included, it is expected to be more stable than using only the constraints. In order to use the original data, we will define two different forms for the *regularization term*. In Section II-B, maximizing data global variance (PCA criterion) will be used that would lead to a global structure preserving semisupervised dimensionality reduction, which is equivalent to the method in [15]. In Section II-C, a sparse reconstruction weight matrix will be calculated by sparse representation, and then, the *regularization term* can be defined based on the idea of minimizing the reconstruction errors in the low-dimensional space, which lead to SSDRsp.

In order to find the optimal projective vectors $\mathbf{w}$'s, we defined the global objective function as

$$\boldsymbol{J} = \boldsymbol{J}_D + \beta \bullet \boldsymbol{J}_R \quad (5)$$

where $\boldsymbol{J}_R$ is some definition of the *regularization term* and parameter β is added to tune the tradeoff between *discrimination term* $\boldsymbol{J}_D$ and *regularization term* $\boldsymbol{J}_R$. The final objective function $\boldsymbol{J}$ is to preserve the some property of the original data set as well as the pairwise constraints in sets $\mathbf{M}$ and $\mathbf{C}$. In the following two sections (Sections II-B and C), two semisupervised dimensionality reduction will be introduced according to our basic framework.

B. SSDRpca

The method in [15] (named global structure preserving semisupervised dimensionality reduction (SSDRpca) here) is one of the most widely used semisupervised dimensionality reduction methods. The goal of SSDRpca is to preserve the global covariance of the original data set as well as the pairwise constraints. The objective of SSDRpca is defined as maximizing $\boldsymbol{J_{SSDRpca}}(\mathbf{w})$ w.r.t. $\mathbf{w}^{\boldsymbol{T}}\mathbf{w} = 1$, where $\boldsymbol{T}$ denotes the transpose operator and $\boldsymbol{J_{SSDRpca}}(\mathbf{w})$ can be written as follows:

$$\boldsymbol{J_{SSDRpca}}(\mathbf{w}) = \frac{\mathbf{1}}{\boldsymbol{2n}_C} \sum_{(i,j)\in C} \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_j\right)^{\mathbf{2}} - \frac{\alpha}{\boldsymbol{2n}_M} \times \sum_{(i,j)\in M} \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \mathbf{w}^{\mathrm{T}}\mathbf{x}_j\right)^{\mathbf{2}} + \frac{\beta}{\boldsymbol{2N^2}} \times \sum_{i,j} \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \mathbf{w}^{\mathrm{T}}\mathbf{x}_j\right)^{\mathbf{2}}. \quad (6)$$

The last term of (6) is equivalent to the PCA criterion with some simple algebraic derivation. As shown in [15], the objective of SSDRpca is efficient and has a closed-form solution of an eigen problem.

In terms of our semisupervised dimensionality reduction framework proposed in Section II-A, the first two terms of (6) is equivalent to the *discrimination term*, and the last term is equivalent to the *regularization term* using PCA criterion. Thus, SSDRpca can be seen as a special case of our semisupervised dimensionality reduction framework.

C. SSDRsp

Sparse representation is initially proposed as an extension to traditional signal representations such as Fourier representation and wavelet representation [17]. Additionally, in mathematical form, sparse representation can be briefly expressed as follows [16], [17]:

$$\min_{s_i} \|\mathbf{s}_i\|_1 \quad \boldsymbol{s.t.} \quad \mathbf{x}_i = \mathbf{X}\mathbf{s}_i \quad \text{and} \quad 1 = \mathbf{1}^{\mathrm{T}}\mathbf{s}_i \tag{7}$$

where $\mathbf{1} \in \mathbf{R}^{\boldsymbol{N}}$ is a vector of all ones. $\boldsymbol{s}_i$ is the sparse reconstructive weight vector for $\mathbf{x}_i$, and $\mathbf{s}_i = \{\mathbf{s}_{ij}\}$, where the elements $\mathbf{s}_{ij}(j \neq i)$ summarize the contribution of each $\mathbf{x}_j$ to reconstructing $\mathbf{x}_i$. All the sparse reconstructive weight vectors compose a reconstructive weight matrix $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N\}$. The sparse weight matrix $\mathbf{S}$ can reflect the intrinsic geometric properties of the data to some extent and naturally preserve potential discriminant information [17]. We therefore expect that the characterization in the original high-dimensional embedding space can be preserved in the low-dimensional embedding space. In particular, the same weights $\mathbf{s}_{ij}$ that reconstruct the ith data in high dimensions should also reconstruct its low-dimensional representations. Here, the *regularization term* $\boldsymbol{J}_R$ can be defined as follows:

$$\boldsymbol{J}_{\boldsymbol{R}}(\mathbf{w}) = -\frac{\mathbf{1}}{N}\left(\sum_i \|\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \mathbf{w}^{\mathrm{T}}\mathbf{X}\mathbf{s}_i\|^2\right). \tag{8}$$

