Extension of ISOMAP for Imperfect Manifolds

Chao Shao, Haitao Hu
School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou, China
Email: shaochao051227@gmail.com, frank.h@139.com

Abstract— **As one of the most promising nonlinear dimensionality reduction techniques, Isometric Mapping (ISOMAP) performs well only when the data belong to a single well-sampled manifold, where geodesic distances can be well approximated by the corresponding shortest path distances in a suitable neighborhood graph. Unfortunately, the approximation gets less and less precise generally as the number of edges of the corresponding shortest path increases, which makes ISOMAP tend to overlap or overcluster the data, especially for disjoint or imperfect manifolds. To alleviate this problem, this paper presented a variant of ISOMAP, i.e. Edge Number-based ISOMAP (EN-ISOMAP), which uses a new variant of Multidimensional Scaling (MDS), i.e. Edge Number-based Multidimensional Scaling (EN-MDS), instead of the classical Multidimensional Scaling (CMDS) to map the data into the low-dimensional embedding space. As a nonlinear variant of MDS, EN-MDS gives larger weight to the distances with fewer edges, which are generally better approximated and then more trustworthy than those with more edges, and thus can preserve the more trustworthy distances more precisely. Finally, experimental results verify that not only imperfect manifolds but also intrinsically curved manifold can be visualized by EN-ISOMAP well.**

Index Terms— **ISOMAP, EN-ISOMAP, EN-MDS, imperfect manifolds, geodesic distance, shortest path distance**

I. Introduction

Nowadays, the explosive growth in the amount of data and its dimensionality makes data visualization more and more important in the data mining process. According to No Free Lunch (NFL) Theorem [1], the structure information of the data should be taken into account to select the more suitable algorithm for data analysis/processing. For high-dimensional data, the useful structure information cannot be seen by eyes directly, but can be obtained by data visualization approaches easily. During the last decades of years, lots of approaches have been presented to visualize high-dimensional data, and most of them fall into the following five categories:

1) Several sub-windows are used to visualize the data in different subsets of dimensions respectively, such as scatterplot matrices and pixel-oriented techniques [2];

2) All the dimension axes are rearranged non-orthogonally in a low-dimensional space, such as parallel coordinates [3] and star coordinates [4];

3) The dimensions of the data are embedded each other to partition a low-dimensional space hierarchically, such as dimensional stacking [5] and treemap [6];

4) Certain objects (or icons) with several visual features are used to represent the high-dimensional data, where each visual feature stands for one dimension of the data, such as stick figures [7] and star icons [8];

5) The dimensionality of the data is reduced to two or three by dimensionality reduction techniques, such as Principal Component Analysis (PCA), Multidimensional Scaling [9], [10] (MDS), Self-Organizing Map [11], [12] (SOM), Isometric Mapping [13]–[15] (ISOMAP), Locally Linear Embedding [16], [17] (LLE), Laplacian Eigenmap [18] and Hessian Eigenmap [19] *etc*.

Unlike the other approaches, dimensionality reduction techniques try to preserve the high-dimensional relationship between the data in a low-dimensional space, and thus can visually represent the structure information of the data well. In addition, dimensionality reduction techniques can also be used to avoid ”the curse of dimensionality” and improve the performance and efficiency of the subsequent data analysis/processing algorithms.

