Journal of Pattern Recognition Research 1 (2007) 1-16

Growing Locally Linear Embedding for Manifold Learning

Junsong Yin *jsyin@nudt.edu.cn*
Dewen Hu *dwhu@nudt.edu.cn*
Zongtan Zhou *narcz@163.com*
College of Mechatronics and Automation, National University of Defense Technology
Changsha Hunan, 410073, China P. R.

Received: 18 May 2006 Accepted: 17 December 2006 Published online: 17 March 2007

Abstract

Locally linear embedding is an effective nonlinear dimensionality reduction method for exploring the intrinsic characteristics of high dimensional data. This paper proposes a new manifold learning method, which is based on locally linear embedding and growing neural gas and is termed growing locally linear embedding (GLLE). GLLE overcomes the major limitations of the original locally linear embedding, which are intrinsic dimensionality estimation, selection of the number of nearest neighbors, and computational complexity. By embedding the topology learning mechanism in growing neural gas, the proposed GLLE algorithm preserves global topological structures and geometric characteristics of input patterns, which makes the projections more stable. The performed theoretical analysis and experimental simulations show that GLLE results in a faster learning procedure and a lower reconstruction error, which widens the applicability of manifold learning.

Keywords: Nonlinear dimensionality reduction, Global topology preservation, Locally linear embedding, Growing neural gas, Growing locally linear embedding.

1. Introduction

A central problem in machine learning and pattern recognition is the development of appropriate representations of complex data. Most real data lie on a low dimensional manifold embedded in a high dimensional space. High dimensional data contain redundancies and correlations that hide important relationships. Data analysis can be used to eliminate these redundancies and to reduce data complexities. A dimensionality reduction algorithm maps high dimensional data into a low dimensional space, revealing the underlying structure in the data.

Dimensionality reduction, including linear and nonlinear methods, is a useful operation of data visualization and feature extraction in clustering and pattern recognition. Well known principle component analysis (PCA) [1] and linear discriminant analysis (LDA) [3] are linear dimensionality reduction methods. Multi-dimensional scaling [2], isometric maps (ISOMAP) [4], locally linear embedding (LLE) [5], Laplacian eigenmaps [6], and self-organizing maps (SOM) [7] are examples of nonlinear dimensionality reduction methods. All of those methods can reduce the redundancies while retaining the primary characteristics.

LLE is an effective nonlinear dimensionality reduction algorithm proposed first by Roweis in 2000 [8]. Compared to the other methods, the LLE algorithm requires only two parameters to be determined and attains global minimization of the reconstruction error while avoiding plunging into local extrema. Although the LLE algorithm was demonstrated on a number of artificial and real-world data sets, several limitations restrict its wide application. The two parameters that

© 2007 JPRR. All rights reserved.

have to be specified are the intrinsic dimension d and the number of nearest neighbors, K. Improper values of these parameters greatly influence the results. On one hand, a large value of the intrinsic dimension d, amplifies noise effects, while a low value leads to overlaps in mapping results (excessively reduced). On the other hand, a low number of nearest neighbors K cannot make the reconstruction to reveal the global features of the original data, while a large K causes a manifold to loose the nonlinear feature and behave like the traditional PCA [9]. Furthermore, LLE was found to be sensitive to the amount of the initial data [24]. When there is insufficient data (poorly-sampled manifolds), local characteristics are lost. An excessive data amount results in an incomplete reconstruction and a long computational time.

Researchers have proposed several modifications to the LLE algorithm. One modification uses the selection of the optimal value of parameter K [10]. Another modification is a supervised locally linear embedding (SLLE) algorithm [11, 12]. Yet another modification uses a substitute algorithm HLLE based on Hessian eigenmaps [13]. These modified algorithms do reduce the learning time, enhance the space partition ability, and reduce reconstruction error. However, they cannot make LLE an adaptively optimal map, and are also restricted when the manifold is noisy.

We proposed a new algorithm called growing locally linear embedding (GLLE) [24]. GLLE embeds the global topology learning mechanism in growing neural gas (GNG) network and competitive Hebbian learning (CHL) rule. When topology learning is introduced, the improved GNG algorithm can not only map the probability distribution of the input manifold, but also reveal its intrinsic dimensionality [15]. The number of neural nodes in the neural network is fewer than the number of input patterns. The proposed GLLE algorithm is capable of identifying the two parameters adaptively and reduces the time consumption by decreasing the samples reasonably and preserving the global topological and geometrical structures. This nonlinear dimensionality reduction method, unifying a differential geometric operator and topology learning, can keep the local features and preserve the global topology at the same time. Additionally, a linear transformation of the spectrum matrix $\tilde{M} = I - M$ makes the selection of the lowest eigenvectors become the selection of the highest ones with the same eigenvalues, which makes the algorithm more stable.

The paper addresses the three limitations of LLE: intrinsic dimensionality estimation, selection of nearest neighbors, and time consumption. The main advantages of the proposed algorithm are determining the two parameters adaptively, decreasing the computational complexity, introducing the preservation of the global topology, and improving the robustness of the algorithm.

The remainder of this paper is organized as follows. Section 2 briefly reviews the original LLE algorithm and the topological version of GNG. In Section 3 the limitations of LLE are discussed, and the improved manifold learning algorithm is presented. By embedding the topology learning mechanism in the improved GNG, GLLE gives a satisfied solution to the several shortcomings of LLE. Section 4 analyzes the performance of LLE and GLLE. In Section 5 experimental results of handwritten digits and multi-pose faces recognition are presented. Section 6 offers conclusions and discussions.

2. Overview of LLE and GNG

This section briefly describes the original LLE algorithm and the GNG method.

2.1 The LLE Algorithm

Locally linear embedding was first proposed by Roweis and Saul in 2000. LLE's primary idea is to reconstruct a nonlinear manifold by embedding a local linear hyperplane [8]. LLE is an unsupervised learning algorithm. It preserves the relationships between neighbors in manifold

data and represents high dimensional data in a lower dimensional Euclidean space. LLE maps a data set $X=\{X_1, X_2, \ldots, X_N\}$, $X_i \in R^D$ globally to a low dimensional set $Y=\{Y_1, Y_2, \ldots, Y_N\}$, $Y_i \in R^d$ with $d<D$. The algorithm has three steps:

Algorithm 1 The original LLE [8].

1. Find K-nearest neighbors of each point X_i. The vector X_i and its vicinity are set to fit a hyperplane.
2. Compute constrained weights W_{ij} that best linearly reconstruct X_i from its neighbors $X_i \leftarrow \sum W_{ij} X_j$. Reconstruction errors are measured by the cost function:

$$\varepsilon(W)=\sum_i |X_i-\sum_j W_{ij} X_j|^2 . \tag{1}$$

The optimal weights W_{ij} subject to constraints are found by solving the following least squares problem:
3. Compute low dimensional embedding vectors $Y_i \in R^d$ best reconstructed by W_{ij} minimizing equation

$$\phi(Y)=\sum_i |Y_i-\sum_j W_{ij} Y_j|^2 , \tag{2}$$

under the constraints $\sum_i Y_i=0$ and $1/N \sum_i Y_i^T Y_i=I$, where N is the number of samples.