Equation (8) is equivalent to the objective function of sparsity preserving projections (SPP) [17]. With some simple algebraic derivation, (8) can be rewritten into

$$\begin{aligned}\boldsymbol{J}_{\boldsymbol{R}}(\mathbf{W}) &= \mathbf{w}^{\mathrm{T}}\mathbf{X}\left(\frac{\mathbf{1}}{\boldsymbol{N}}(\mathbf{S}^{\mathrm{T}} + \mathbf{S} - \mathbf{S}^{\mathrm{T}}\mathbf{S} - \mathbf{I})\right)\mathbf{X}^{\mathrm{T}}\mathbf{w} \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{X}\boldsymbol{\Omega}\mathbf{X}^{\mathrm{T}}\mathbf{w}\end{aligned} \tag{9}$$

where $\mathbf{S}$ is the reconstructive weight matrix and $\mathbf{I}$ is the identity matrix. According to (4) and (9), the objective function of SSDRsp can be defined as maximizing $\boldsymbol{J}_{\boldsymbol{SSDRsp}}(\mathbf{w})$ w.r.t.$\mathbf{w}^{\boldsymbol{T}}\mathbf{w} = 1$, and $\boldsymbol{J}_{\boldsymbol{SSDRsp}}(\mathbf{w})$ can be written as follows:

$$\begin{aligned}\boldsymbol{J}_{\boldsymbol{SSDRsp}}(\mathbf{w}) &= \boldsymbol{J}_D(\mathbf{w}) + \beta \bullet \boldsymbol{J}_R(\mathbf{w}) \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{X}(\Delta - \mathbf{A} + \beta\boldsymbol{\Omega})\mathbf{X}^{\mathrm{T}}\mathbf{w} \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{X}\tilde{\mathbf{L}}\mathbf{X}^{\mathrm{T}}\mathbf{w}\end{aligned} \tag{10}$$

where $\tilde{\mathbf{L}} = \boldsymbol{\Delta} - \mathbf{A} + \beta\boldsymbol{\Omega}$. Then, from (10), the objective function of SSDRsp can be expressed to be a typical eigen problem. Then, the optimal $\mathbf{w}$'s are the eigenvectors corresponding to the largest d eigenvalues. The pseudocode for the algorithm is shown in Fig. 1.

Algorithm: SSDRsp
Input: Set of data points $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N\}$;
Set of must-link constraints $\mathbf{M}$;
Set of cannot-link constraints $\mathbf{C}$.
Output: projective vectors $\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_d\}$
Method:
1. Initialize parameters: α, β, $n_M=|\mathbf{M}|$, $n_C=|\mathbf{C}|$;
2. Calculate the coefficient matrix using Eq. (3);
3. Construct the reconstructive weight matrix $\mathbf{S}$ using Eq. (7);
4. Obtain the projective vectors $\mathbf{w}$'s from maximizing $\boldsymbol{J}_{\boldsymbol{SSDRsp}}(\mathbf{w})$ in Eq. (10).

Fig. 1. SSDRsp algorithm.

III. Experimental Results

A. Data Sets and Experimental Setting

Experiments were conducted on two hyperspectral image data sets: Indian Pines 92AV3C and Washington DC Mall [18]. For Indian Pines 92AV3C, we chose two sets of all 16 classes: one for training and the other for testing, respectively, sampling 1398 instances for training and 1401 instances for testing from the original data set randomly. Similar methodology was used for Washington DC Mall, but 2107 instances were selected for training and 3045 instances were selected for testing. Table I summarizes the properties of the data sets: the number of samples (#samples), the number of spectral bands (#bands), and the number of classes (#classes).

