

Manifold Learning using Growing Locally Linear Embedding

Junsong Yin, Dewen Hu, *Senior Member, IEEE*, Zongtan Zhou

Abstract—Locally Linear Embedding (LLE) is an effective nonlinear dimensionality reduction method for exploring the intrinsic characteristics of high dimensional data. This paper mainly proposes a hierarchical framework manifold learning method, based on LLE and Growing Neural Gas (GNG), named Growing Locally Linear Embedding (GLLE). First, we address the major limitations of the original LLE: intrinsic dimensionality estimation, neighborhood number selection and computational complexity. Then by embedding the topology learning mechanism in GNG, the proposed GLLE algorithm is able to preserve the global topological structures and hold the geometric characteristics of the input patterns, which make the projections more stable and robust. Theoretical analysis and experimental simulations show that GLLE with global topology preservation tackles the three limitations, gives faster learning procedure and lower reconstruction error, and stimulates the wide applications of manifold learning.

I. INTRODUCTION

A central problem in machine learning and data mining involves the development of appropriate representations of complex data. Most real data lies 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 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. Typically, Principle Component Analysis (PCA) [1], Multi-Dimensional Scaling [2], Linear Discriminant Analysis (LDA) [3], and etc. are linear dimensionality reduction methods, while Isometric Maps (ISOMAP) [4], Locally linear Embedding (LLE) [5], Laplacian Eigenmaps [6], Self-Organizing Maps [7], and etc. are nonlinear dimensionality reduction methods. All of those operations can reduce the redundancies and retain the primary characteristics. Of these, LLE is an effective nonlinear dimensionality reduction algorithm, proposed first by Roweis in 2000 [8]. Compared with others, the notable advantages of LLE are: i) only 2 parameters have to be defined; ii) it can reach global minimization of the reconstruction error while

avoiding plunging into a local extremum.

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 have to be set are the intrinsic dimension d and the neighborhood number K , which greatly influence the results obtained. On the one hand, for the intrinsic dimension d , a high value magnifies noise effects, while a low value leads to overlaps in mapping results (excessively reduced). On the other hand, for the neighborhood number K , a lower neighborhood number cannot make reconstructions reveal the global characters of original data, while a larger neighborhood number causes a manifold to lose a nonlinear feature and to behave like traditional PCA [9]. Furthermore, it is found in our research that LLE is sensitive to initial data amount [24]. A lack of data cannot cover the support fields of manifold when local characteristics are lost, while excessive data will lead to incomplete reconstruction (also relative to the neighborhood number) and heavy computational assumption.

Previous researchers have proposed a few improved LLE algorithms. One algorithm gives the selection of the optimal parameter K value [10], one gives a supervised algorithm SLLE [11], [12], and another gives a substitute algorithm HLLE based on Hessian Eigenmaps [13]. These algorithms do reduce the learning time, enhance the space partition ability, and reduce reconstruction error but cannot make LLE an adaptively optimal map, and are also restricted when a manifold is noised.

We propose an algorithm called Growing Locally Linear Embedding (GLLE), which 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]. What's more, the number of neural nodes in neural network is fewer than that of the input pattern. That is to say, our proposed GLLE algorithm is capable of identifying the two parameters adaptively, reducing the time consumption by decreasing the samples reasonably and preserving the global topological and geometrical structures. It can be said that GLLE has greatly improved the applicability of LLE in nonlinear dimensionality reduction. The 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. On the same way, a linear transformation on spectrum matrix

Junsong Yin is with National University of Defense Technology, Changsha, Hunan, 410073, P.R. China (corresponding author to provide phone: 867314574977; fax: 867314574992; e-mail: jsyin@nudt.edu.cn).

Dewen Hu is with National University of Defense Technology, Changsha, Hunan, 410073, P.R. China (e-mail: dwhu@nudt.edu.cn).

Zongtan Zhou is with National University of Defense Technology, Changsha, Hunan, 410073, P.R. China (e-mail: Narcz@163.com).

$\tilde{M} = I - M$ makes the selection of the lowest eigenvectors become the selection of highest ones with the same eigenvalues, which makes the algorithm more stable and robust.

