2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Integrating Local and Global Manifold Structures for Unsupervised Dimensionality Reduction

Xiaochen Chen, Jia Wei*, Jinhai Li and Xiaodong Zhang

Abstract— **Recently there has been a lot of interest in geometrically motivated approaches dealing with data in high dimensional spaces. We consider the case where data is sampled from a low dimensional manifold which is embedded in high dimensional Euclidean space. In this paper, we propose a novel unsupervised linear subspace learning algorithm called Local and Global Manifold Preserving Embedding (LGMPE). Different from existing manifold learning based linear subspace learning algorithms which aims at preserving either single kind of local manifold structure or single kind of global manifold structure on the data manifold, LGMPE can preserve different local and global manifold structures simultaneously in the graph embedding framework. Several experiments on real face datasets demonstrate the effectiveness of the proposed algorithm.**

I. INTRODUCTION

In many real world applications, such as face recognition, information retrieval, bioinformatics, etc, one is often confronted with high dimensional data [30], [6], [1]. However, high dimensionality is a major cause of the practical limitations of many pattern recognition technologies. Moreover, it has been observed that a large number of features may actually degrade the performance of classifiers if the number of the training samples is small relative to the number of features. This is called the "curse of dimensionality" [8]. Fortunately, there might be reason to suspect that the naturally generated high dimensional data probably reside on a lower dimensional manifold. This leads one to consider methods of dimensionality reduction that allow one to represent the data in a lower dimensional space.

The goal of dimensionality reduction is to reduce the complexity of the input data while some desired intrinsic information of the data is preserved [21], [14], [25], [27], [7], [16], [13], [26], [22]. Two of the most popular methods for dimensionality reduction are Principal Component Analysis (PCA) [21] and Linear Discriminant Analysis (LDA) [14], which are unsupervised and supervised respectively. PCA tries to preserve the global covariance structure of the data in low dimensional projection subspace without knowing the class labels of the data; while LDA aims to minimize the within-class similarity and maximize the between-class similarity simultaneously in low dimensional projection subspace when the class labels of the data are available.

X. Chen, J. Wei and J. Li are with the School of Computer Science and Engineering, South China University of Technology, China. (email: 1092441751@qq.com, csjwei@scut.edu.cn, 260835954@qq.com)

X. Zhang is with the Department of Information, The Third Affiliated Hospital, Sun Yat-sen University, China. (email: zhangxd@139.com)

This work is supported by the Guangdong Natural Science Foundation (S2012040008022) and the National Natural Science Foundation of China (61273363).

Unsupervised dimensionality reduction method PCA is optimal only when the data space is linear, however, there is no evidence that the data is sampled from a linear subspace. For example, it is always believed that the face images are sampled from a nonlinear low-dimensional manifold which is embedded in the high-dimensional ambient space [12]. This motivates us to consider manifold based techniques for dimensionality reduction. So, various manifold learning techniques, such as ISOMAP [20], Locally Linear Embedding (LLE) [17], Laplacian Eigenmap (LE) [2], Local Tangent Space Alignment (LTSA) [29], Maximum Variance Unfolding (MVU) [24] and its relaxed variance Distinguishing Variance Embedding (DVE) [23], have been proposed which reduce the dimensionality of a fixed training set in a way that maximally preserves certain inter-point relationships. LLE, LE and LTSA are local methods which attempt to preserve local geometry of the data; essentially, they seek to map nearby points on the manifold to nearby points in the low dimensional representation. ISOMAP, MVU and DVE are global methods which consider preserving both local and global geometry of the data, mapping nearby points on the manifold to nearby points in low dimensional space to preserve local geometry, and faraway points to faraway points in low dimensional space to preserve global geometry. One of the major limitations of these methods is that they do not generally provide a functional mapping between the high and low dimensional spaces that are valid both on and off the training data.

To address the out-of-sample problem by explicitly defining an embedding function, linear versions of the manifold learning methods were proposed, such as Isometric Projection (IsoProjection) [3], Neighborhood Preserving Embedding (NPE) [10], Locality Preserving Projections (LPP) [11], Linear Local Tangent Space Alignment (LLTSA) [28], and so on. IsoProjection can be viewed as the linear approximation of ISOMAP, which can be obtained by preserving the pairwise distances on the manifold. NPE can be viewed as the linear approximation of LLE, which aims at preserving the local neighborhood structure on the data manifold. LPP is the linear version of LE, which is obtained by finding the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. LLTSA is the linear version of LTSA, which uses the tangent space in the neighborhood of a data point to represent the local geometry, and then aligns those local tangent spaces in the low-dimensional space which is linearly mapped from the raw high-dimensional space. These methods are defined everywhere, rather than only on the training data points.