TABLE I
Data Sets Used in the Experiments

properties	Indian Pines 92AV3C		Washington DC Mall	
	Training	Testing	Training	Testing
#samples	1398	1401	2107	3045
#bands	220	220	191	191
#classes	16	16	7	7

TABLE II
Classification Accuracies (%) on Indian Pines 92AV3C With Different Constraints ($\#C$) and Fixed Number of Features (15)

#C	*Classifier*	*PCA*	*NWFE*	*SPP*	*JD*	*SSDRpca*	*SSDRsp*
10	NN	21.91	50.75	47.11	68.77	**69.30**	68.02
	SVM	14.99	48.75	57.89	72.45	**73.22**	72.97
	QDC	3.43	43.61	48.18	59.91	**62.66**	59.41
100	NN	21.91	50.75	47.11	69.49	69.59	**72.42**
	SVM	14.99	48.75	57.89	73.10	73.73	**77.19**
	QDC	3.43	43.61	48.18	62.71	66.98	**67.85**

TABLE III
Classification Accuracies (%) on Washington DC Mall With Different Constraints ($\#C$) and Fixed Number of Features (15)

#C	*Classifier*	*PCA*	*NWFE*	*SPP*	*JD*	*SSDRpca*	*SSDRsp*
10	NN	72.46	**97.51**	79.99	96.25	96.54	96.79
	SVM	7.14	94.50	92.60	**94.65**	94.42	93.45
	QDC	52.42	**98.20**	92.96	96.51	96.87	97.19
100	NN	72.46	**97.51**	79.99	97.01	96.52	97.20
	SVM	7.14	94.50	92.60	94.80	94.55	**97.10**
	QDC	52.42	**98.20**	92.96	97.19	96.86	97.36

Since the method proposed in this letter is used for classification of hyperspectral image data, we therefore use classification accuracy to evaluate the dimensionality reduction results. Three widely used classifiers were used here: nearest neighbor classifier (NN), quadratic discriminant classifier (QDC), and support vector machines (SVM). The methods compared in the experiments include the following:

1) **PCA**[7]: principal component analysis;
2) **NWFE**[2]: nonparametric weighted feature extraction;
3) **SPP**[17], which uses only the *regularization term* $\boldsymbol{J}_R$ described in Section II-C;
4) **JD**, which uses only the *discrimination term* $\boldsymbol{J}_D$ described in Section II-A;

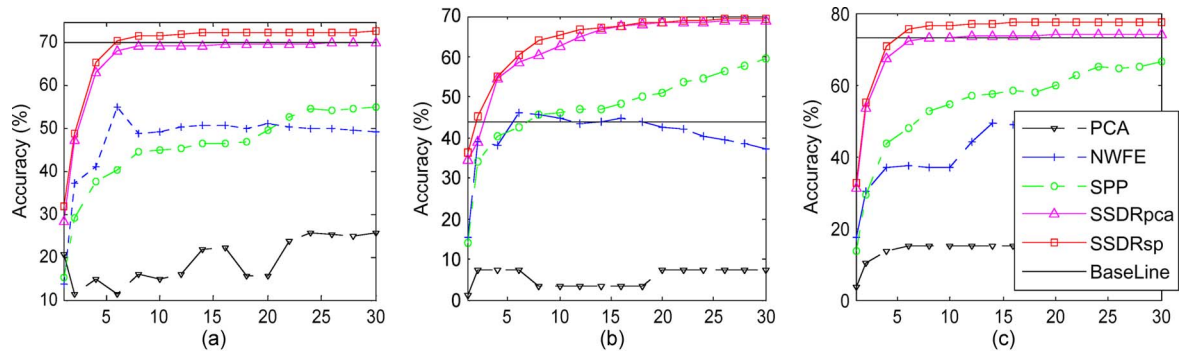

Fig. 2. Accuracies (%) on Indian Pines 92AV3C when the number of constraints is 100. (a) Results of NN. (b) Results of QDC. (c) Results of SVM.

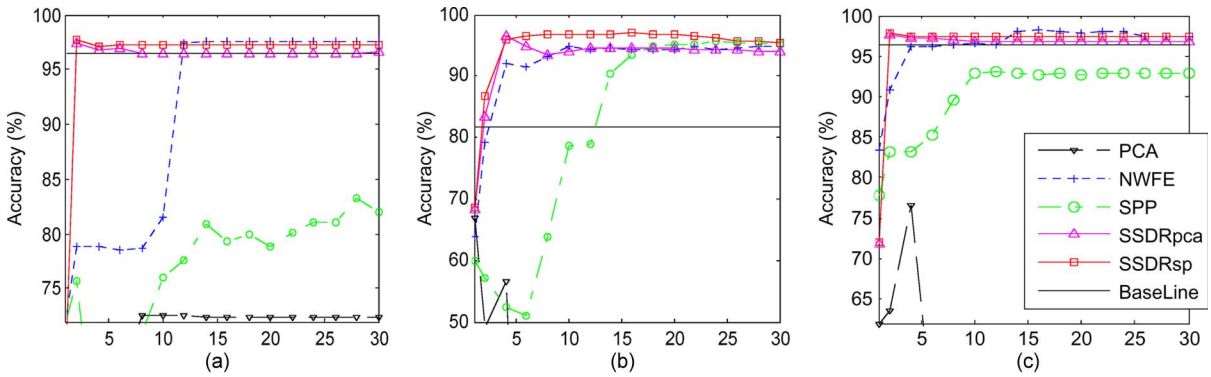

Fig. 3. Accuracies (%) on Washington DC mall when the number of constraints is 100. (a)Results of NN. (b) Results of QDC. (c) Results of SVM.