Our research addresses the three limitations of LLE: intrinsic dimensionality estimation, selection of neighbors, and time consumption. Aiming at the limitations, the most contributions of this paper are to define the two parameters adaptively, decrease the computational complexity, preserve the global topology when locally embedding, and improve the robustness of the nonlinear dimensionality reduction algorithm.

The remainder of this paper is organized as follows. Section 2 reviews the original LLE and the topological version of GNG briefly. In Section 3 we analyze the three limitations of LLE and present the improved manifold learning algorithm GLLE. Section 4 provides the performance analysis of LLE and GLLE by comparison. Section 5 presents experimental results of handwritten digits and multi-pose faces. Finally, Section 6 offers our conclusions and discussions.

II. REVIEW OF THE LLE AND GNG

A. The LLE Algorithm

Locally Linear Embedding (LLE) was first proposed by Roweis and Saul in Science in 2000, and its primary idea is to reconstruct a nonlinear manifold by embedding a local linear hyperplane [8]. The major characteristics are that LLE is an unsupervised learning algorithm, can preserve the relationships between neighbors in manifold data and map the high dimensional data in low dimensional Euclidean space. LLE maps a data set $X = \{X_1, X_2, \dots, X_N\}, X_i \in \mathcal{R}^D$ globally to a low dimensional set $Y = \{Y_1, Y_2, \dots, Y_N\}, Y_i \in \mathcal{R}^d$, obeys $d < D$. The algorithm can be described in three steps:

(1) Find local neighbors of each point X_i using k -nearest neighbors.

(2) Compute constrained weights W_{ij} that best linearly reconstruct X_i from its neighbors. Reconstruction errors are measured by the cost function:

$$\varepsilon(W) = \sum_i \left| X_i - \sum_j W_{ij} X_{ij} \right|^2 \quad (1)$$

(3) Compute low dimensional embedding vectors $Y_i \in \mathcal{R}^d$ best reconstructed by W_{ij} minimizing equation

$$\phi(Y) = \sum_i \left| Y_i - \sum_j W_{ij} Y_{ij} \right|^2 \quad (2)$$

under the constraints $\sum_i Y_i = 0$ and $\frac{1}{N} \sum_i Y_i^T Y_i = I$. The projecting cost function can be revised as

$$\phi(Y) = \sum_i \left| Y_i - \sum_j W_{ij} Y_{ij} \right|^2 = \sum_i \left| (I - W) Y_i \right|^2 = \text{tr}(Y^T M Y) \quad (3)$$

where $M \in \mathcal{R}^{N \times N}$ and $M = (I - W)^T (I - W)$. Now the LLE embedding problems are transformed to compute the bottom d non-zero eigenvalues of matrix M , in order to obtain the homoeomorphism maps covering the manifold data in high dimensional space [9].

B. GNG and its Improvement

B. Fritzke [14], [16] proposed Growing Neural Gas (GNG) in 1995. As an unsupervised non-topology preservation self-organizing neural network, it initializes with random node numbers and little aprior information. By the dynamic node growing and dying mechanism, GNG can cluster and partition the input space satisfactorily. On the other hand, GNG can merely reflect the probability distribution of the input patterns, but cannot reveal the topological structure of the embedding manifold. Considering the properties of the original LLE, traditional GNG is improved on the topology preservation, which was inspired by Bruske et al [26]. Besides the growing mechanism in GNG, Competitive Hebbian Learning (CHL) rule [17] and edge-aging mechanism [15], similar to the Rival Penalized Competitive Learning [25], are introduced to model the constitution and update the topological structure. The improved GNG can preserve the geometric characteristics, 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. The comparison of the original GNG and the improved GNG with topological structure is shown in Fig. 1.

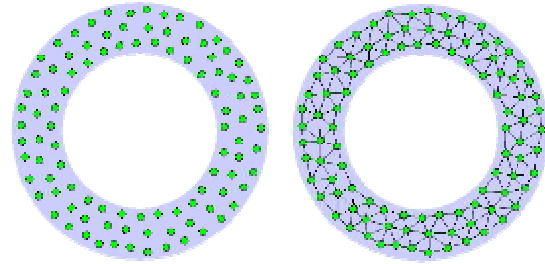


Fig. 1: Mapping results of GNG and the topological GNG. (a) mapping result of the traditional GNG with no links; (b) mapping result of the improved GNG with topological connections.