The embedding cost function can be rewritten as

$$\phi(Y)=\sum_i |Y_i-\sum_j W_{ij} Y_j|^2=\sum_i |(I-W)Y_i|^2=\mathrm{tr}(Y^T MY) , \tag{3}$$

where $M \in R^{N\times N}$ and $M=(I-W)^T(I-W)$. Now the LLE embedding problem is transformed into the computation of the bottom d non-zero eigenvalues of matrix M, [9].

2.1 Growing Neural Gas and Its Improvements

Fritzke [14, 16] proposed the growing neural gas (GNG) method in 1995 as an unsupervised non-topology preservation self-organizing neural network. GNG is initialized with a random number of nodes and little a prior information. By the dynamic node growing and dying mechanism, GNG can cluster and partition the input space satisfactorily. GNG can merely reflect the probability distribution of the input patterns, but cannot reveal the topological structure of the embedding manifold. Bruske and Sommer [26] added topology preservation to the traditional GNG. Later competitive Hebbian learning rule [17] and edge-aging mechanism [15], similar to the rival penalized competitive learning [25], were incorporated to model the graph connectivity and update the topological structure. The improved GNG can preserve the geometric features, and reveal the intrinsic topology and dimensional properties of the manifold data when those mechanisms are introduced [15, 22].

The improved GNG introduces the CHL rule and an edge-aging mechanism to preserve the formation of topology and form a relationship between the two nearest nodes to the current input. As shown in Fig. 1, there is a cross between connection i,h and connection j,k. The improved GNG deletes the unreasonable connection i,h using the CHL rule and edge aging mechanism, and establishes a new connection i,j. The comparison of the original GNG and the improved GNG with topological structure is shown in Fig. 2.

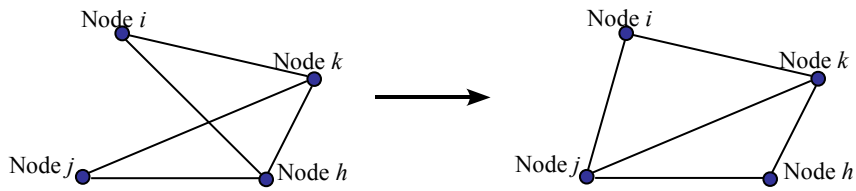


Fig. 1: The updating process of a topological structure in the improved GNG. With the CHL rule and edge-aging mechanism, the connections become reasonable: some crosses and unexpected edges will be deleted adaptively in the topology construction, and some related nodes will be connected.

3. Growing LLE Algorithm

Although, there are only two parameters, the intrinsic dimension and the neighborhood number, to be preselected in the original LLE algorithm, they must be identified through experimental estimation or by minimizing the reconstruction error. Because LLE computes an embedding for the training points obtained from the principal eigenvectors of a symmetric matrix, it can hardly project the embedding manifold into affine subspace without distortion. These limitations, as well as the time complexity, restrict the applications of LLE [18, 19]. In this paper, the improved GNG with topological connections is introduced into the original LLE method, resulting in a new algorithm named growing locally linear embedding. This section will describe the improvement in GLLE, including the selection of the two parameters, compare the time consumption and analyze the reconstruction error of GLLE. In the following analysis, it can be seen that the proposed algorithm can not only identify the two parameters adaptively, but also reduce the time consumption and memory usage. GLLE improves the the original LLE algorithm on adaptability and applicability.

The proposed GLLE is an unsupervised nonlinear manifold learning method that comprises six steps:

Algorithm 2 The GLLE algorithm.

Input: *Data matrix* $X=\{X_1, X_2, \ldots, X_N\}$ *with* N *samples on the manifold.*

Output: *Reduced data matrix* Y^L *and topological link matrix* D_L *.*

1. Compute the topological structure in the training of data matrix X *by the improved topological GNG that gives the reduced samples* X^L *and topological distance matrix* D_L *, where* D_L *defines* K *and* d *used in LLE adaptively. By self-organizing the input samples a new topological structure will be formed, with reduced samples* X^L *and edges between them, covering the manifold with their Voronoi field.*

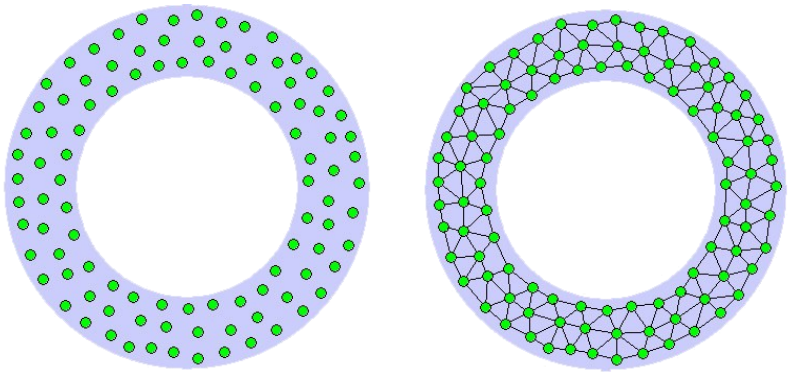
2. Compute the weights W'_{ij} *that best reconstruct each data point* X^L *from the linked neighbors as* D_L *gives, with reconstruction errors are measured by the cost function*

$$\varepsilon(W')=\sum_i |X_i^L-\sum_j W'_{ij}X_j^L|^2 . \tag{4}$$

3. Compute the vectors Y^L *best reconstructed by the weights* W'*, minimizing equation*

$$\phi(Y^L)=\sum_i |Y_i^L-\sum_j W'_{ij}Y_j^L|^2=\mathrm{tr}(Y^{LT}M'Y^L) , \tag{5}$$

where $M'=(I-W')^T(I-W')$*,* $M'\in R^{N\times N}$ *and* $N_w=\mathrm{rank}(X^L)$ *is the number of reduced samples.*



(a) (b)

Fig. 2: Mapping results of GNG and the topological GNG: (a) mapping result of the traditional GNG with no links and (b) mapping result of the improved GNG with topology connectivity.