5) **SSDRpca** [15], which uses both the *discrimination term* $\boldsymbol{J}_D$ and the *regularization term* $\boldsymbol{J}_R$ described in Section II-B;
6) **SSDRsp**, which uses both the *discrimination term* $\boldsymbol{J}_D$ and the *regularization term* $\boldsymbol{J}_R$ described in Section II-C;
7) **Baseline**: the accuracy on the original data.

We selected the same number of must-link constraints and cannot-link constraints from the training sets randomly according to the ground truth in our experiments. The same constraints were used for JD, SSDRpca, and SSDRsp in each run, and the results were averaged over ten runs. In the experiments, the value of α was always fixed to one, and the value of β was searched in the range of [0, 3] based on the standard grid search, which is an exhaustive search method to determine the best value over a supplied range of parameter values with a user-defined step size, i.e., 0.2 in our experiments. The reconstructive weight matrix $\mathbf{S}$ in Section II-C of SSDRsp was calculated by the sparse learning package (SLEP)[1]—a software that can solve the L1-regularized sparse learning problem [(7)] efficiently. The dimensionality reduction methods were run on the training data sets, but the accuracies were calculated only on the testing data sets.

B. Results and Discussion

We first compared all the dimensionality reduction methods when the number of features was fixed at 15. For the three methods (JD, SSDRpca, and SSDRsp) derived from the semisupervised dimensionality reduction framework, we displayed their results with different numbers of constraints: ten little constraints and 100 lots of constraints independently. Tables II and III show the results on two hyperspectral image data sets, respectively. The symbol $\#C$ in Tables II and III denotes the number of constraints (where $\boldsymbol{n}_M = \boldsymbol{n}_C = \#C$), and the symbol *Classifier* denotes the classifiers used in experiments. From the results of the two tables, the two semisupervised methods are superior to the unsupervised methods. For the Indian Pines 92AV3C (Table II), both SSDRpca and SSDRsp are all superior to JD that uses only the *discrimination term*, which shows that the *regularization term* can improve the performance of dimensionality reduction. However, for two semisupervised methods (SSDRpca and SSDRsp) on Washington DC mall (Table III), only SSDRsp is superior to JD, which shows that the methods proposed in this letter are more stable than SSDRpca.

In the second set of experiments, we evaluated the performance of all dimensionality reduction methods when different numbers of features were used. Figs. 2 and 3 demonstrate the results of three classifiers on two hyperspectral image data sets, where the number of constraints is 100 for JD, SSDRpca, and SSDRsp. It can be seen from Figs. 2 and 3 that SSDRsp is superior to other methods in nearly all cases. Furthermore, Figs. 2 and 3 indicate that SSDRsp and SSDRpca still work well when only a few features are used (about ten features for Indian Pines 92AV3C and four features for Washington DC mall according to the curves) in comparison to other approaches.

In the next experiments, we study the sensitivity of SSDRsp with the parameter β for all three classifiers (NN, QDC, and SVM). Fig. 4 shows the changing curves when the value of β

[1] www.public.asu.edu/~jye02/Software/SLEP/

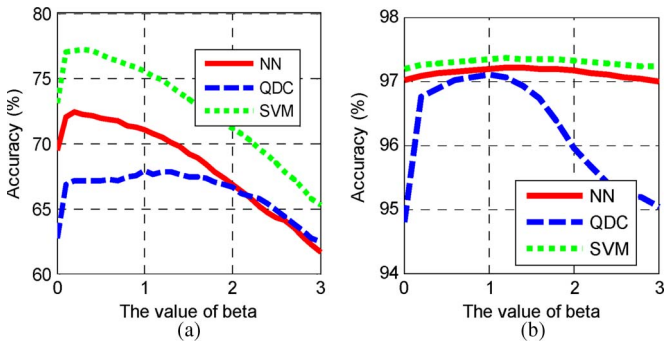

Fig. 4. Sensitivity analysis of β on different data sets. (a) Indian Pines 92AV3C. (b) Washington DC mall.

