A Modified Locality-Preserving Projection Approach for Hyperspectral Image Classification

Yongguang Zhai, Lifu Zhang, *Senior Member, IEEE*, Nan Wang, Yi Guo, Yi Cen, Taixia Wu, and Qingxi Tong

***Abstract*—Locality-preserving projection (LPP) is a typical manifold-based dimensionality reduction (DR) method, which has been successfully applied to some pattern recognition tasks. However, LPP depends on an underlying adjacency graph, which has several problems when it is applied to hyperspectral image (HSI) processing. The adjacency graph is artificially created in advance, which may not be suitable for the following DR and classification. It is also difficult to determine an appropriate neighborhood size in graph construction. Additionally, only the information of local neighboring data points is considered in LPP, which is limited for improving classification accuracy. To address these problems, a modified version of the original LPP called MLPP is proposed for hyperspectral remote-sensing image classification. The idea is to select a different number of nearest neighbors for each data point adaptively and to focus on maximizing the distance between nonnearest neighboring points. This not only preserves the intrinsic geometric structure of the data but also increases the separability among ground objects with different spectral characteristics. Moreover, MLPP does not depend on any parameters or prior knowledge. Experiments on two real HSIs from different sensors demonstrate that MLPP is remarkably superior to other conventional DR methods in enhancing classification performance.**

***Index Terms*—Dimensionality reduction (DR), hyperspectral remote sensing, image classification, unsupervised learning.**

I. Introduction

HYPERSPECTRAL images (HSIs) provide a large amount of spectral information for discriminating subtle differences among ground objects [1]. However, the fact that adjacent bands are highly correlated also brings redundancy [2]. Furthermore, the number of spectral bands in HSI is usually quite large, which is challenging for conventional classification techniques due to the curse of dimensionality, i.e., the Hughes phenomenon [3]. Therefore, reducing the dimensionality of hyperspectral data without losing valuable information about objects of interest is an important issue in remote-sensing data analysis [4]. Numerous dimensionality reduction (DR) approaches for hyperspectral data classification have been proposed. They can be divided into three types: supervised, semisupervised, and unsupervised methods. Supervised methods depend on labeled samples to increase the separability among different classes, such as classic linear discriminant analysis (LDA) [5] and nonparametric weighted feature extraction (NWFE) [6]. Semisupervised methods, which aim to improve the classification accuracy by using both limited labeled and unlabeled data, have been developed for the case where only a fewer labeled samples are available [7]–[9]. In practical applications, labeled data are usually limited, and labeling large quantities of data is often very expensive [10], [11]. Thus, unsupervised idea plays a very important role in DR. Unsupervised approaches project the original data to a lower dimensional space without label information, where data points with similar spectral features stay close while very different data points are far apart. Principal component analysis (PCA) is a typical unsupervised DR widely used for feature extraction [12]. PCA is based on the simple linear model, so it has difficulties in handling nonlinearity. However, nonlinearity does exist in the hyperspectral remote-sensing image due to multiscattering between targets and sensors [13]. Therefore, nonlinear DR methods have gained much attention. Locally linear embedding [14] and Laplacian eigenmap (LE) [15] are commonly used nonlinear DR methods. Their extensions such as neighborhood preserving embedding (NPE) [16] and locality-preserving projection (LPP) [17] have been widely used in some pattern recognition tasks. LPP builds a graph exploring local geometric structures and defines a linear projection from the original space to the low-dimensional space. This linear mapping is incorporated into LE, leading to a simple linear system which can be easily solved by eigendecomposition. Several extensions [18]–[21] to LPP have been proposed in recent years.

However, three problems must be addressed in LPP before it can be applied to HSI classification. First, in the process of DR, the neighborhood graph is artificially created in advance and thus does not necessarily benefit the following DR and classification. Second, it only considers the similarities among neighboring data points, while dissimilarities among data points that are further apart are ignored. However, incorporating dissimilarities among far apart data points will certainly improve the separability of classes, i.e., ground objects with different spectral characteristics in HSI. Third, it is not clear how to

Manuscript received January 24, 2016; revised April 20, 2016; accepted April 28, 2016. Date of publication May 25, 2016; date of current version July 20, 2016. This work was supported in part by the Natural Science Foundation of China through Projects 41371362, 41501391, and 41272364 and in part by the Major Special Project—the China High-Resolution Earth Observation System project and the National High-Tech R&D Program of China (863 Program) under Grant 2013AA12A302.