However, the linear manifold learning methods mentioned above can only preserve single kind of manifold structure either local or global. In this paper, we propose a novel unsupervised dimensionality reduction algorithm called Local and Global Manifold Preserving Embedding (LGMPE), which can preserve different local and global manifold structures simultaneously in the graph embedding framework. Given a set of data points in the ambient space, we first construct a nearest neighbor graph of the observed data. We then build two manifold preserving matrixes. One describes the local manifold structure using linear reconstruction coefficients of the neighboring data points which are connected in the nearest neighbor graph. The other describes the global manifold structure using shortest paths of pairs of data points which are not connected in the nearest neighbor graph. We then find an optimal embedding such that the different local and global manifold structures can be preserved simultaneously in the dimensionality reduced space.

The organization of this paper is as follows. In Section 2, we show how to derive the LGMPE algorithm from the nearest neighbor graph and its complementary graph by preserving different local and global manifold structures. In Section 3, the relationship and comparison with other methods are discussed. In Section 4, the proposed method is tested and the results are compared with those of popular manifold learning based linear subspace learning algorithms. Last, we conclude our work in Section 5.

II. Local and Global Manifold Preserving Embedding

A. Unsupervised Linear Dimensionality Reduction Problem

The generic problem of unsupervised linear dimensionality reduction is the following: Given a set of points $x_1, x_2, ..., x_m$ in R^n, find a transformation matrix A that maps these m points to a set of points $y_1, y_2, ..., y_m$ in R^d, where $d << n$, such that y_i "represents" x_i, namely $y_i = A^T x_i$, where A is a $n \times d$ matrix.

B. Local Manifold Structure Preserving

To preserve local manifold structure, we first construct a K nearest neighbor graph G with m nodes on the original data set. The i-th node corresponds to the data point x_i. We put an edge between node i and j if x_j is among the K nearest neighbors of x_i or x_i is among the K nearest neighbors of x_j. In many cases, the data points might reside on a nonlinear submanifold, but it might be reasonable to assume that each local neighborhood is linear as Roweis et al. did in LLE [17]. Thus, we can characterize the local geometry of these patches by linear coefficients that reconstruct each data point from its neighbors. Reconstruction errors which are measured by the cost function should be minimized:

$$\phi(W) = \sum_i \|x_i - \sum_{j:x_j \in N_K(x_i)} W_{ij} x_j\| \tag{1}$$

where $N_K(x_i)$ represents the K nearest neighbors of x_i. We constrain $\sum_{j:x_j \in N_K(x_i)} W_{ij} = 1$, with $W_{ij} = 0$ if x_j does not belong to the set of K nearest neighbors of x_i. Please see [17] for the details about how to solve the above minimization problem to compute the linear coefficients matrix W. In the low dimensional space, to preserve local manifold structure, we want to minimize the cost function by fixing W as following:

$$\begin{aligned} \Phi(Y) &= \textstyle\sum_i \|y_i - \sum_{j:x_j \in N_K(x_i)} W_{ij} y_j\| \\ &= trace(YMY^T) \\ &= trace(A^T XMX^T A) \end{aligned} \tag{2}$$

where

$$Y = (y_1, y_2, ..., y_m)$$

$$X = (x_1, x_2, ..., x_m)$$

$$M = (I - W)^T (I - W)$$

$$I = diag(\underbrace{1, ..., 1}_{m})$$

It is easy to check that M is symmetric and semi-positive definite. In one dimensional case, the optimal projection is given by solving the following minimization problem:

$$a^* = \underset{a}{argmin}(a^T XMX^T a) \tag{3}$$

where a is a $n \times 1$ vector.

C. Global Manifold Structure Preserving

To preserve global manifold structure, we first construct the complementary graph G^c of the K nearest neighbor graph G. In graph G^c, an edge is added between nodes i and j if they are not connected in graph G, otherwise not. Noting that $G + G^c$ is a complete graph. Thus, we can characterize the global geometry of the original space by finding an embedding function f which can preserve the geodesic distances in the reduced embedding Euclidean space if two samples are connected in graph G^c. That is:

$$f^* = \underset{f}{argmin} \sum_{(x_i, x_j) \in G^c} (d_M(x_i, x_j) - d(f(x_i), f(x_j)))^2 \tag{4}$$