III. GROWING LLE ALGORITHM

Only two parameters, intrinsic dimension and neighborhood number, have to be pre-set in the original LLE, but the parameters can be identified only through experimental estimation or by reconstructing error. Because LLE computes an embedding for the training points obtained from the principal eigenvectors of a symmetric matrix, it can hardly project 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, and a novel algorithm named GLLE is proposed. This section will expatiate the improvement in GLLE, including the two parameters selection, comparison of the time consumption and analysis of 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, analyzed in this section. It can be said that GLLE improves the algorithm on adaptability and application.

The proposed GLLE, an unsupervised nonlinear manifold learning method introducing topological learning theory, can be cast in the following framework firstly.

Algorithm: GLLE algorithm

Input: Data matrix $X=\{X_1, X_2, \dots, X_N\}$ 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; give the reduced samples X^L and topological distance matrix D_L , where D_L defines parameter 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 Vorinoi 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:

$$\varepsilon(W') = \sum_i \left| X_i^L - \sum_j W'_{ij} X_{ij}^L \right|^2.$$

3. Compute the vectors Y^L best reconstructed by the weights W' , minimizing equation $\phi(Y^L) = \sum_i \left| Y_i^L - \sum_j W'_{ij} Y_{ij}^L \right|^2$.

As in (3) $\phi(Y^L) = tr(Y^{L^T} M' Y^L)$, where matrix $M' \in \mathcal{R}^{N_w \times N_w}$ is $M' = (I - W')^T (I - W')$. $N_w = \text{rank}(X^L)$, means the number of reduced samples.

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

$$\arg \min_{Y^L} Y^{L^T} M' Y^L \quad (4)$$

$Y^{L^T} Y^L = NI$

Using Lagrange Multiplier Method, where the Lagrange function now has the form

$$\mathcal{L}_2(Y^L, \lambda) = tr(Y^{L^T} M' Y^L) + tr(\Lambda Y^{L^T} Y^L - N \times I) \quad (5)$$

where the 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.

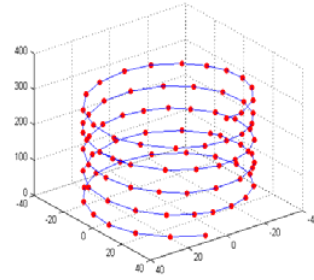
5. Construct a symmetric matrix $\tilde{M} = I - M'$, with

$\tilde{M}_{ij} = W'_{ij} + W'_{ji} - \sum_k W'_{ki} W'_{kj}$. Then the selections of bottom d non-zero eigenvalues and eigenvectors of M' are transformed into the selection of top d non-one ones of \tilde{M} .

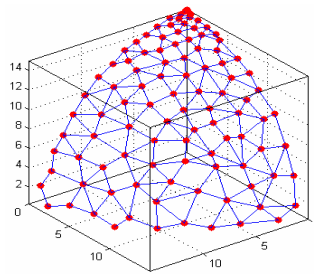
6. Compute the d -dimensional projection data matrix Y^L with minimizing the embedding error.

A. Estimation of Intrinsic Dimensionality

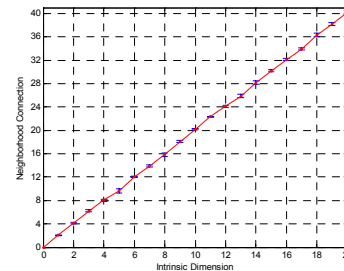
The estimation of intrinsic dimensionality is a precondition problem that dimensionality reduction should envisage. The estimation of intrinsic dimensionality has been studied for years and some results have been presented, using maximum likelihood estimator, eigenvalues of the covariance matrix, et al [27], [28]. However, those proposed methods basically are experimental or identifying the residential variance curves after redundant computation [10], [20], which are just preprocessing procedures, and not suitable for nonlinear dimensionality reduction.



(a)



(b)



(c)

Fig. 2: Estimation of the intrinsic dimensionality by topological connection. (a) One-dimensional mapping of two-dimension. (b) Two-dimensional mapping of three-dimension. (c) Statistical analysis of the relationship between connections (y-axis) and intrinsic dimensionality (x-axis).

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 an average each node connects to other 2 nodes in 1-dimensional space, while connects to 4 nodes in 2-dimensional space (GridTop). Analogically there are similar results in other integer dimensional spaces.