Y. Zhai is with the State Key Laboratory of Remote Sensing Sciences, Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, Beijing 100101, China, and also with the College of Water Conservancy and Civil Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China (e-mail: zhaiyg@radi.ac.cn).

L. Zhang, N. Wang, Y. Cen, T. Wu, and Q. Tong are with the State Key Laboratory of Remote Sensing Sciences, Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, Beijing 100101, China (e-mail: zhanglf@radi.ac.cn; wangnan@radi.ac.cn; cenyi@radi.ac.cn; wutx@radi.ac.cn; tqxi@263.net).

Y. Guo is with the School of Computing, Engineering and Mathematics Parramatta South Campus Penrith South, Western Sydney University, Richmond, N.S.W. 2751, Australia (e-mail: y.guo@westernsydney.edu.au).

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.2016.2564993

1545-598X © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

determine the value of one of the most important parameters in LPP, the number of nearest neighbors used to construct the adjacency graph [22].

In this letter, a weighted graph is proposed, and a discriminant criterion is incorporated in the original LPP to solve the aforementioned problems for HSI classification. The goal is to find a projection that can increase the class separability by maximizing the distances between points that are not similar. Moreover, instead of fixing the number of nearest neighbors in LPP's adjacency graph construction for all data points, the modified LPP (MLPP) chooses the number of neighbors for each point adaptively.

This letter is organized as follows. The LPP and MLPP are introduced in Section II. Section III presents experimental results to show its effectiveness. Finally, some conclusions are drawn in Section IV.

II. Methodology

A. LPP

Given a data set $X = [x_1, x_2, \ldots, x_n]$ and $x_i \in R^d (i = 1, \ldots, n)$, where n is the number of sample points and d is the dimensionality of the data, the purpose of DR is to find a transformation f from R^d to R^t such that $y_i = f(x_i)$, $y_i \in R^t$, and $t \ll d$. If it is a linear mapping, then a matrix $A \in R^{d\times t}$ is usually obtained such that $y_i = A^T x_i$ [17].

In LPP, an affinity graph W is built exploiting the local neighborhood structure of the data set. By using W, LPP seeks a low-dimensional subspace where the local neighborhood information contained in X can be preserved. The objective function of LPP in (1) is to ensure that, if x_i and x_j are "close," then the corresponding y_i and y_j are also "close" as well

$$\min \sum_{ij} \|y_i - y_j\|^2 W_{ij} = \min \sum_{ij} \|A^T x_i - A^T x_j\|^2 W_{ij} \tag{1}$$

where W_{ij} is the *affinity* between x_i and x_j. If x_i is one of the nearest neighbors of x_j and vice versa, then

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{s}} \tag{2}$$

where parameter s is a prespecified constant. For simplicity, W_{ij} can also be set as 1 when x_i and x_j are nearest neighbors and 0 otherwise. Minimizing the objective function given in (1), it can be converted to a generalized eigendecomposition problem

$$X^T L X a_i = \lambda_i X^T D X a_i \tag{3}$$

where $L = D - W$ is the so-called Laplacian matrix, D is a diagonal matrix whose jth diagonal entry is $\Sigma_j W_{ij}$, and λ_i is the ith smallest eigenvalue corresponding to a_i for $i = 1, \ldots, t$.

B. Modification of LPP

LPP attempts to preserve the locality of data. However, for HSIs, LPP often fails. This is caused by the fact that the number of nearest neighbors in LPP's affinity graph construction is fixed for all data points without taking into account that the number of data points in each class can vary greatly. Particularly for a data point from small size classes, its neighborhood may contain data points from different classes. Moreover, how to determine the number of nearest neighbors is a crucial issue in LPP and likewise in other related methods.