where $d_M(x_i, x_j)$ means the geodesic distance between x_i and x_j on the manifold M, $d(f(x_i), f(x_j))$ represents the Euclidean distance between $f(x_i)$ and $f(x_j)$ in the reduced embedding space. In real world data set, the underlying manifold M is often unknown and hence the geodesic distance measure is also unknown. So, we can only compute the shortest path distances $d_G(x_i, x_j)$ on the graph G to estimate the geodesic distances $d_M(x_i, x_j)$. The procedure of computing shortest path is as follows: Initialize $d_G(x_i, x_j) = d(x_i, x_j)$, if x_i and x_j are linked by an edge; $d_G(x_i, x_j) = \infty$, otherwise. Then for each value of $p = 1, 2, ..., m$ in turn, replace all entries $d_G(x_i, x_j)$ by: $min\{d_G(x_i, x_j), d_G(x_i, x_p) + d_G(x_p, x_j)\}$. The matrix of final values $D_G = \{d_G(x_i, x_j)\}$ will contain the shortest path distances between all pairs of points in G. This procedure is named Floyd-Warshall algorithm [5]. For the convenience of following use, we define:

$$D^c_{G_{ij}} = \begin{cases} D_{G_{ij}} & if\ (x_i, x_j) \in G^c \\ 0 & otherwise \end{cases} \tag{5}$$

In the following, we apply techniques from Multi-Dimensional Scaling (MDS) to convert distances to inner products, which uniquely characterize the geometry of the data in a form that supports efficient optimization [20]. Specifically, let D be the distance matrix such that D_{ij} is the distance between x_i and x_i. Define matrix S where $S_{ij} = D_{ij}^2$ and $H = I - \frac{1}{m}ee^T$ which is often called "centering matrix", where $I = diag(\underbrace{1,...,1}_{m})$, $e = (\underbrace{1,...,1}_{m})^T$. It can be shown that $\tau(D) = -HSH/2$ is the inner product matrix. That is $D_{ij}^2 = \tau(D)_{ii} + \tau(D)_{jj} - 2\tau(D)_{ij}$.

Let D_Y denote the Euclidean distance matrix in the reduced embedding space. Thus, the objective function in equation (4) becomes minimizing the following equation:

$$\|\tau(D_G^c) - \tau(D_Y)\|_{L^2} \tag{6}$$

where $\|Q\|_{L^2}$ means the L^2 matrix norm $\sqrt{\sum_{ij} Q_{ij}^2}$.

Consider a linear embedding function $f(x) = a^T x$, where a is a $n \times 1$ vector. Let $y_i = f(x_i)$, $Y = a^T X$. We have:

$$\tau(D_Y) = Y^T Y = X^T a a^T X \tag{7}$$

The optimal projection is given by solving the following minimization problem:

$$a^* = \underset{a}{argmin} \|\tau(D_G^c) - X^T a a^T X\|_{L^2}^2 \tag{8}$$

We can get that:

$$\begin{aligned} &\|\tau(D_G^c) - X^T a a^T X\|_{L^2}^2 \\ &= trace((\tau(D_G^c) - X^T a a^T X)(\tau(D_G^c) - X^T a a^T X)^T) \\ &= trace(\tau(D_G^c)\tau(D_G^c)^T - X^T a a^T X \tau(D_G^c)^T \\ &-\tau(D_G^c) X^T a a^T X + X^T a a^T X X^T a a^T X) \\ &= trace(\tau(D_G^c)\tau(D_G^c)^T) - 2trace(a^T X \tau(D_G^c) X^T a) \\ &+trace(a^T X X^T a a^T X X^T a) \end{aligned} \tag{9}$$

Because $\tau(D_G^c)$ is already known, the minimization problem of equation (8) can be rewritten as follows:

$$a^* = \underset{a}{argmax}(a^T X \tau(D_G^c) X^T a - a^T X X^T a a^T X X^T a) \tag{10}$$