We conducted many virtual manifold data experiments with the intrinsic dimensionality $0 < d < 40$. From the Monte-Carlo statistical analysis results shown in Fig. 2, it can be obviously seen that topological structures are truly capable of estimating the dimensionality. The difference of connecting edges is much different from other dimensionalities, as shown in Fig. 2(a) and 2(b), and in the same dimensional space, the variance of connecting edges is small enough. It is practical to deduce the intrinsic dimensionality of the input manifold data by formed topology. Points in Fig. 2(c) mean the average values of the network edges in various dimensionalities, while the variance bar of 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. Thus it can be said the intrinsic dimensionality can be estimated by computing the average connections of each node in the topology.

B. Dynamic Selection of Neighborhood Numbers

Like the parameter of the dimension d , the selection of the neighborhood number K is also important to the original LLE. If the neighborhood number K is larger, the algorithm will ignore even lose the local nonlinear features on the manifold, just as the traditional PCA performs. In contrast, if the neighborhood number K is defined smaller, LLE will split the continuous manifold into detached locality pieces, because the global characteristics are lost. The selection of K value is another key of dimensionality reduction, and there have been a lot of papers giving various identifying methods [10], [21]. But these papers explain only the relationship of neighborhood number and embedding dimensionality: $K > d$, rather than give the optimal selection of K value, including the analysis of interior and edges of the manifold, according to the input topological spaces. Under CHL rule [17], the proposed GLLE algorithm defines the optimal neighborhood of each node to optimizing the local linearity.

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

$$K_{opt} = \arg \min_K (1 - \rho_{D_X D_Y}^2) \quad (6)$$

where D_X and D_Y are the Euclidean distance (between pairs of points) matrix of X and Y separately, and ρ is the standard linearly relative coefficient. Theoretically, less residential variance leads to better embedding effect. Because the

Euclidean distance is irresponsible, the reconstruction errors go up and down along with K value. For example, there is a global minimal value of parameter K when $K < 5$ in face dataset, which is because when K is little, all of its neighbors is in the vicinity. The K value increases, the error decreases, so K_{opt} is not in $K < 5$. Until $K=22$ the optimal neighborhood nodes are affirmed [10], as shown in Fig. 3.

By computations and statistical experiments the parameter K value of the topological neighbors in the proposed algorithm is according to the K_{opt} basically; moreover, for the adaptability of neighborhood selection, GLLE works with less error than fixed K_{opt} .

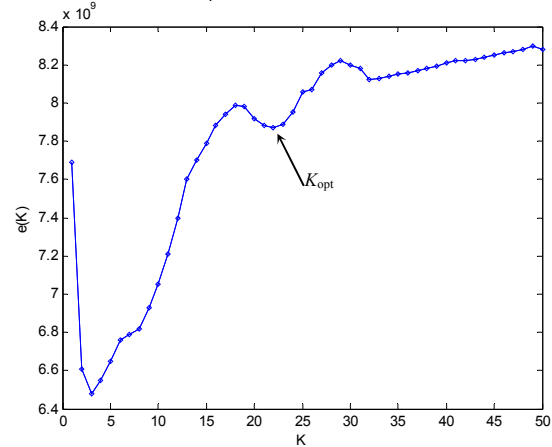


Fig. 3. Neighbor-error curve and the value K_{opt}

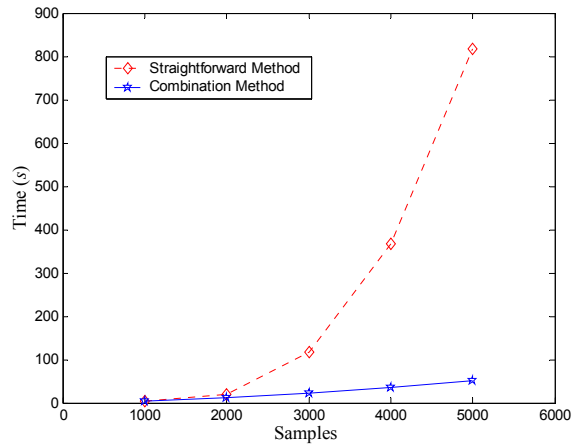


Fig. 4. Comparison of the time consumptions

C. Reduce the Complexity of Algorithm

The original LLE needs to compute all the distances between a couple of nodes. Because the input space needs to be cover wholly, the initializing sample points should be created to distribute uniformly. Thus it will consume considerable memorizing space and time, lowering its adaptability. After self-organization of GNG, the improved algorithm covers the support field of the manifold with fewer reference vectors, preserves the most properties and reduces time and space consumptions greatly.

i. Time complexity of the original LLE [10]. Parameter N is the number of the initializing samples, D is the dimensionality of the high dimensional space, and K is the neighborhood value. Thus the time consumption is described as follows.