To address the aforementioned issues, an MLPP is proposed here for HSIs. MLPP focuses more on enlarging the differences among ground objects, namely, the nonnearest neighbor relationship. The main purpose of MLPP was to maximize distances between nonnearest neighborhoods while maintaining closeness within each individual neighborhood. In MLPP, Euclidean distance is used to determine the neighbors of each point. Since neighboring points in the original space should have high spectral similarity, the point and its nearest neighbors should be from the same ground object. To achieve this goal, we devise a novel method to build an adjacency graph so that, for a given data point x_i, it is very likely that its nearest neighbors belonged to the same class as that of x_i while others do not. We then pay more attention to distances between nonnearest neighborhoods in order to increase class separability. Furthermore, the MLPP algorithm chooses the nearest neighbors adaptively so that misclassification error is greatly reduced. Thus, MLPP is more suitable for the hyperspectral remote-sensing image.

1) Adjacency Graph Construction and Weights: MLPP utilizes a "graph growing" strategy to construct the adjacency graph to increase the possibility that any given data point and its neighbors are from the same ground object. To start with, MLPP chooses only one nearest point, easing the problem of how to determine the size of the neighborhood. However, considering only one nearest neighbor will cause the distribution of some classes to be scattered in the low-dimensional space. To reduce this effect, a two-step procedure is adopted in MLPP. The first step is to extend the neighboring points. Assume that, for x_i, x_j is its nearest neighbor. For x_j, however, x_k is its nearest neighbor. In this situation, both x_i and x_k are considered to be the nearest points of x_j. The second step is to merge neighborhoods. Following the previous example, x_i and x_k are also regarded as nearest neighbors to each other. The outcome of this procedure is neighborhoods formed with various sizes on which an adjacency graph is constructed without any parameter. Fig. 1(b) illustrates this procedure clearly.

The intention of MLPP is to increase the difference among ground objects by maximizing distances between nonnearest neighborhoods. Therefore, in MLPP, the nearest points are given less weight, more specifically 0, to maintain the nearest neighbor relationship. The nonnearest points are given greater weight. If x_i is not in the neighborhood of x_j, then the weight between them, i.e., W_{ij}, is set to be 1 or by using (2) (we use 1 throughout our experiments). By referring to the adjacency graph, we form the weighted matrix W.

2) Objective Function of MLPP: The core idea of the MLPP algorithm is finding a linear mapping from the original space to the low-dimensional space such that, in the low-dimensional space, the distances between nonneighboring points are maximized and the distances between nearest neighbors, i.e., locality, are preserved. Before we find the linear mapping, we can obtain the low-dimensional embeddings for the original data first as a by-product. This is reflected in the following objective function to be optimized:

$$\max \sum_{ij} \|y_i - y_j\|^2 W_{ij} \tag{4}$$

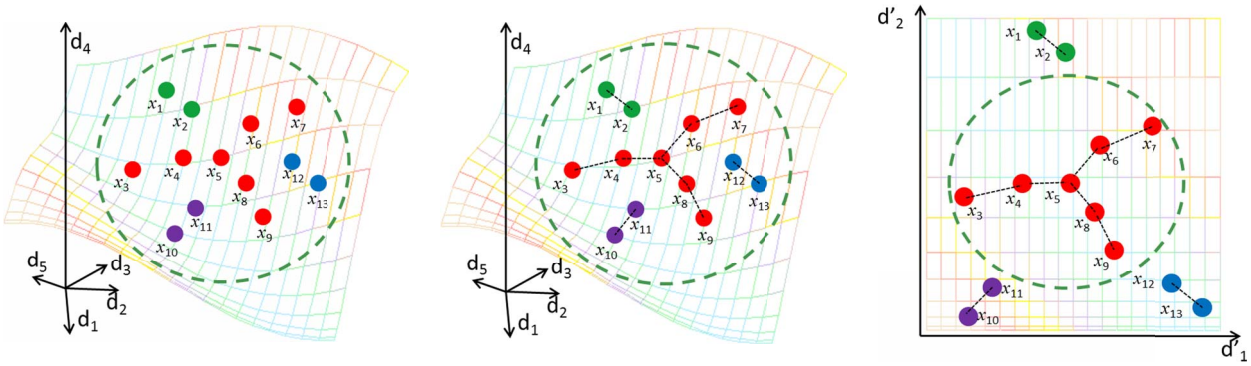

Fig. 1. MLPP effect drawing. (a) Original high-dimensional feature space distribution. x_1 to x_{13} are random points distributed in high-dimensional space. (b) MLPP optimization neighbor relationship. Through the two-step optimization of MLPP, x_1 and x_2, x_3 to x_9, x_{10} and x_{11}, x_{12} and x_{13} respectively become the nearest neighbor relationship. (c) MLPP projected low-dimensional space distribution. In the low-dimensional mapping space, the nonnearest neighbor relationship is further enhanced.

where W is the weight matrix constructed from the adjacency graph derived from the procedure described in the previous section.

The maximization in (4) ensures that, if x_i and x_j are not nearest neighbors, y_i and y_j should be far away. Fig. 1(a)–(c) shows how MLPP is working on a toy example. Now, we derive a linear mapping $a \in R^d$ from R^d to lD space such that $y^T = a^T X$. The objective function to be maximized is

$$\begin{aligned}\frac{1}{2}\sum_{ij}(y_i - y_j)^2 W_{ij} &= \frac{1}{2}\sum_{ij}(a^T x_i - a^T x_j)^2 W_{ij} \\ &= \sum_i a^T x_i D_{ii} x_i^T a - \sum_{ij} a^T x_i W_{ij} x_j^T a \\ &= a^T X(D-W)X^T a = a^T X L X^T a \end{aligned} \tag{5}$$

where D and L are similar to those in (3). It is worth mentioning that the larger the value of D_{ii}, the more the weight on y_i. To remove scaling indeterminacy, we impose the following constraint:

$$y^T D y = 1 \Rightarrow a^T X D X^T a = 1. \tag{6}$$

Therefore, the final optimization problem of MLPP becomes

$$\max_{a^T X D X^T a = 1} a^T X L X^T a. \tag{7}$$

The optimum a that maximizes the objective function is given by the maximum eigenvalue solution to the following generalized eigendecomposition problem:

$$X L X^T a = \lambda X D X^T a. \tag{8}$$

It is easy to show that XLX^T and XDX^T are symmetric and are positive semidefinite. Let column vectors $a_1, \ldots, a_t$ be the eigenvectors of (8), ordered according to their corresponding eigenvalues, $\lambda_1 > \cdots > \lambda_t$. The final linear mapping of MLPP is as follows:

$$A = [a_1, a_2, \ldots, a_t] \tag{9}$$

and $y_i = A^T x_i$, where $y_i \in R^t$ is the low-dimensional embedding of x_i and A is a $d \times t$ matrix.

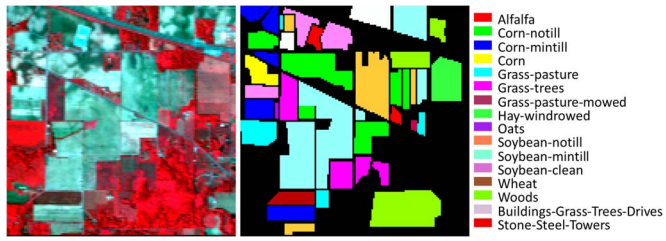

Fig. 2. Indian Pine data set and ground truth.

TABLE I
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON AVIRIS INDIAN PINE DATA SET

DR + Classifier		Gra-pas	Oats	OA(%)	Kappa
Original	NN	5.00	12.50	72.86	0.6981
	SVM	0	0	73.25	0.7012
LDA	NN	35.00	12.50	77.51	0.7428
	SVM	0	0	78.34	0.7481
LPP	NN	85.00	56.25	81.37	0.7877
	SVM	0	0	80.72	0.7768
NPE	NN	80.00	68.75	84.93	0.8280
	SVM	0	0	80.14	0.7692
NWFE	NN	90.00	81.25	85.99	0.8402
	SVM	0	0	81.51	0.7864
MLPP	NN	**100**	**100**	**89.70**	**0.8825**
	SVM	**53.85**	**20.00**	**82.20**	**0.7952**