D. The Algorithm

Having the above preparation we can get the optimal projection of LGMPE from equation (3) and (10):

$$\begin{aligned} a^* = \underset{a}{argmax}(\alpha(a^T X \tau(D_G^c) X^T a - a^T X X^T a a^T X X^T a) \\ -\beta a^T X M X^T a) \end{aligned} \tag{11}$$

where α and β are scaling parameters to balance the contributions of the global and local manifold structures. On one hand, since $a^T X \tau(D_G^c) X^T a - a^T X X^T a a^T X X^T a$ can measures to what extent can Euclidean distances in the reduced embedding space provide an approximation to the geodesic distances on manifold M if two samples are connected in graph G^c, it should be the larger the better. On the other hand, a small $a^T X M X^T a$ implies that every data point in the reduced embedding space can be represented as a good linear combination of its neighbors with respect to the linear coefficients matrix W that reconstruct each data point from its neighbors in graph G. Thus, equation (11) should be as large as possible, which indicates that both the global and local manifold structures are well preserved in the dimensionality reduced space. In order to remove an arbitrary scaling factor in the projection, we impose a constraint as follows:

$$a^T X X^T a = 1 \tag{12}$$

Thus, equation (11) can be written as follows:

$$a^* = \underset{a^T X X^T a=1}{argmax} \; (a^T X(\alpha\tau(D_G^c) - \beta M) X^T a - \alpha) \tag{13}$$

The vectors $a_i (i = 1, 2, ..., d)$ that maximize the above objective function are given by the eigenvectors corresponding to the maximum eigenvalues of the generalized eigenproblem:

$$X(\alpha\tau(D_G^c) - \beta M) X^T a = \lambda X X^T a \tag{14}$$

Let $A = (a_1, a_2, ..., a_d)$, the linear embedding is as follows:

$$x \rightarrow y = A^T x \tag{15}$$

where y is a d-dimensional representation of the high dimensional data point x. A is the $n \times d$ transformation matrix.

To get a stable solution of the above eigenproblem, the matrix XX^T is required to be non-singular which is not true when the number of features is larger than the number of samples. In this case, we first project the original data into the PCA subspace to overcome the small sample size problem and to throw away the smallest principal components, and then LGMPE can be applied in the reduced subspace.

III. Comparison With Related Methods

A. Comparison With Nonlinear Manifold Learning Methods

Compared with nonlinear manifold learning methods such as Isomap, LLE, LE, LTSA, MVU, DVE and so on, LGMPE provides an explicit linear transformation matrix to project the original data to the reduced low dimensional space, which aims to preserving both the global and local structure of the underlying data manifold. LGMPE is defined everywhere, while nonlinear manifold learning methods can't provide. Therefore, query points can also be mapped into the low-dimensional representation space in which retrieval, clustering and classification may be performed.

B. Comparison With NPE

NPE is the linear approximation of LLE which emphasize on preserving the local structure of the data set, but ignore the global structure of the data set. However, researchers have recently found that both global structure and local structures are important for learning from high dimensional data [27], [4], [19]. LGMPE aims to preserving both global and local structures of the underlying data manifold. And it is easy to see that NPE is the special case of LGMPE with $\alpha = 0$, so its capability is limited. In the same way, LGMPE is also different from other manifold learning based linear subspace learning methods such as LPP and LLTSA which preserve local structure only.

C. Comparison With IsoProjection

IsoProjection is the linear approximation of ISOMAP. It can only preserve local and global distances structures in the embedded space. LGMPE aims to preserve the local linear reconstruction coefficients between samples which are nearest neighbors and preserve global distances between samples which are not nearest neighbors. Different from IsoProjection, LGMPE considers that local linear reconstruction coefficients are more valuable than local distances when preserving local structure. Besides, LGMPE is more flexible to balance the contribution of global and local manifold structures through α and β.

IV. Experiments

In this section, the performance of LGMPE is evaluated on the AR [15], PIE [18], and Extended Yale-B [9] face image databases and compared with the classification performances of Baseline, PCA, LPP, NPE and IsoProjection. We use the preprocessed versions of the PIE database and the Extended Yale-B database which are publicly available from the web page of D. Cai [1]. The AR database is available from the web page of A. M. Martinez [2].

A. Database Description

The AR database contains 126 subjects (70 men and 56 women) and each subject has 26 face images taken in two sessions separated by 2 weeks time. For each session, 13 face images with varying facial expression, illumination and occlusion were captured. For the experiments reported in this section, 100 different individuals were randomly selected from this database. The images are morphed to the 42×30 pixel arrays, segmented using an oval-shaped mask, and converted to 256 gray-level images by adding all three color channels.

The PIE face database contains 41,368 images of 68 people, each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. In this experiment, our dataset only contains five near frontal poses (C05, C07, C09, C27, C29) and all the images under different illuminations and expressions. That is there are 170 images for each individual. All the face images are aligned and cropped. The cropped images are 32×32 pixels, with 256 gray levels per pixel.