- Searching for neighbors: $O(DN^2)$
- Computing the reconstruction weights: $O(DNK^4)$
- Computing the minimal eigenvalues: $O(KdN^3)$

ii. Time complexity of the GLE. Parameter N_w is the node number after self-organizing, K_{mean} is the mean of the dynamic neighbors, and T_N is the training times, with $T_N \leq 10$ generally.

- Self-organizing time consumption of GNG, concluding training time, searching for nearest neighbors and updating the weights: $T_N \times O(N_w \log N_w) + O(DN_w^2)$
- Computing the reconstruction weights: $O(DN_w K_{mean}^4)$
- Computing the minimal eigenvalues: $O(K_{mean} d N_w^2)$

Comparing the time consumption between the LLE and the proposed GLE, because $N_w \ll N$ and $K_{mean} \approx K_{opt}$, the time consumption of GLE is much less than that of LLE. The time consumption of the original LLE algorithm ascends polynomially with the increasing of the initializing samples, while that of GLE ascends linearly. The comparison of the time consumption between the two algorithms is shown in Fig. 4, with abscissa meaning initializing samples and y-axis meaning time (s), under the condition of CPU: Athlon XP 2500+, Matlab 6.5 in Windows XP sp2.

IV. PERFORMANCE ANALYSIS OF GLE

This section will show some performance analyses on LLE and GLE. Some typical manifold data like Swiss Roll and S-curve [8], [19] are applied in the analyses. We will observe the difference between the nonlinear dimensionality reduction methods while topology preservation and Hessian operator are introduced. What's more, a manifold of multi-pose and multi-expression faces is analyzed.

A. Dimensionality Reduction with Topology Preservation

In the performance analyses the unraveling objects is smooth sub-manifolds of Swiss Roll and S-curve embedding in 3-dimensional space. The unraveling results of the original LLE can be seen in reference [8], [19]. It can be seen that the manifolds are unfolded flat, but some parts are compressed overmuch. While GLE covers the support fields of the manifolds by self-organizing learning, and the results without any contractive instances are much better than the original LLE, as shown in Fig. 5 with the training times $T_N=5$.

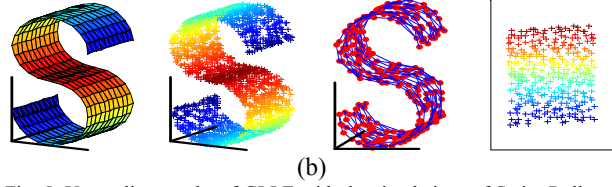
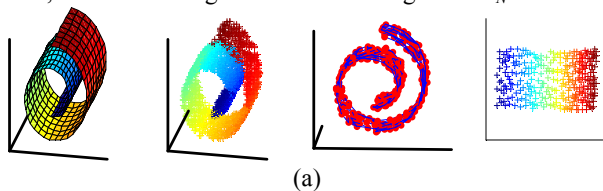


Fig. 5: Unraveling results of GLE with the simulations of Swiss Roll and S-curve. Obviously seen that the results are much regular and plain in 2-D observed space. (a) Unraveling results of Swiss Roll by GLE; (b) Unraveling results of S-curve by GLE

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. 6. That is to say, the manifold can be described as the neural nodes spatially.