III. EXPERIMENTS AND RESULTS

A. Experimental Data and Setup

We used two real hyperspectral imagery data sets to evaluate the proposed algorithm. The first hyperspectral data set was captured by the National Aeronautics and Space Administration's Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over northwestern Indiana in the United States in June 1992. The whole scene contains 145 × 145 pixels, with 220 spectral bands in the wavelength range from 400 to 2500 nm and a spatial resolution of approximately 20 m × 20 m. After removing noisy and water absorption bands, it was left with 169 spectral channels. Sixteen ground-truth classes, with a total of 10 249 samples, were considered in our experiments.

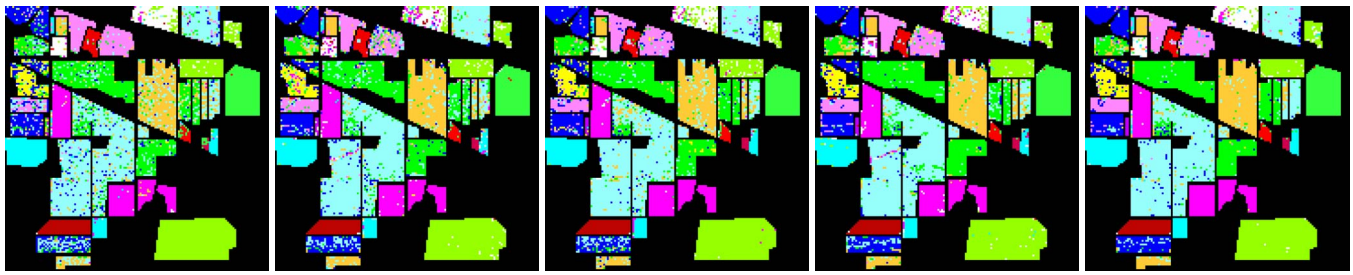

Fig. 3. Thematic maps of classification results of different methods. (a) LDA+1NN. (b) LPP+1NN. (c) NPE+1NN. (d) NWFE+1NN. (e) MLPP+1NN.

Fig. 2 shows the pseudocolor image and ground truth of the Indian Pine data.

The second image data set was acquired by the airborne Reflective Optics System Imaging Spectrometer (ROSIS) over the urban test area of Pavia in northern Italy. The whole scene contains 610 × 340 pixels and is centered at the University of Pavia. It has 103 spectral bands after the removal of 12 bands due to noisy and water absorption. A total of 42 776 pixels from nine classes were used for our experiments.

We compared our algorithm against NPE, LDA, LPP, and NWFE. For simplicity, all the algorithms reduced the data to c-1 dimensional space, where c is the number of classes. The performance of these algorithms was evaluated by classification accuracy and Kappa statistics as all of the DR methods were combined with the 1 Nearest Neighbor (1NN) classifier and support vector machine (SVM) classifier with a radial basis function (RBF) kernel. The SVM with the RBF kernel was performed by using LIBSVM [23].

All the samples in these data sets were divided into two parts for classification: 10% for training and 90% for testing. Note that, in these five algorithms, LDA and NWFE are supervised approaches. Based on our past experiences from LPP and NPE, the number of nearest neighbors (k) is still uncertain. We experimented k in the range of [1, 15] and chose the value with the best overall classification accuracy.

B. Results

The classification accuracies of using these five DR methods for the Indian Pine images are shown in Table I.

The best overall accuracy (OA) is from MLPP: The OA of MLPP+1NN was 89.70% and the kappa index was 0.8825, while the OA of MLPP+SVM was 82.20% and the kappa index was 0.7952. The performance of the MLPP algorithm for small size classes is much better than other algorithms. For example, for Grass-pasture (called "Gra-pas" in the table, 28 samples) and Oats (20 samples) in the Indian Pine data set, the OA of MLPP+1NN is 100%, which is way ahead of the other algorithms, and so is MLPP+SVM. Fig. 3 shows the thematic maps of the classification results of all 5 DR methods plus the 1NN classifier.