The Extended Yale-B face database contains 16128 images of 28 human subjects under 9 poses and 64 illumination conditions. In this experiment, we choose the frontal pose and use all the images under different illumination, thus we get 64 images for each person. All the face images are aligned and cropped. The cropped images are 32×32 pixels, with 256 gray levels per pixel.

B. Experimental Settings

In our experiments, we first project the face images into a PCA subspace by retaining 99% of the principal component to deal with small sample size problem. Second, we carry out these algorithms on the training face images and learn the transformation matrix. Third, each test face image is mapped into a low-dimensional subspace via the transformation matrix. Finally, we classify the test images by the nearest neighbor classifier. For the baseline method, the recognition is simply performed in the input space without dimensionality reduction.

In all of the experiments, LGMPE, IsoProjection, NPE and LPP use K nearest neighbor to construct the adjacency graphs. For LPP, we use the simple-minded way to choose the weights of the adjacency graph: $W_{ij} = 1$ if and only if vertices i and j are connected by an edge, otherwise $W_{ij} = 0$. For LGMPE, we set $\alpha = 1$ and β is searched from the grid: $\{1, 10, 10^2, ..., 10^{12}\}$ with Hold-Out method over 20 random splits of training and testing samples.

In our experiments, all the results are the average over 20 random splits of training and testing samples. For convenience, we make some notations used in illustrations: Tr - Number of training samples of each class, Dims - Number of reduced dimensions, K - Nearest neighborhood size.

C. Experimental Results

The recognition results of AR are illustrated in Figure 1. We can see that LGMPE has higher recognition rates than those of other methods for most cases with the variations of Tr, Dims and K. This says that integrating both the global geodesic distances and local linear reconstruction coefficients does have positive effect for the dimensionality reduction problem.

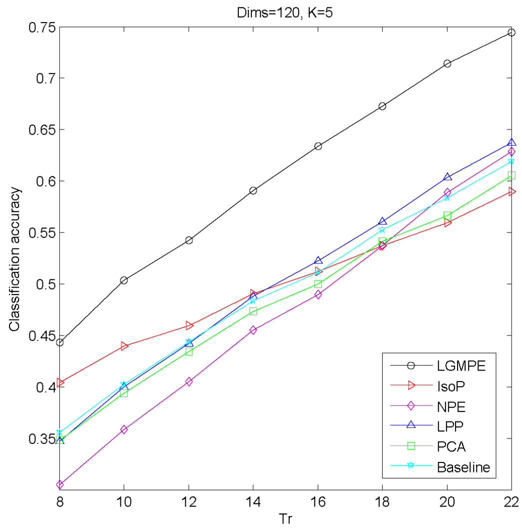
The recognition results of PIE are illustrated in Figure 2. From Figure 2(c), we can see that LGMPE is better than IsoProjection and NPE with the variation of K. Although LPP is better than LGMPE when $K = 2$, LGMPE is more stable than LPP with the increase of K. The property of LGMPE which is not sensitive to K is very useful for practical use.

The recognition results of Extended Yale-B are illustrated in Figure 3. Again, we can see that LGMPE has higher recognition rates than those of other methods for most cases, and we can get from Figure 3(c) that LGMPE is stable while LPP is very sensitive to the increase of K.

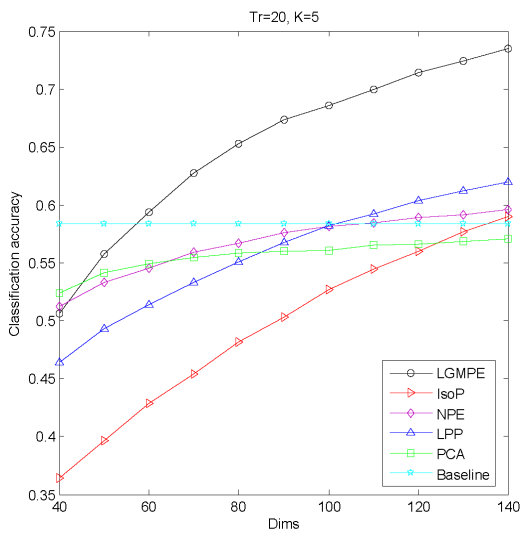
D. Paramater Analysis of β

This experiment is done to evaluate the influence of β which can determine the contribution of the local manifold structure in LGMPE. As can be seen from Figure 4 (The X-axis denotes the index of β with the radix equals 10), the performance of LGMPE is close to that of IsoProjection when β is small, and the performance of LGMPE is close to that of NPE when β is large enough. The most important is that the performance of LGMPE is better than both IsoProjection and NPE when β is assigned to an appropriate value. This explains that why LGMPE can get good results in face recognition tasks.