4. With the constraints, it is possible to show that this minimization problem reduces to solving

$$\arg \min_{Y^L, (Y^L)^T Y^L = N I} (Y^L)^T M' Y^L \tag{6}$$

using the Lagrange multiplier method, where the Lagrange function now has the form

$$L_2(Y^L, \lambda) = \mathrm{tr}((Y^L)^T M\ Y^L) + \mathrm{tr}(\Lambda (Y^L)^T Y^L - N \times I)\ . \tag{7}$$

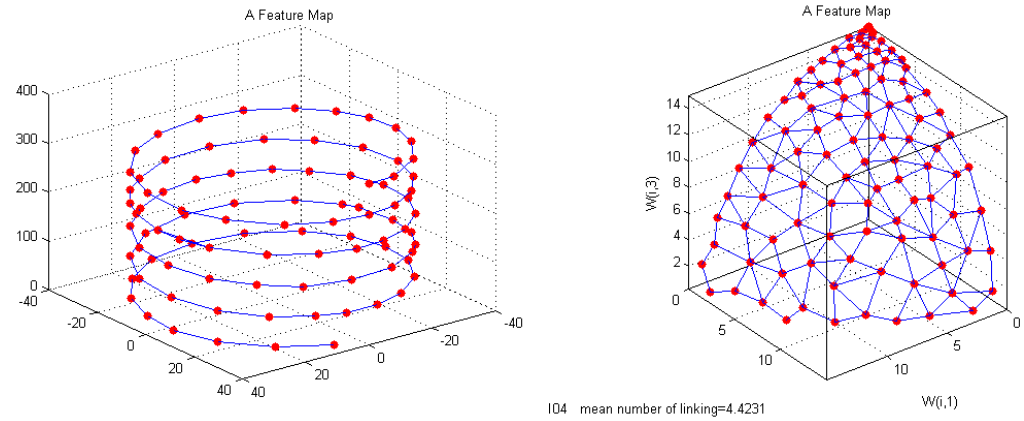
Matrix Λ contains the Lagrange multipliers as its diagonal elements. According to the other constraint $\sum_i Y_i^L = 0$, we discard eigenvector with the smallest eigenvalue.
5. Construct a symmetric matrix $\tilde{M} = I - M'$, with $\tilde{M}_{ij} = W'_{ij} + W'_{ji} - \sum_k W'_{ki} W'_{kj}$. Then the selection of the bottom d non-zero eigenvalues and eigenvectors of M' is transformed into the selection of the top d ones of $\tilde{M}$ except 1.
6. Compute the d-dimensional projection data matrix Y^L that minimizes the embedding error.

3.1 Estimation of Intrinsic Dimensionality

The estimation of the intrinsic dimensionality is a precondition problem that dimensionality reduction should envisage. The estimation of intrinsic dimensionality has been studied for years and some results have been reported, for example, using maximum likelihood estimators and eigenvalues of the covariance matrix [27, 28]. However, these methods are experimental or based on identifying the residual variance curves after redundant computation [10, 20], which are just preprocessing procedures, and not suitable for nonlinear dimensionality reduction.

In GLLE, the intrinsic dimensionality can be estimated by the formed stable topological structure. Because CHL rule is introduced, GNG is not only capable of revealing the probability distribution, but also mapping the geometric features, furthermore, forming an optimal topological structure [17]. With Monte-Carlo experiments, we can obviously find the relationship between connection and dimension. On average, each node connects to other 2 nodes in 1-dimensional space, while in 2-dimensional space each node connects to 4 nodes (GridTop). Analogically, there are similar results in other integer dimensional spaces.

We have conducted multiple virtual manifold data experiments with the intrinsic dimensionality $0 < d < 40$. From the Monte-Carlo statistical analysis results shown in Fig. 3, it can be seen that topological structures are truly capable of estimating the dimensionality. The



(a) (b)

Fig. 3: Estimation of the intrinsic dimensionality by topological connection: (a) one-dimensional mapping of two-dimension and (b) two-dimensional mapping of three-dimension.

connectivity number in the one-dimensional manifold is significantly different from that in the two-dimensional manifold, as shown in Fig. 3(a) and (b). The variance of the connectivity number is small enough to allow accurate estimation of the manifold dimension, as shown in Fig. 4. It is practical to deduce the intrinsic dimensionality of the input manifold data by formed topology. The dots in Fig. 4 designate the average values of the network edges in manifolds of various dimensionality, with the variance bar at each point is also shown. After self-organizing the manifold data, GNG can expediently estimate its intrinsic dimensionality, which can identify the parameter of embedding dimensionality. Therefore, the intrinsic dimensionality can be estimated by computing the average connections of each node in the topology.

3.2 Dynamic Selection of the Number of Nearest Neighbors

The selection of the number of nearest neighbors K is equally important in the original LLE algorithm. If K is too large, the LLE algorithm will ignore local nonlinear features on the manifold, as the traditional PCA does. If K is too small, LLE will split a continuous manifold into detached pieces, because the global features are lost. The selection of the appropriate value of K is another key of dimensionality reduction. Many research papers address this problem, for example [10, 21]. However, most approaches only explain the relationship between the number of nearest neighbors and the embedding dimensionality, rather than provide a procedure for the

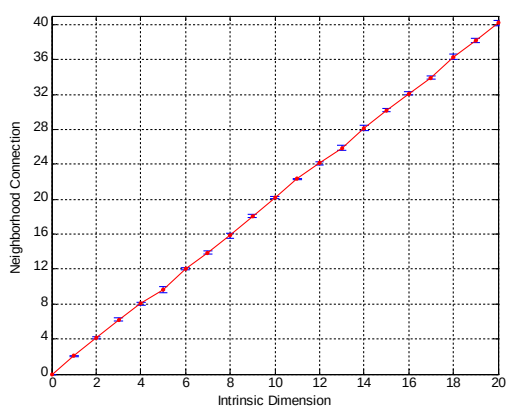


Fig. 4: Relationship between connections (y-axis) and intrinsic dimensionality (x-axis).

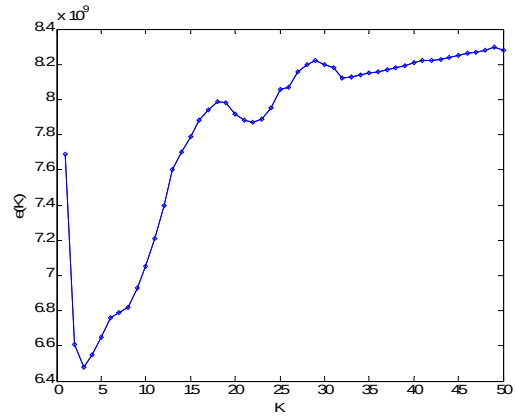


Fig. 5: Neighbor-error curve and the optimal value of K.

selection of optimal K. With the CHL rule [17], the proposed GLLE algorithm determines an optimal neighborhood for each node, which optimizes the local linearity.

For a uniform distribution of input vectors, the optimal value of the parameter K should minimize the residual variance:

$$K_{\text{opt}} = \arg\min_{K} \left(1 - \rho^2_{D_X D_Y}\right), \tag{8}$$

where D_x and D_y are the Euclidean distance (between pairs of points) matrix of X and Y separately, and ρ is the correlation coefficient. Theoretically, less residual variance leads to better a embedding effect. Because the Euclidean distance is not sufficient in a high dimensional space, the reconstruction errors go up and down along with K values.

As illustrated in Fig. 5, a global minimum of parameter K exists for $K<5$. Such behavior of the residual variance curve is explained by the fact that for the majority of points, first few neighbors of a point are close to that point and, hence, the addition of a new neighbor decreases the reconstruction error. However, as K continues to increase, the reconstruction error starts to increase as well and then begins to alternate (rise and fall), because the Euclidean distance is no longer a reliable indicator for proximity.