C. Test of the Validity of MLPP

To understand the proposed method more intuitively, we selected samples from three classes, soybean-notill, soybean-mintill, and corn-notill from the Indian Pine data set, with 100 samples for each class. The spectral curves of these classes are plotted in Fig. 4, one curve per class in order not to clutter the figure. For simplicity, the target dimensionality of all the algorithms is two.

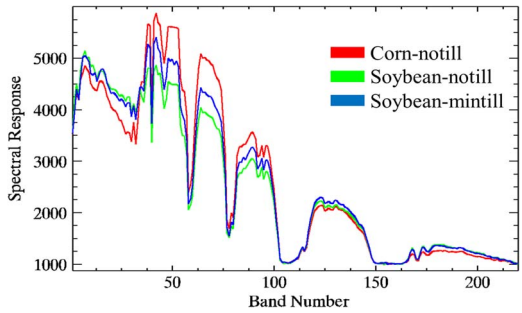

Fig. 4. Spectral curves of the three agricultural classes in the Indian Pine data set. One spectrum per class.

Fig. 5(a) shows the initial data distribution for bands 21 and 55. Fig. 5(b)–(f) displays the projected data obtained by using different DR algorithms. It is quite obvious that the results of LPP, LDA, and NPE are very poor in terms of separating these classes. LDA, a global linear algorithm, has difficulty to capture the nonlinear structure of the data. One possible reason for the poor performance of LPP and NPE is that not all of the nearest neighbors belong to the same class due to the limitation of fixed number of neighbors in constructing the adjacency graph. The NWFE algorithm works well and has the closest performance to that of MLPP, but there still exists overlapped regions between corn-notill and soybean-notill. The MLPP algorithm separates the three classes quite nicely without label information. Although there is some overlapping, it is better than that in NWFE's results.

D. Results Using the Pavia University Data Set

The classification results of the five methods on the Pavia University image are shown in Table II. MLPP has the best OA. MLPP+1NN has an OA of 94.82% and a kappa index of 0.9312; the OA of MLPP+SVM is 87.95%, and the kappa index is 0.8385. Interestingly, MLPP+1NN performed better than MLPP+SVM. The reason is that the construction of the adjacency graph in MLPP indeed produces very "pure" neighborhoods (samples in one neighborhood are mostly from the same class) and therefore enables the 1NN classifier to achieve a high success rate. NWFE has a similar scenario but with much less differences between two classifiers.

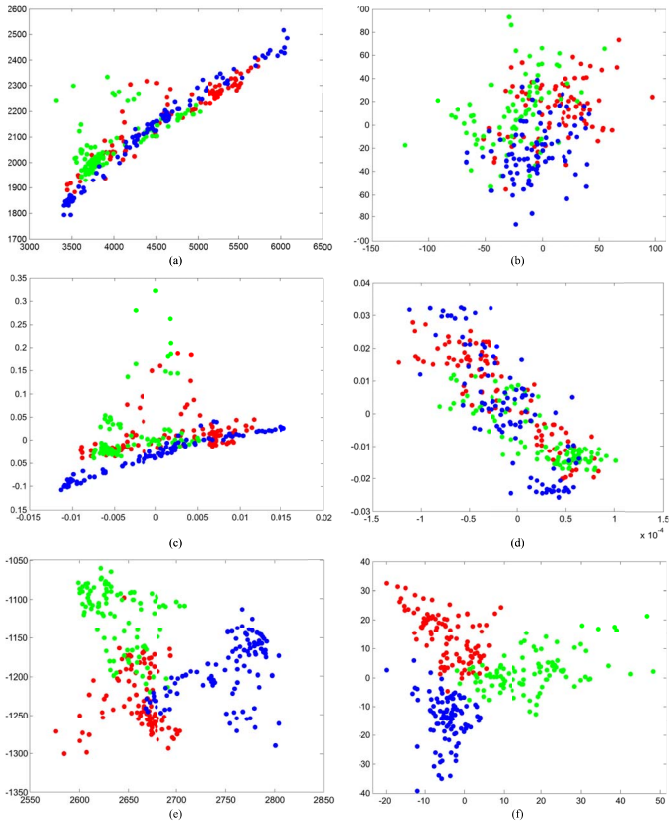

Fig. 5. Projected data produced by different DR algorithms. Different colors stand for different classes. (a) Original (bands 21 and 54). (b) LDA. (c) NPE ($k = 12$). (d) LPP ($k = 10$). (e) NWFE. (f) MLPP.