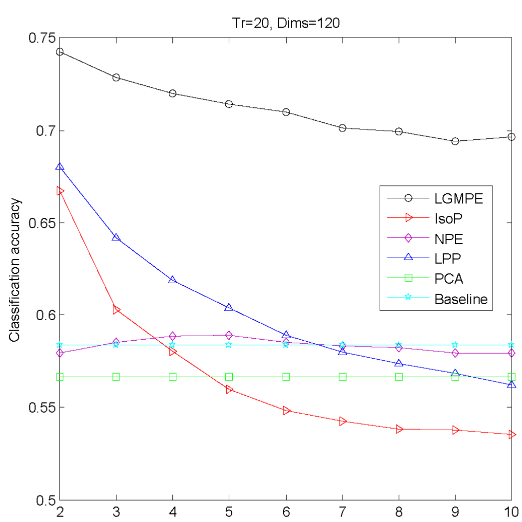
[1]http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html
[2]http://www2.ece.ohio-state.edu/ aleix/ARdatabase.html



(a)

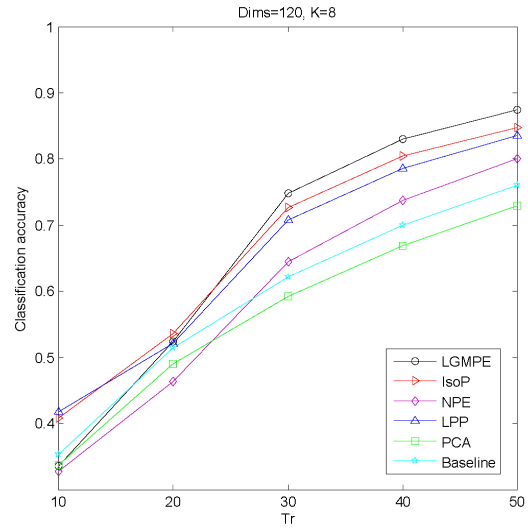


(b)

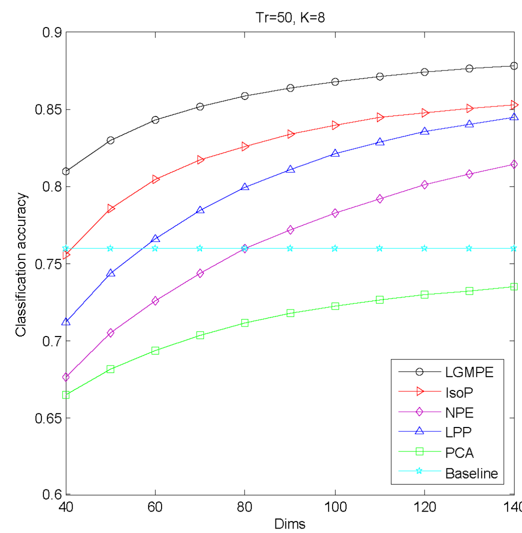


(c)

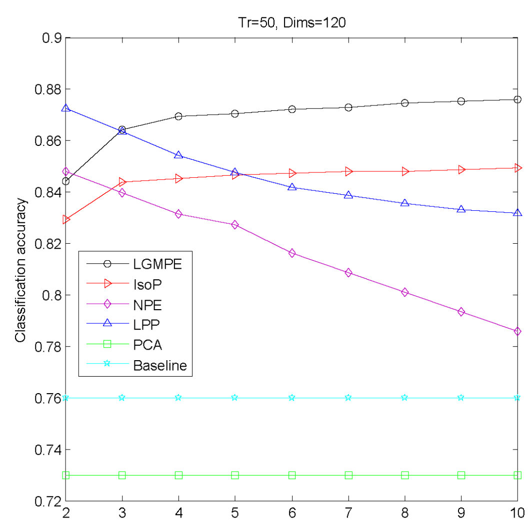
Fig. 1. Experimental results on AR.