After ten statistical experiments, the number the topological neighbors K in the proposed algorithm is close to K_{opt}. Moreover, thanks to the adaptive selection by the neighborhood connectivity, GLLE achieves an error smaller than that of the algorithm with fixed K_{opt}.

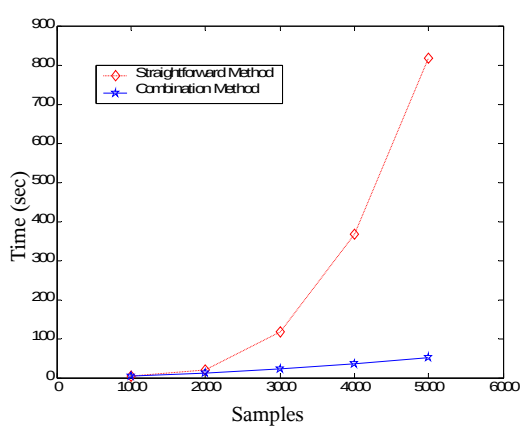


Fig. 6: Comparison of the time consumptions.

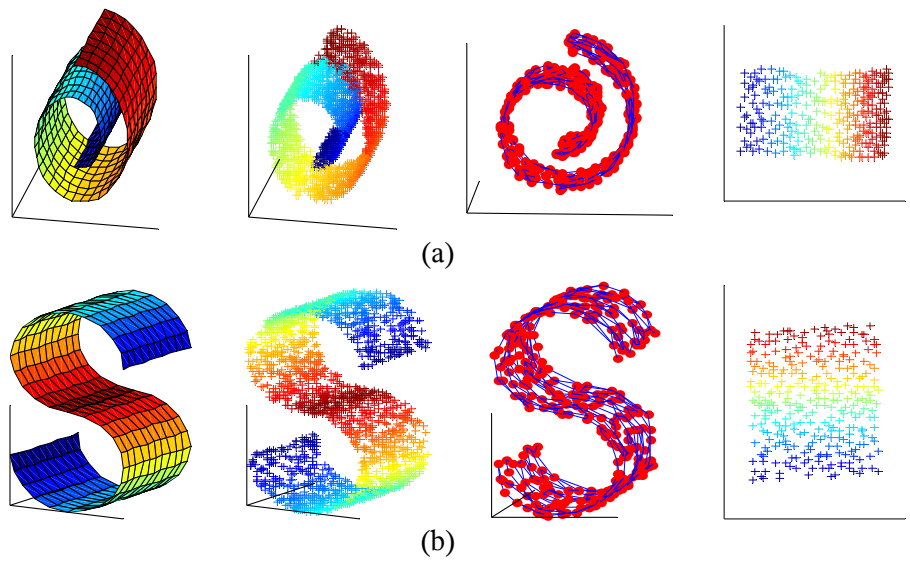


Fig. 7: Unraveling results of GLLE with the simulations of Swiss roll (a) and S-curve (b). The results are more regular and plain in the 2-D input space.

3.3 Reduced Time Complexity of GLLE

The original LLE computes all distances between each pair of nodes. Because the input space needs to be covered entirely, the initializing sample points should be distributed uniformly. This process consumes considerable memory space and time and lowers adaptability. After the self-organization step of GNG, the improved algorithm covers the support field of the manifold with fewer vectors while preserving the most features, and reduces the time and space requirements greatly.

The time complexity of the individual steps of the original LLE [10] is as follows. Searching for neighbors is $O(DN^2)$, computing the reconstruction weights is $O(DNK^4)$, and computing the minimal eigenvalues is $O(KdN^3)$, where N is the number of the initializing samples, D is the dimensionality of the high dimensional space, and K is the number of nearest neighbors.

The time complexity of the proposed GLLE is as follows. Self-organizing time consumption of

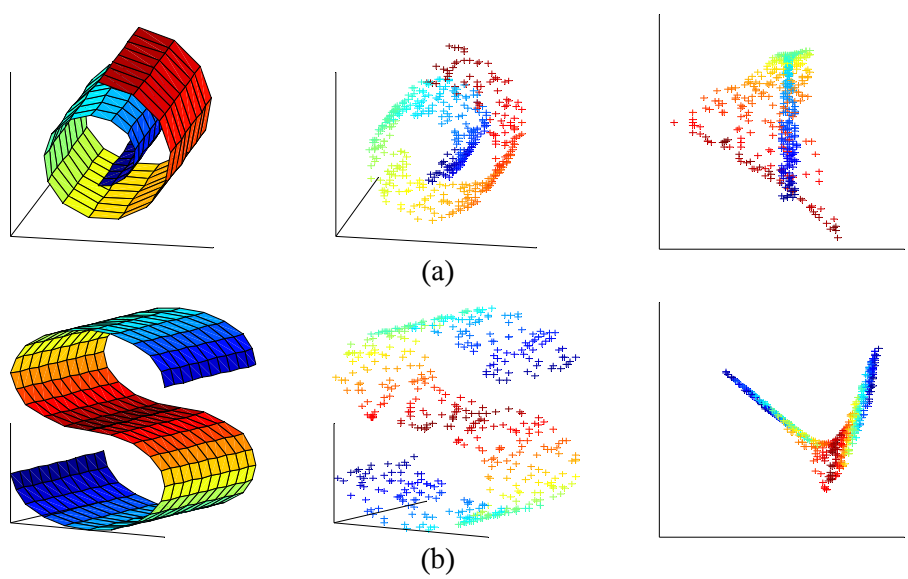
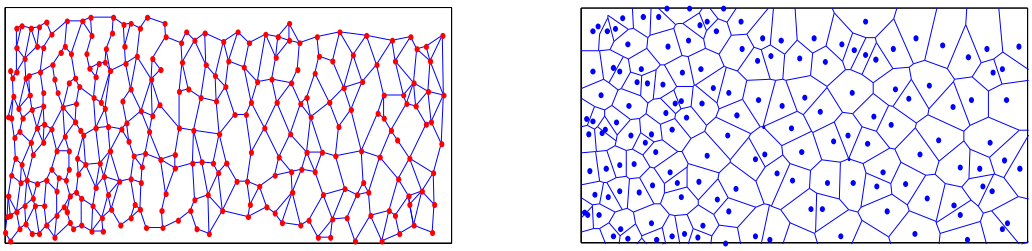


Fig. 8: Results for Swiss roll (a) and S-curve (b) with fewer samples Nt=500.



(a) (b)

Fig. 9: The unraveling topological structure in 2-D observed space (a) and the Voronoi region of the GLLE nodes (b) covering the support field of the whole Swiss Roll manifold in the Fig. 7 (a).