TABLE IV
COMPUTATIONAL COST (THE UNIT OF MEASUREMENT IS SECOND) OF DIFFERENT APPROACHES. *Set1* DENOTES INDIAN PINES 92AV3C DATA SET, AND *Set2* DENOTES WASHINGTON DC MALL DATA SET

Dataset	*PCA*	*NWFE*	*SPP*	*SSDRpca*	*SSDRsp*
Set1	0.062s	663.6s	50.49s	94.7s	177.4s
Set2	0.047s	1803s	84.17s	194.1s	359.4s

was varied from zero to three. The other parameter α was not studied here because it is fixed to be one in our experiments following [15]. According to our framework in Section II, the value of β is used to balance the contribution of *regularization term*. Particularly, when the value of β is zero, SSDRsp has not balance the *regularization term* and only utilizes the pairwise constraints. From the results in Fig. 4, all curves have the similar changing tendency that they rise quickly at first and then begin to descend, and the peak value arises always with a small value of β. According to this observation, we advise that the value of β can be selected within a limited range near zero, i.e., zero to three in our experiments.

At last, we compared the computational cost (related to the process of data extraction only) of different approaches on two hyperspectral image data sets. We carried out the experiments on an Intel (R) Xeon (R) CPU E5410 at 2.33 GHz processor, and the time consumed on parameter determination is also included here for the computational cost of β determination in SSDRpca and SSDRsp methods. From the results in Table IV, PCA and SPP are faster than SSDRpca and SSDRsp, respectively, and SSDRpca is more efficient than SSDRsp. The reason for the former is because β determination in SSDRpca and SSDRsp is time consuming, and the reason for the latter is because constructing the weight matrix in SSDRsp is time consuming. It's noteworthy that both SSDRpca and SSDRsp are much more efficient than NWFE.

IV. CONCLUSION AND FUTURE WORK

This letter proposed a general semisupervised dimensionality reduction framework based on pairwise constraints for hyperspectral image classification. It contains two terms: a *discrimination term* and a *regularization term*. From the framework, we also developed a novel semisupervised dimensionality reduction method called SSDRsp based on sparse representation. Experimental results on two well-known hyperspectral image data sets show the effectiveness of our proposed method for hyperspectral image classification.

In the future, we will further evaluate our methods on more hyperspectral image data sets, e.g., NASA John F. Kennedy Space Center and the Okavango Delta, Botswana data sets, etc. Other interesting topics for future research include nonlinear extension of the proposed semisupervised dimensionality reduction framework, selection of most informative pairwise constraints, and definition of novel *regularization term* for the semisupervised framework.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions. The authors would also like to thank D. Landgrebe for providing the hyperspectral image data sets.

REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley Interscience, 2001.
[2] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.
[3] Q. Du, "Modified Fisher's linear discriminant analysis for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 503–507, Oct. 2007.
[4] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.
[5] J.-M. Yang, P.-T. Yu, and B.-C. Kuo, "A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1279–1293, Mar. 2010.
[6] C.-H. Li, B.-C. Kuo, C.-T. Lin, and C.-C. Hung, "Dimension reduction for hyperspectral image classification via support vector based feature extraction," in *Proc. IEEE IGARSS*, Boston, MA, Jul. 2008, pp. V-389–V-392.
[7] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, pp. 417–441, 1933.
[8] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.
[9] R. D. Phillips, W. T. Watson, R. H. Wynne, and C. E. Blinn, "Feature reduction using a singular value decomposition for the iterative guided spectral class rejection hybrid classifier," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 1, pp. 107–116, Jan. 2009.
[10] M. He and S. Mei, "Dimension reduction by random projection for endmember extraction," in *Proc. IEEE Conf. Ind. Electron. Appl.*, Jun. 15–17, 2010, pp. 2323–2327.
[11] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.
[12] R. Dianat and S. Kasaei, "Dimension reduction of optical remote sensing images via minimum change rate deviation method," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 198–206, Jan. 2010.
[13] X. Zhu, "Semi-supervised learning literature survey," Dept. Comput. Sci., Univ. Wisconsin Madison, Madison, WI, Comput. Sci. TR 1530, Jul. 19, 2008.
[14] B. Guo, S. R. Gunn, R. I. Damper, and J. B. D. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.
[15] D. Zhang, Z. Zhou, and S. Chen, "Semi-supervised dimensionality reduction," in *Proc. 7th SIAM Int. Conf. Data Mining (SDM)*, Minneapolis, MN, Apr. 2007, pp. 629–634.
[16] J. Wright, A. Yang, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[17] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognit.*, vol. 43, no. 1, pp. 331–341, Jan. 2010.
[18] [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/documentation.html