TABLE II
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON ROSIS PAVIA UNIVERSITY DATA SET

DR + Classifier		OA(%)	kappa
Original	NN	80.16	0.7705
	SVM	81.87	0.7791
LDA	NN	87.58	0.8347
	SVM	83.62	0.7755
LPP	NN	91.66	0.8896
	SVM	86.07	0.8109
NPE	NN	91.64	0.8891
	SVM	84.05	0.7817
NWFE	NN	88.27	0.8431
	SVM	87.62	0.8328
MLPP	NN	**94.82**	**0.9312**
	SVM	**87.95**	**0.8385**

IV. CONCLUSION

In this letter, we proposed a modified version of LPP called MLPP for hyperspectral remote-sensing image classification. Different from the original LPP, MLPP is totally parameter free owing to the adaptive adjacency graph generation and simple weight scheme. MLPP maximizes the distances between nonnearest neighbor points while preserving locality. As a result, separability among classes is greatly enhanced in the low-dimensional space. Experiments on two data sets from different hyperspectral sensors demonstrated that MLPP has superior performance to other DR methods in terms of improving classification accuracy.

REFERENCES

[1] Q. Zhang, L. Zhang, Y. Yang, Y. Tian, and L. Weng, "Local patch discriminative metric learning for hyperspectral image feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 612–616, Mar. 2014.

[2] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.

[3] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[4] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.

[5] C.-I. Chang and H. Ren, "An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1044–1063, Mar. 2000.

[6] 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.

[7] S. Yang *et al.*, "Semisupervised dual-geometric subspace projection for dimensionality reduction of hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3587–3593, Jun. 2014.

[8] Z. Feng, S. Yang, S. Wang, and L. Jiao, "Discriminative spectral-spatial margin-based semisupervised dimensionality reduction of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 224–228, Feb. 2015.

[9] S. Chen and D. Zhang, "Semisupervised dimensionality reduction with pairwise constraints for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 369–373, Mar. 2011.

[10] Q. Shi, L. Zhang, and B. Du, "Semisupervised discriminative locally enhanced alignment for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4800–4815, Sep. 2013.

[11] Y. Song, F. Nie, C. Zhang, and S. Xiang, "A unified framework for semi-supervised dimensionality reduction," *Pattern Recognit.*, vol. 41, no. 9, pp. 2789–2799, Sep. 2008.

[12] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surv. Land Inf. Sci.*, vol. 62, pp. 115–122, 2002.

[13] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, Mar. 2005.

[14] D. de Ridder and R. P. Duin, "Locally linear embedding for classification," Pattern Recog. Group, Dept. Imaging Sci. Technol., Delft Univ. Technol., Delft, The Netherlands, Tech. Rep. PH-2002-01, 2002, pp. 1–12.

[15] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[16] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE 10th ICCV*, 2005, pp. 1208–1213.

[17] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*. Chicago, IL, USA: Univ. Chicago Press, 2004, pp. 153–160.

[18] J. Cheng, Q. Liu, H. Lu, and Y.-W. Chen, "Supervised kernel locality preserving projections for face recognition," *Neurocomputing*, vol. 67, pp. 443–449, Aug. 2005.

[19] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image Vis. Comput.*, vol. 24, no. 3, pp. 239–248, Mar. 2006.

[20] L. Zhang, L. Qiao, and S. Chen, "Graph-optimized locality preserving projections," *Pattern Recognit.*, vol. 43, no. 6, pp. 1993–2002, Jun. 2010.

[21] L. Yang, W. Gong, X. Gu, W. Li, and Y. Liang, "Null space discriminant locality preserving projections for face recognition," *Neurocomputing*, vol. 71, no. 18, pp. 3644–3649, Oct. 2008.

[22] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.*, vol. 4, pp. 119–155, Dec. 2003.

[23] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–39, Apr. 2011.