GNG, including training time, searching for nearest neighbors and updating the weights is $T_N O(N_w \log(N_w)) + O(D N_w^2)$, computing the reconstruction weights is $O(D N_w K_{\text{mean}}^4)$, computing the minimal eigenvalues is $O(K_{\text{mean}} d N_w^2)$, where N_w is the number of nodes after self-organizing, K_{mean} is the mean value of the number of nearest neighbors, and T_N is the training time.

Because of $N_w = N$ and $K_{\text{mean}} = K_{\text{opt}}$, the time consumption of GLLE is much less than that of LLE. The time consumption of the original LLE algorithm increase polynomially with the increasing number of the initializing samples, while that of GLLE increases linearly. The comparison of the time consumption between the two algorithms is shown in Fig. 6, with the computations performed on a PC with an AMD Athlon XP 2500+ processor, Matlab 6.5 and Windows XP SP2.

4. Performance Analysis of GLLE

This section reports the results of a performance analysis of LLE and GLLE. Some typical manifold datasets, such as Swiss roll and S-curve [8, 19], are used in this analysis. A significant difference between the nonlinear dimensionality reduction methods is observed. Additionally, a manifold of multi-pose and multi- expression faces of a single person is analyzed.

4.1 Dimensionality Reduction with Topology Preservation

In the performance analysis, the smooth sub-manifolds of a Swiss roll and an S-curve embedded into the 3-dimensional space are used. The unraveling results of the original LLE were reported in [8, 19]. It can be seen that the manifolds are unfolded flat, but some parts are compressed too much. GLLE covers the support fields of the manifolds by self-organizing learning, and the results without any contractive instances are better than those for the original LLE, as shown in Fig. 7 with the training times $T_N = 5$.

When there are not enough samples to overlay the manifolds, as shown in Fig. 8, some holes and breaks appear in the manifolds. The condition of the local linearity is lost; therefore, LLE cannot perform successfully. The unsuccessful unraveling results are shown in Fig. 8. In contrast, GLLE can unravel the manifold data correctly. There are topological connections in the network after self-organizing learning, and the nodes cover the entire manifold with their Voronoi fields, as shown in Fig. 9. That is to say, the manifold can be described as the neural nodes spatially.

4.2 Visualization of High-dimensional Data

LLE and other manifold learning methods have been applied to multi-pose face recognition and

Fig. 10: Typical images in the Frey face dataset.

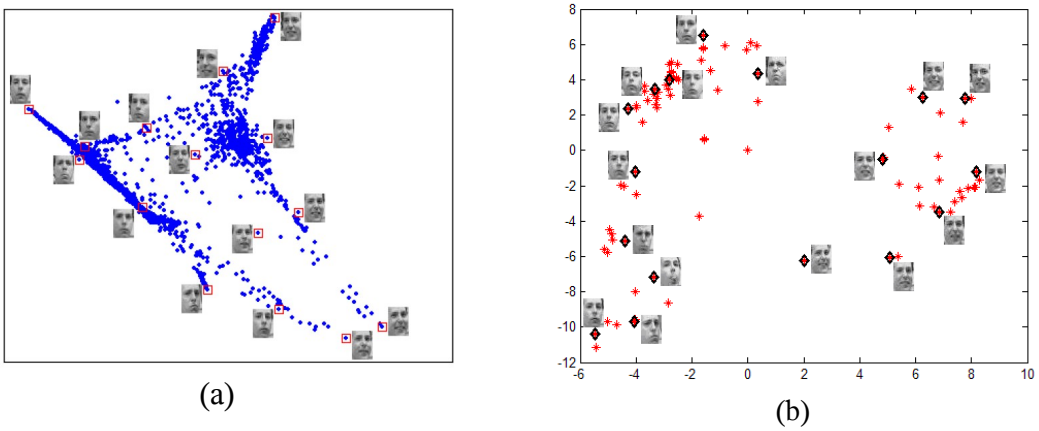


Fig. 11: Simulation results of Frey face dataset by LLE (a) and GLLE (b).

video-based face tracking with satisfactory results [23]. However, some problems still remain unsolved. For example, the original LLE is not applicable to non-convex manifolds and to manifolds with irregular appearance. In fact, uniform distribution of multi-pose face sequences on a manifold is impossible without presence of holes and noise. The GLLE algorithm has solved this problem. The following experiment demonstrates that GLLE preserves the arranging directions, while LLE cannot unfold the face sequences completely and some superposition appears.

The Frey face dataset (available at http://www.cs.toronto.edu/~roweis/data.html) is used in this example. The dataset contains 1965 face images. Each image is a 20×28 pixel gray-scale picture. Some typical faces are shown in Fig. 10. Rearranging the pixels of each image into a vector with dimension $D=560$, the face sequences are transformed into 2-D matrix X. The simulation results mapped by LLE and GLLE are shown in Fig. 11. The result of the original LLE is hardly satisfying for its superposition and holes, with some different faces mapped into the same vicinity. The result of GLLE is quite satisfying with the multi-pose face changing directions as shown in the figure clearly. The face images are mapped to cover the field more uniformly, with partition of the open mouth faces and the closed mouth faces. The results of GLLE can reflect the intrinsic properties, i.e. pose and expression, better than LLE. The mapping result of all 1965 face images by LLE is shown in Fig. 11 (a).

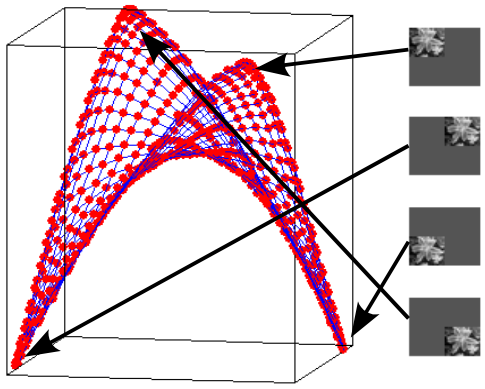


Fig. 12: Dimensionality reduction result of high dimensional images by moving a lily image (from Matlab images) on a noisy background.

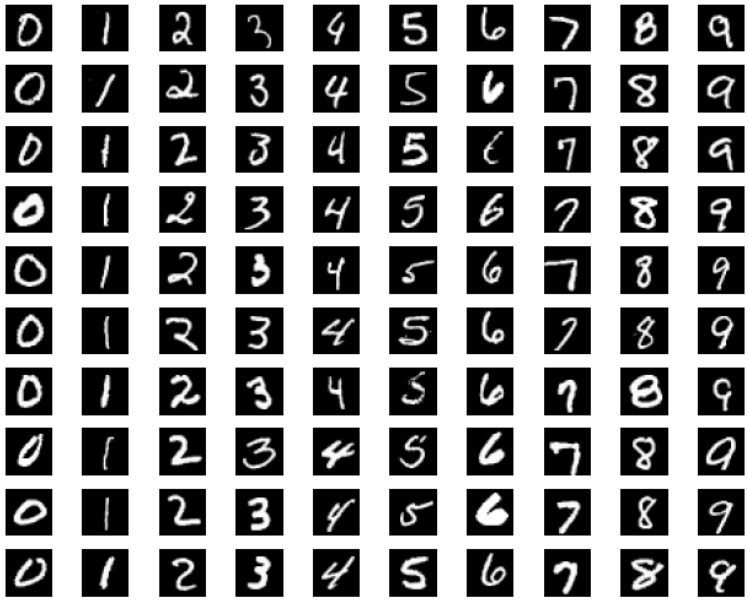


Fig. 13: Digit samples in MNIST data set. In this experiment we only use digits 4, 7 and 9.