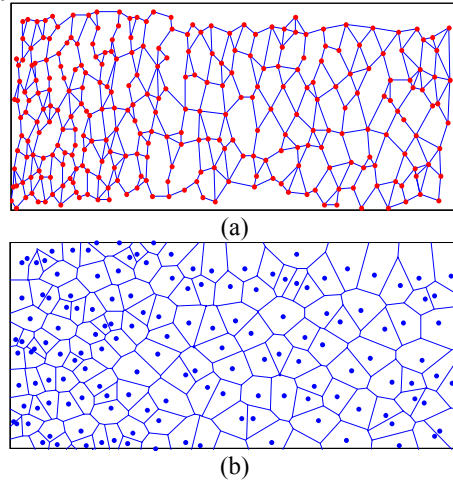


Fig. 6: The unraveling topological structure in 2-D observed space. (a) the Voronoi region of the GLE nodes; (b) covering the support field of the whole Swiss Roll manifold in the Fig. 5 (a).

B. Visualization of High-dimensional Data

These years some manifold learning methods like LLE have been applied in multi-pose face recognition and video-based face track, and some satisfactory results are achieved [23]. However, there are still some problems never solved. For example, the original LLE is hard to deal with non-convex manifolds and mass irregular appearance manifolds. In fact, it is impossible that multi-pose face sequences distribute on a manifold uniformly, for some holes and noises existing. The GLE algorithm has solved this problem. In the following experiments it is obviously seen that GLE preserves the arranging directions, while LLE cannot unfold the face sequences completely and some superposition appears.

In the following experiment the Frey Face dataset (available at <http://www.cs.toronto.edu/~roweis/data.html>) has been chosen, with 1965 images in it. Each image is a gray picture of 20×28 pixels. Some typical faces are shown in Fig. 7. Setting each image to a high-dimensional vector $D=560$, the face sequences are transformed to matrix X . then the simulation results mapped by LLE and GLE are shown in Fig. 8 separately. The result of the original LLE is hardly

satisfying for its superposition and holes, with some different faces mapped into vicinity. On the right side, the result of GLLE is satisfying with the multi-pose face changing directions shown in the figure clearly. As can be seen, the face images are mapped covering the whole field, 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. 8(a).

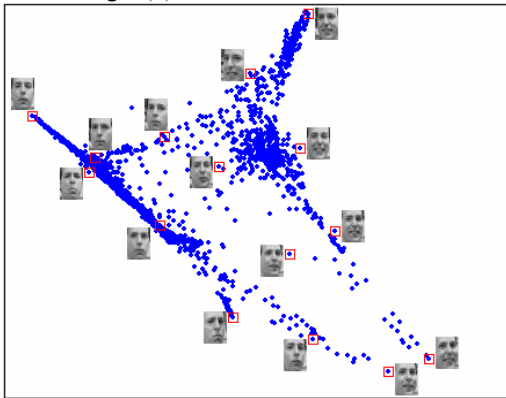


Fig. 7: Typical images in Frey Face dataset

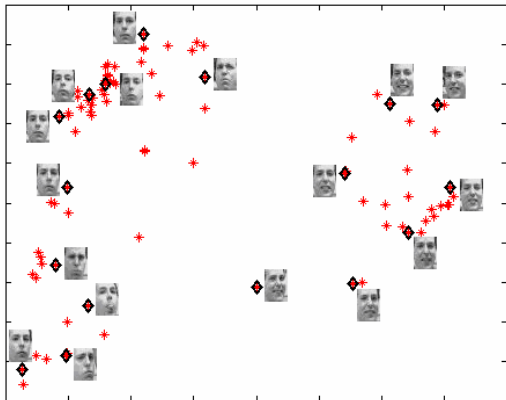
Supposing the embedding error $\varepsilon \rightarrow 0$ in GLLE, then the reconstruction face is obtained as follows:

$$Y_i = \sum_j W_{ij} Y_{ij} \quad (7)$$

It is noted that, because the proposed algorithm reflects the probability distribution, the reconstruction faces may not be the exactly true faces. That is to say, we reconstruct the faces linearly with the nearest neighbors in this experiment. Results are shown in Fig. 8(b).



(a)



(b)

Fig. 8: Simulation results of Frey Face dataset by LLE and GLLE separately. (a) simulation results of LLE; (b) simulation results of GLLE

V. SIMULATION RESULTS

In this section, the datasets comprise of some commonly used datasets, the MNIST dataset and the ORL face dataset, and the simulation results are presented. The accuracies and

time consumptions on these datasets are compared with the PCA [1], LLE, LDA [3] and PCA+LDA, which are classic and familiar algorithms. Different pattern classifiers have been applied for face recognition, including nearest-neighbor [30], Bayesian [31], Support Vector Machine [32], etc. In this paper, without loss of generality, we use the k -Nearest-Neighbor (k -NN) algorithm [30] as the classifier.