(a)

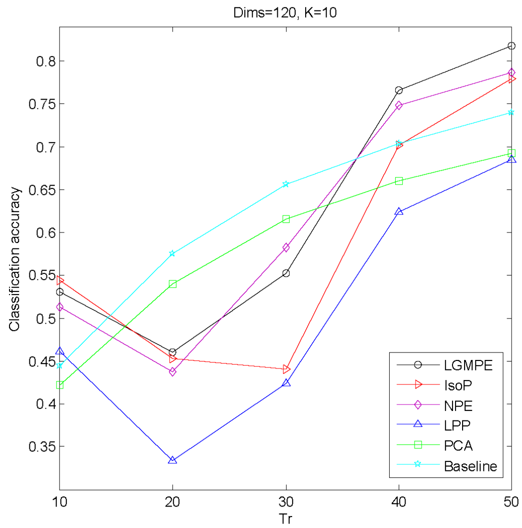


(b)

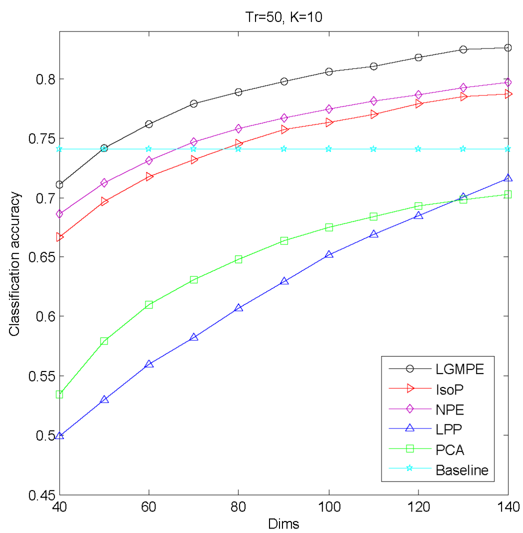


(c)

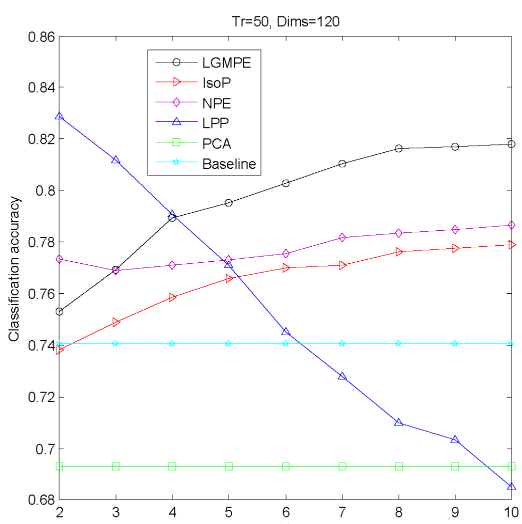
Fig. 2. Experimental results on PIE.



(a)

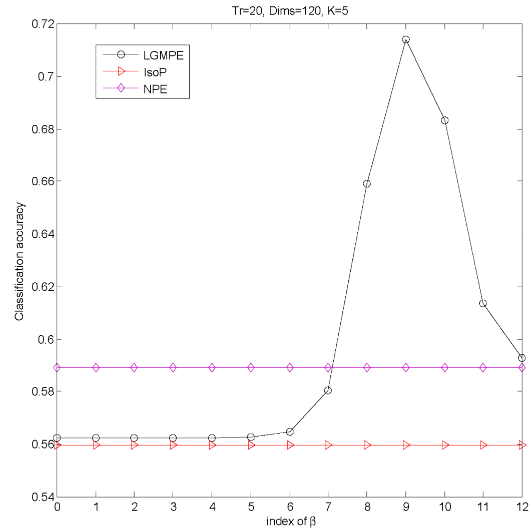


(b)

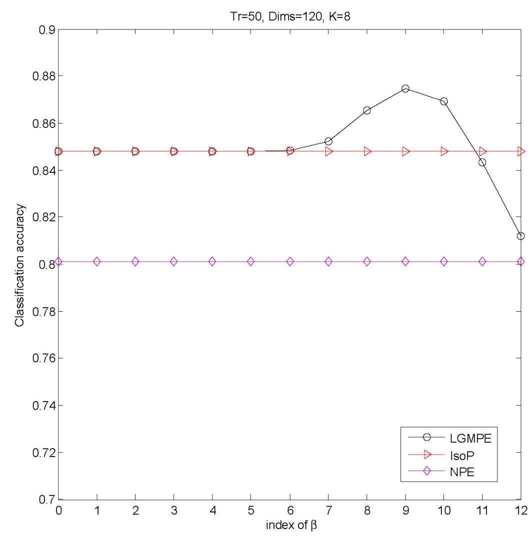


(c)