Assuming the embedding error $\epsilon \to 0$ in GLLE, the reconstruction face is obtained as follows:

$$Y_i = \sum_j W_{ij} Y_j \,. \tag{9}$$

Because the proposed algorithm reflects the probability distribution, the reconstructed faces may not be the exactly the true faces. That is to say, we reconstruct the faces linearly with the nearest neighbors in this experiment. The results are shown in Fig. 11 (b).

In order to identify the topology preservation, we select a lily image with 23×19 gray pixels and move it on a noisy background, forming 480 gray images with 46×38 dimension, as samples on four corners shown in Fig. 12 on the right side. From the saddle surface result we observe that the manifold learning algorithm preserves the topological structure of the images distributing in high dimensional space. The structure after dimensionality reduction using PCA will be distorted [5].

5. Simulation Results of GLLE

In this section, several experiments are carried out to demonstrate the proposed GLLE method for pattern representation and recognition on the MNIST dataset [33] and the ORL face dataset [34]. The simulation results of the accuracy and time consumption are presented along with the results for widely used traditional algorithms such as PCA [1], LLE, ISOMAP [4] and LPP [31]. It should be noted that, since the focus in this paper is on feature representation, we use the K-nearest neighbor (K-NN) algorithm [30] as the classifier.

5.1 MNIST Digit Classification

The MNIST database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. The digits have been size-normalized and centered in 28×28 pixels gray-scale images. The dataset comprises handwriting digit 0-9. Some sample images from the dataset are shown in Fig. 13. It is a challenge in these experiments to select the digits 4, 7, 9 as the input samples, because of the small distinctions between them. In the experiments, the training sets are randomly selected from the MNIST training set, containing 100 images of each digit. The test

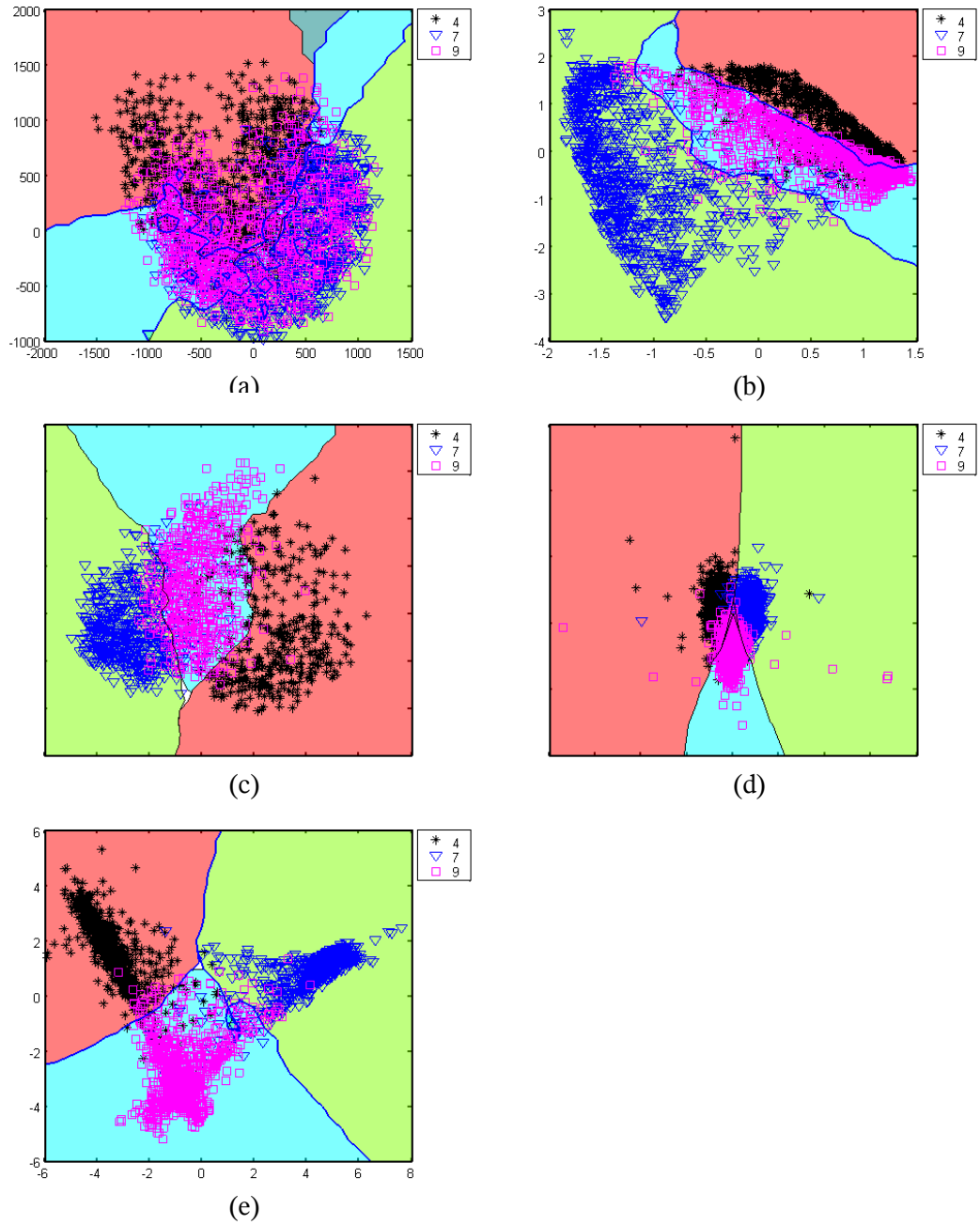
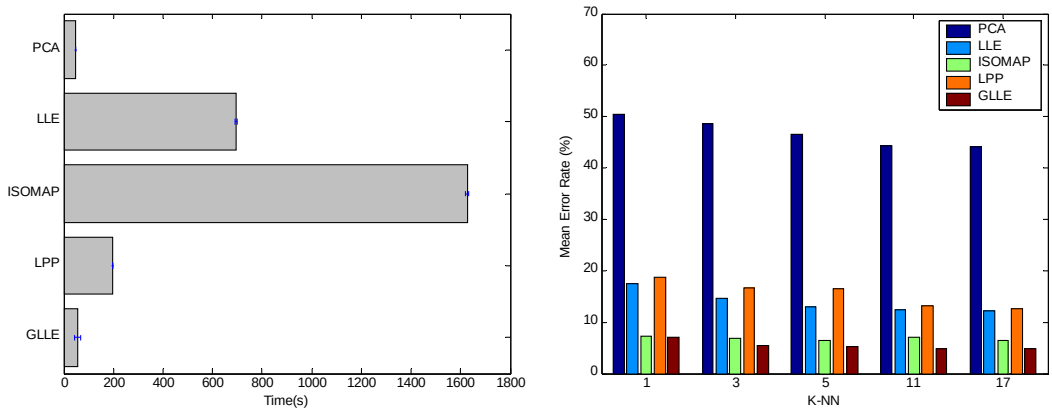


Fig. 14: The classification results of digit 4, 7, 9 in MNIST dataset by the five considered algorithms and K-NN classifier: (a) PCA; (b) LLE; (c) ISOMAP; (d) LPP; and (e) GLLE.

sets are all images of digits 4, 7, and 9 in the MNIST test set. K-nearest neighbor (K-NN) algorithm is used as the classifier, with various K values of 1, 3, 5, 11, and 17.