A. Applied in Pattern Classification of MNIST Digits

The MNIST database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in 28×28 pixels gray-scale images. The data set are comprised of handwriting digit 0~9 with its category information. Samples from the dataset are shown in Fig. 9. 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 experiments, the training sets are randomly selected from the training set with each digit 1000 images, while the test sets are all images of digits 4, 7, and 9 in testing set. K -Nearest-Neighbor (K -NN) algorithm is used as the classifier in common, with various K values of 1, 3, 5, 11, and 17.

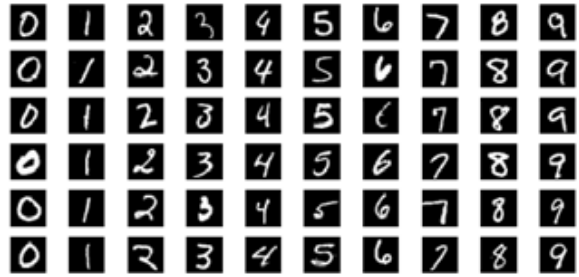


Fig. 9: Digit samples in MNIST data set. In this experiment we only use digit 4, 7 and 9.

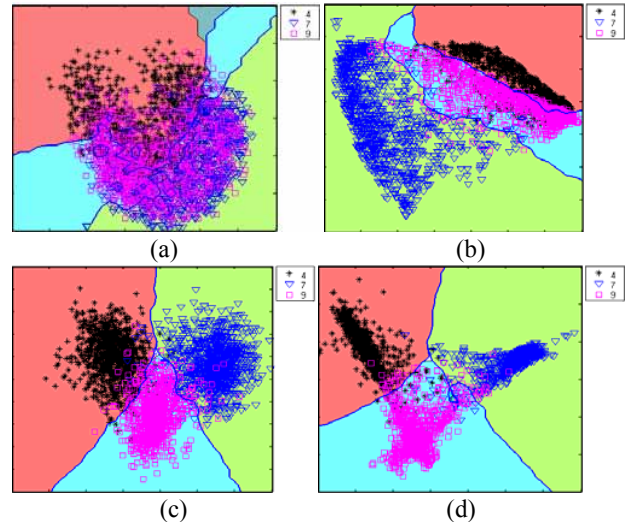


Fig. 10: The classification results of digit 4, 7, 9 in MNIST dataset by the 5 algorithms and K -NN classifier. (a) PCA; (b) LLE; (c) PCA+LDA; (d) GLLE

In experiments each algorithm is performed on the data set 10 times, and the mean accuracies with stds in brackets are

shown in Table 1.

Table I. Classification Accuracies (%)

K-N N	PCA	LLE	PCA+LD A	GLLE
1	49.447	82.577	89.346	93.004
3	51.423	85.426	91.024	94.505
5	53.423	86.982	91.602	94.823
11	55.593	87.568	92.032	95.134
17	55.933	87.822	92.222	95.280

From the results it is obviously seen that GLLE performs excellently on this data set. In theoretical analyzing, the reason is supposed the data set has a character of large without-class variation and small within-class variation in initial high dimensional space. Thus mapping node weight links of within-class samples are stronger and denser than those of without-class samples. Also it can be seen that the accuracy of PCA+LDA is also high. The reason is that the LDA algorithm has used the category information of data, while others merely extract features based on statistical information.

In the experiment we also find that the time consumption of LLE is larger than those of others, while the time consumption of GLLE in dimensionality reduction is as much as those linear algorithms. It can be said that our proposed GLLE outperforms LLE completely.

VI. CONCLUSIONS AND DISCUSSIONS

With various improving methods presented, 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]. This paper proposed an improved Growing Locally Linear Embedding algorithm (GLLE), 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. It is conformed that our proposed GLLE has overcome the 3 primary shortcomings [10] of the original algorithm, stimulating the applications of LLE. In manifold learning, it is supposed 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 has avoided 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.

ACKNOWLEDGMENT

Junsong Yin would thank Shuang Chen and B. Fritzke for discussing Growing Neural Gas.