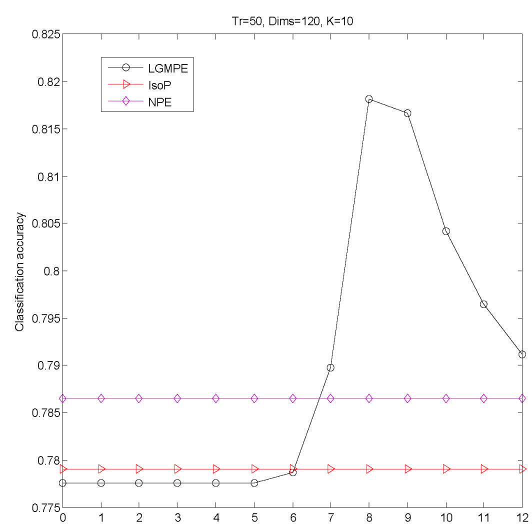
Fig. 3. Experimental results on Extended Yale-B.



(a) AR



(b) PIE



(c) Extended Yale-B

Fig. 4. Paramater Analysis of β.

V. Conclusions

In this paper, we propose a novel linear dimensionality reduction algorithm called Local and Global Manifold Preserving Embedding (LGMPE). LGMPE can preserve local linear reconstruction structure and global geodesic distance structure simultaneously, and can balance the contributions of the two parts. It yields a map which is simple, linear, and defined everywhere. Experiments on real face datasets have been conducted to demonstrate the effectiveness of our algorithm. In our future work, we will combine other local and global structures in this way, and find out whether this way can still work.

References

[1] C. Bartenhagen, H. U. Klein, C. Ruckert, X. Jiang, and M. Dugas. Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data. *BMC Bioinformatics*, 11:567, 2010.

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[3] D. Cai, X. He, and J. Han. Isometric projection. In *Proc. of the 22nd AAAI Conference on Artificial Intelligence*, 2007.

[4] J. Chen, J. Ye, and Q. Li. Integrating global and local structures: A least squares framework for dimensionality reduction. In *Proc. of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, Cambridge, 2nd edition, 2001.

[6] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 2008.

[7] W. Deng, J. Hu, J. Guo, H. Zhang, and C. Zhang. Comments on "globally maximizing, locally minimizing: Unsupervised discriminant projection with application to face and palm biometrics". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(8):1503–1504, 2008.

[8] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, New York, 2nd edition, 2001.

[9] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[10] X. He, D. Cai, S. Yan, and H. J. Zhang. Neighborhood preserving embedding. In *Proc. of the IEEE International Conf. on Computer Vision*, 2005.

[11] X. He and P. Niyogi. Locality preserving projections. In *Advances in Neural Information Processing Systems 16*, pages 153–160, 2004.

[12] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using laplacianfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

[13] H. Li, T. Jiang, and K. Zhang. Efficient and robust feature extraction by maximum margin criterion. *IEEE Trans. on Neural Networks*, 17(1):157–165, 2006.

[14] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Face recognition using LDA-based algorithms. *IEEE Trans. on Neural Networks*, 14(1):195–200, 2003.

[15] A. M. Martinez and A. C. Kak. PCA versus LDA. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.

[16] L. Qiao, S. Chen, and X. Tan. Sparsity preserving projections with applications to face recognition. *Pattern Recognition*, 43:331–341, 2010.

[17] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2327, 2000.

[18] T. Sim, S. Barker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[19] M. Sun, C. Liu, J. Yang, Z. Jin, and J. Yang. A two-step framework for highly nonlinear data unfolding. *Neurocomputing*, 73:1801–1807, 2010.

[20] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[21] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. of the 1991 IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[22] L. J. P. van der Maaten, E. O. Postma, and H. J. van den Herik. Dimensionality reduction: A comparative review. TiCC TR 2009-005, Tilburg University, 2009.

[23] Q. Wang, J. Li, and X. Wang. Distinguishing variance embedding. *Image and Vision Computing*, 28:872–880, 2010.

[24] K. Q. Weinberger and L. K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision*, 70(1):77–90, 2006.

[25] M. Wu, K. Yu, S. Yu, and B. Schölkopf. Local learning projections. In *Proc. of the 24th International Conference on Machine Learning*, 2007.

[26] S. C. 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. on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[27] J. Yang, D. Zhang, J. Y. Yang, and B. Niu. Globally maximizing, locally minimizing: Unsupervised discriminant projection with application to face and palm biometrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(4):650–664, 2007.

[28] T. Zhang, J. Yang, D. Zhao, and X. Ge. Linear local tangent space alignment and application to face recognition. *Neurocomputing*, 70:1547–1553, 2007.

[29] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal of Scientific Computing*, 26(1):313–338, 2004.

[30] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.