Every algorithm has been run 10 times, and the mean accuracy and the standard deviation are summarized in Table 1. The results of 2-D visualization by each algorithm are shown in Fig. 14, and the time consumptions and error rates are shown in Fig. 15. The results showed the superiority of the proposed GLLE algorithm. The GLLE outperforms the original LLE, on both classification and time consumption. In fact, GLLE gives the lowest error rate among the considered methods on this data set when a K-NN classifier is used. Fig. 15 (a) demonstrates large time consumption for the original LLE and ISOMAP algorithm. LLE and ISOMAP require



(a) (b)

Fig. 15: Comparisons in time consumptions and mean error rates: (a) time consumptions and (b) mean error rates.

much time to construct the large scale adjacency graph. The time consumption of GLLE in dimensionality reduction is similar to that of the linear algorithms, because GLLE reduces the number of nodes by the improved GNG effectively.

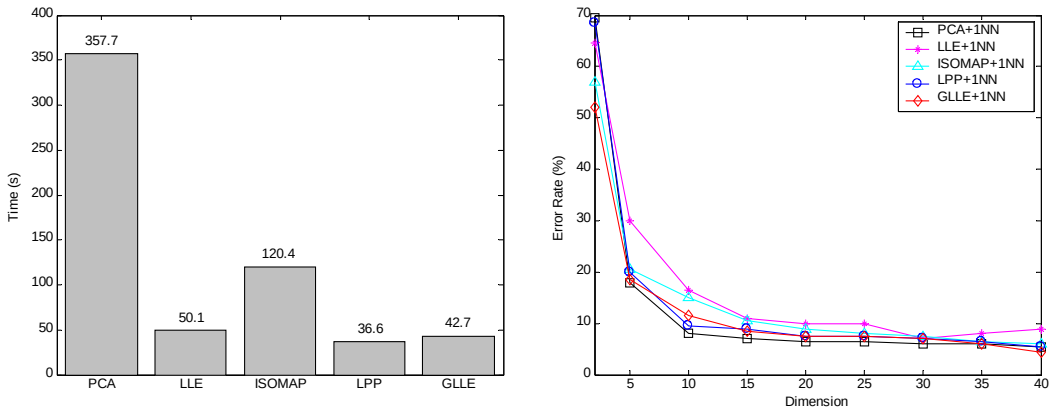
Table 1: Classification accuracies (%) of various reduction algorithms.

K-NN	PCA	LLE	ISOMAP	LPP	GLLE
1	49.45(0.99)	82.58(0.56)	92.83(3.85)	81.35(0.86)	93.00(0.58)
3	51.42(0.75)	85.43(0.43)	93.24(3.09)	83.24(0.66)	94.51(0.40)
5	53.42(0.67)	86.98(0.33)	93.59(2.50)	83.60(0.61)	94.82(0.45)
11	55.60(0.72)	87.57(0.33)	92.87(2.36)	86.86(0.52)	95.14(0.42)
17	55.93(0.53)	87.82(0.41)	93.64(2.57)	87.33(0.35)	95.29(0.40)

5.2 ORL Face Recognition

This part uses the Cambridge Olivetti Research Lab (ORL) face database. The dataset contains 400 face images of 40 individuals. Each individual has 10 gray images with size 92×112. The face images are centralized, with indiscoverable lighting variations, so the major challenge in this data set is the variation of pose and facial expression. In the experiments, the ORL datasets are partitioned into two non-overlapping subsets: 200 training images and 200 test images. Sample images are shown in Fig. 16.

Fig. 16: Face samples in the ORL face dataset.



(a) (b)

Fig. 17: The comparisons of time consumptions and the error rates of different algorithms on the ORL dataset: (a) time consumptions and (b) error rates. Because the dimension of the input space (92×112) is much more greater than the number of samples (400), the time consumption of PCA is much higher than that of the others.

K-NN with $K=1$ was found to perform better, and there is a clear trend of accuracy decreasing with increasing K. That is to say, the error rate of 1-NN classifier is lower than all the others. Similar observation was also made in previous studies, e.g. in [35]. Therefore, in this experiment the 1-NN classifier is used.

Table 2 shows the classification accuracy results for different dimensions of feature spaces on the ORL data sets. The accuracy of GLLE is 95.5% on 40-D vectors, much better than that of LLE. In this data set the number of samples (400) is much smaller than the dimension of image space (92×112); therefore, the manifold learning methods, such as LLE, ISOMAP, LPP and GLLE, construct the adjacency graph quickly. Since PCA needs to compute the covariance matrix of size 1030×10304, the time consumption is much larger, as shown in Fig. 17.

From the experimental results it can be seen that the GLLE algorithm achieves the best classification accuracy of all the tested algorithms. When the number of samples increases, both the accuracy and the time consumption increase. Considering the classification rate and time consumption, the proposed GLLE algorithm is an effective unsupervised nonlinear dimensionality reduction method.

Table 2: Classification accuracies (%) of various dimensional spaces.

Dimension	PCA	LLE	ISOMAP	LPP	GLLE
2	30.50	35.50	43.00	31.50	48.00
5	82.00	70.00	79.50	80.00	81.50
10	92.00	83.50	85.00	91.50	88.50
15	93.00	89.00	89.50	93.00	91.50
20	93.50	90.00	91.00	92.50	92.50
25	93.50	90.00	92.00	92.50	92.50
30	94.00	93.00	92.50	93.00	93.00
35	94.00	92.00	93.50	93.50	94.00
40	94.50	91.00	94.00	94.50	95.50

6. Conclusions and discussions

With various improvements, the LLE algorithm can perform the dimensionality reduction better and better, and its applications are wider and wider in the visualization of high

dimensional data and pattern recognitions [10, 11]. In this paper an improved growing locally linear embedding algorithm is proposed, by introducing the CHL rule to preserve the topological structures in self-organizing learning, based on the traditional LLE and GNG. The proposed algorithm compresses the redundant information in manifolds and preserves most intrinsic properties at the same time. The proposed GLLE overcomes the main shortcomings [10] of the original LLE algorithm, stimulating the applications of LLE. In manifold learning, it is assumed that the input samples are lying on a smooth convex manifold, and then the current LLE is applied. If the manifold is not smooth, or there are some out-of-sample inputs [29], the results will face the probability of collapse. GLLE avoids this limitation by giving a judgment before unraveling by global topology preservation. Furthermore, the highlight of global preservation can also illuminate ISOMAP [4] as the selection of landmarks.