REFERENCES

- [1] I.T. Joloffe, *Principal Component Analysis*. New York: Speinger-Verlag, 1989.
- [2] I. Borg, P. Groenen, *Modern Multidimensional Scaling*. Springer, 1997.
- [3] S. Balakrishnama, A. Ganapathiraju, "Linear discriminant analysis. Institute for Signal and Information Processing," *Mississippi State University*, 1998. 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, J.C. Langford, "The isomap algorithm and topological stability," *Science*, vol. 295, pp. 7a, 2002.
- [5] L.K. Saul, S.T. Roweis, An introduction to Locally Linear Embedding, Technical report, *Gatsby Computational Neuroscience Unit*, UCL, 2001
- [6] Mikhail Belkin, Partha 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* (Eds.2), Springer, 1995
- [8] S.T. Roweis, L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, 2000.
- [9] D. de Ridder, R.P.W. Duin, "Locally linear embedding for classification," Technical Report PH-2002-01, Pattern Recognition Group, Dept. of Imaging Science & Technology, Delft University of Technology, Delft, The Netherlands, 2002.
- [10] O. Kouropteva, O. Okun, M. Pietikäinen, "Selection of the optimal parameter value for the locally linear embedding algorithm," *Proc. of the 1st International Conference on Fuzzy Systems and Knowledge Discovery*, Singapore, pp. 359-363, 2002.
- [11] D. de Ritter, O. Kouropteva, O. Okun, M. Pietikäinen, R.P.W. Duin, "Supervised locally linear embedding," *Int. Conf. on Artificial Neural Networks and Neural Information Processing*, LNCS: 2714, Springer, 2003: 333-341.
- [12] O. Kouropteva, O. Okun, M. Pietikäinen, "Supervised locally linear embedding algorithm for pattern recognition," *Proc. of the 1st Iberian Conference on Pattern Recognition and Image Analysis*, LNCS: 2652, Spain, 2003: 384-394.
- [13] L. Donoho David, Carrie 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] Shuang Chen, Zongtan Zhou, Dewen Hu, "Diffusion and growing self-organizing map: a nitric oxide based neural model," *ISNN2004*, LNCS: 3173, 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. Martinez, "Competitive hebbian learning rule forms perfectly topology preserving maps," *ICANN93*, Amsterdam, pp. 427-434, 1993.
- [18] O. Kouropteva, O. Okun, M. Pietikäinen, "Classification of handwritten digits using supervised locally linear embedding algorithm and support vector machine," *Proc. of the 11th European Symposium on Artificial Neural Networks*, Belgium, pp. 229-234, 2003.
- [19] J.B. Tenenbaum, V. de Silva, J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319-2323, 2000.

- [20] D. Hundley, M. Kirby, "Estimation of topological dimension," *the 3rd SIAM International Conference on Data Mining*, California, pp. 194-202, 2003.
- [21] Meila Marina, Jianbo Shi, "Learning segmentation by random walks," *Advances in NIPS 13 (NIPS2000)*, Cambridge, pp. 873-879, 2001.
- [22] Junsong Yin, Dewen Hu, Shuang Chen, Zongtan 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] Changshui Zhang, Jun Wang, Nanyuan Zhao, David 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] Jian Xiao, Zongtan Zhou, Dewen Hu, Junsong Yin, Shuang Chen, "Self-organizing locally linear embedding for nonlinear dimensionality reduction," *ICNC2005, LNCS: 3610*, Changsha, pp. 101-109, 2005.
- [25] L. Xu, "Rival penalized competitive learning, finite mixture, and multisets clustering," *Proc. Int. Joint Conf. on Neural Networks*, pp. 2525-2530, 1998.
- [26] J. Bruske, 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] Elizaveta Levina, J. Bickel Peter, "Maximum likelihood estimation of intrinsic dimension," *Advances in NIPS 17 (NIPS2004)*. MIT Press, 2005.
- [28] P.J. Verwee, 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)*. Cambridge, MA. MIT Press, 2004
- [30] R.O.Duda, P.E. Hart, D.Stork. *Pattern Classification*. Wiley, 2000.
- [31] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 696-710, 1997.
- [32] P.J. Phillips, "Support vector machines applied to face recognition," *Proc. Conf. Advances in Neural Information Processing Systems 11*, pp. 803-809, 1998.