Because the bottom minimal eigenvalues have been selected in numerical computations, LLE is sometimes not stable. In numerical computations, because the selected minimal eigenvalues are very close to 0, it is easy to induce singular eigenvalues and eigenvectors, which may introduce disturbances in the embedding processes and lead to deformed unraveling of a few samples or a sparse matrix. When the new matrix of selection the largest eigenvalues is introduced, the proposed GLLE can solve this problem satisfactorily, and increase the robustness of the algorithm.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive comments. This work is supported by National Nature Science Foundation (Grant No. 30370416, 60575044, 60675005), National Distinguished Young Scholars Fund of China (Grant No. 60225015), and Specialized Research Fund for the Doctoral Program of Higher Education of China (20049998012).

References

[1] I. T. Jolliffe, *Principal Component Analysis.* New York, NY: Springer-Verlag, 2002.

[2] I. Borg, P. Groenen, *Modern Multidimensional Scaling: Theory and Applications.* New York, NY: Springer, 1996.

[3] S. Balakrishnama and A. Ganapathiraju. (1998). Linear discriminant analysis. *Institute for Signal and Information Processing, Mississippi State University* [Online]. Available: http://www.isip.msstate.edu/publications/reports/isip_internal/1998/linear_discrim_analysis/lda_theory.pdf

[4] M. Balasubramanian, E. L. Schwartz, J. B. Tenenbaum, V. de Silva, and J. C. Langford, "The Isomap algorithm and topological stability," *Science*, vol. 295, no. 5552, pp. 7, 2002.

[5] L. K. Saul and S. T. Roweis. (2001). An introduction to locally linear embedding. [Online]. Available: http://www.cs.toronto.edu/~roweis/lle/papers/lleintro.pdf

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

[7] T. Kohonen, *Self-Organizing Maps*. 3rd Edition, Berlin: Springer-Verlag, 2001.

[8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[9] D. de Ridder and R. P. W. Duin, "Locally linear embedding for classification," Pattern Recognition Group, Dept. of Imaging Science & Technology, Delft University of Technology, Delft, The Netherlands, Tech. Rep. PH-2002-01, 2002.

[10] O. Kouropteva, O. Okun, and M. Pietikäinen, "Selection of the optimal parameter value for the locally linear embedding algorithm," in *Proc. 1st Int. Conf. Fuzzy Systems and Knowledge Discovery*, Singapore, 2002, pp. 359-363.

[11] D. de Ritter, O. Kouropteva, O. Okun, M. Pietikäinen, and R. P. W Duin, "Supervised locally linear embedding," in *Proc. Int. Conf.. Artificial Neural Networks and Neural Information Processing*, Springer LNCS: 2714, 2003, pp. 333-341.

[12] O. Kouropteva, O. Okun, and M. Pietikäinen, "Supervised locally linear embedding algorithm for

pattern recognition," in *Proc. 1st Iberian Conference on Pattern Recognition and Image Analysis*, Spain, Springer LNCS: 2652, 2003, pp. 384-394.

[13] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Applied Mathematics*, vol. 100, no. 10, pp. 5591-5596, 2003.

[14] B. Fritzke, "A growing neural gas network learns topologies," *Advances in Neural Information Processing Systems 7*, Cambridge MA: MIT Press, pp. 625-632, 1995.

[15] S. Chen, Z. Zhou, and D. Hu, "Diffusion and growing self-organizing map: a nitric oxide based neural model," in *F. Yin, J. Wang, and C. Guo (Eds.) ISNN 2004: LNCS 3174*, pp. 199-204, 2004.

[16] B. Fritzke, "Growing cell structures – a self-organizing network for unsupervised and supervised learning," *Neural Networks*, vol. 7, no. 9, pp. 1441-1460, 1994.

[17] T. M. Martinetz, "Competitive Hebbian learning rule forms perfectly topology preserving maps," in *Proc. ICANN'93*, Amsterdam, 1993, pp. 427-434.

[18] O. Kouropteva, O. Okun, and M. Pietikäinen, "Classification of handwritten digits using supervised locally linear embedding algorithm and support vector machine," in *Proc. 11th European Symp. Artificial Neural Networks*, Belgium, 2003, pp. 229-234.

[19] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319-2323, 2000.

[20] D. Hundley and M. Kirby, "Estimation of topological dimension," in *Proc. 3rd SIAM Int. Conf. Data Mining*, California, 2003, pp. 194-202.

[21] M. Meila and J. Shi, "Learning segmentation by random walks," *Advances in NIPS 13 (NIPS2000)*, Cambridge, 2001, pp. 873-879.

[22] J. Yin, D. Hu, S. Chen, and Z. Zhou, "DSOM: a novel self-organizing model based on NO dynamic diffusing mechanism," *Science in China (Series F)*, vol. 48, no. 2, pp. 247-262, 2005.

[23] C. Zhang, J. Wang, N. Zhao, and D. Zhang, "Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction," *Pattern Recognition*, vol. 37, no. 2, pp. 325-336, 2004.

[24] J. Xiao, Z. Zhou, D. Hu, J. Yin, and S. Chen, "Self-organizing locally linear embedding for nonlinear dimensionality reduction," in *L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3610*, pp. 101-109, 2005.

[25] L. Xu, "Rival penalized competitive learning, finite mixture, and multisets clustering," in *Proc. IEEE Int. Joint Conf. Neural Networks*, 1998, pp. 2525-2530.

[26] J. Bruske and G. Sommer, "Intrinsic dimensionality estimation with optimally topology preserving maps," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 572-575, 1998.

[27] E. Levina and J. P. Bickel, "Maximum likelihood estimation of intrinsic dimension," in *Proc. Neural Information Processing Systems: NIPS*, Vancouver, CA, Dec. 2004.

[28] P. J. Verveer and R. P. Duin, "An evaluation of intrinsic dimensionality estimators," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp. 81-86, 1995.

[29] Y. Bengio, J. F. Paiement, and P. Vincent, "Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering," *Advances in Neural Information Processing Systems 16 (NIPS2003)*, Cambrige, MA. MIT Press, 2004.

[30] R. O. Duda, P. E. Hart, and D. Stork, *Pattern Classification*. John Wiley & Sons, Inc., 2000.

[31] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, 2005.

[32] M. H. Yang, "Extended Isomap for classification," in *Proc. 16th Int. Conf. Pattern Recognition*, vol. 3, 2002, pp. 615-618.

[33] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[34] ORL. (1992). The ORL face database at the AT&T (Olivetti) Research Laboratory. Available: http://www.uk. research.att.com/facedatabase.html

[35] J. P. Ye and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no.6, pp. 929- 941, 2005.