As one of nonlinear dimensionality reduction techniques, ISOMAP extends the classical Multidimensional Scaling (CMDS) by replacing the Euclidean distance with the geodesic distance, and thus can visualize intrinsically flat manifold such as Swiss roll [13] well. However, ISOMAP requires certain assumptions about the data, one of which is that the data must belong to a single well-sampled manifold, not disjoint or imperfect manifolds [14], [17], [20]. As we know, ISOMAP uses the shortest path distances in a suitable neighborhood graph to approximate the corresponding geodesic distances between the data. Unfortunately, the approximation gets less and less precise generally as the number of edges of the corresponding shortest path increases, especially for imperfect manifolds. Generally speaking, the distances with many edges are longer than those with few edges; however, CMDS used in ISOMAP is linear and treats all the distances equally, so the generally worse-approximated distances with many edges often dominate the global structure of the result map, and the more trustworthy distances with few edges are often scarified, which makes ISOMAP tend to overlap or overcluster of the data, especially for imperfect manifolds [14], [17], [20]. To alleviate this problem, a solution is presented in [14], but the results still remain poor because some unsuitable or untrustworthy long Euclidean distances are used directly [17]. In this paper, we present a variant of ISOMAP, i.e. Edge Number-based ISOMAP (EN-ISOMAP), which improves ISOMAP by using the nonlinear Edge Number-based MDS (EN-MDS) instead of the linear CMDS. As

This work was supported in part by the Research Programme of Henan Fundamental and Advanced Technology of China (No. 112300410201), and the Key Technologies R&D Programme of Henan Province, China (No. 102102210400, 112102310519).

doi:10.4304/jcp.7.7.1780-1785

a new variant of MDS, EN-MDS can limit the effects of the generally worse-approximated distances with many edges to a certain extent, and thus the more trustworthy distances with few edges can be better preserved. As a result, besides the well-sampled intrinsically flat manifold which ISOMAP can visualize well, not only imperfect manifolds but also intrinsically curved manifold such as uniform fishbowl [24] can be visualized by EN-ISOMAP well.

This paper is organized as follows: In Section 2, we recall ISOMAP and the relevant variants briefly. In Section 3, we present EN-ISOMAP in detail. Finally, experimental results and conclusions are given in Section 4 and Section 5 respectively.

II. ISOMAP and the Relevant Variants

When the global geometric structure of the data is unknown, we are not sure that the Euclidean distance can be used to measure the dissimilarity between the data. Fortunately, the Euclidean distance is trustworthy enough to measure the dissimilarity between the data within a local neighborhood, which is also called the locally Euclidean nature of the manifold. So the global geometric structure of the data can be approximated by these local Euclidean distances, as ISOMAP does.

If the data lies on a single well-sampled manifold, it's proved that the unknown global geodesic distances between the data can be well approximated by the corresponding graph distances, i.e. the shortest path distances, in a suitable neighborhood graph which represents the right neighborhood structure of the data [21]. After using the geodesic distance instead of the Euclidean distance, the linear CMDS can be used to map the data into the low-dimensional embedding space successfully for intrinsically flat manifold, which is called ISOMAP described briefly as follows [13]:

1) Select n representative data points randomly or using vector quantization (with better results [22]) for very large data sets to keep subsequent computation tractable;
2) Construct a suitable neighborhood graph (should be connected for data visualization) using the k-nearest neighbors method with a suitable neighborhood size k (k is more natural to select than ϵ in the ϵ-ball method [23]);
3) Compute all the shortest path distances in this neighborhood graph;
4) Apply CMDS with these shortest path distances to map the data into the low-dimensional embedding space.

As described above, ISOMAP tends to overlap the data for imperfect manifolds. This problem can be alleviated by constructing the neighborhood graph with the edges between each data point and its $\frac{k}{2}$ farthest data points besides its $\frac{k}{2}$ nearest data points, instead of only its k nearest neighbors [14], which we call Global ISOMAP (G-ISOMAP). This is likely to obtain enhanced visualization results, since some of the original global information of the data can also be preserved in the low-dimensional embedding space; however, compared with ISOMAP, G-ISOMAP is unsuitable to visualize highly nonlinear data sets such as Swiss roll, because some unsuitable or untrustworthy long Euclidean distances are used directly and then the local feature is lost.

For intrinsically curved manifold such as fishbowl [24], the linear CMDS used in ISOMAP is unsuitable because the embedding is not isometric any more. In order to still apply the linear CMDS to map the data into the low-dimensional embedding space, each edge in the neighborhood graph has to be scaled in length by a scale factor, so that the corresponding angles can be preserved in the low-dimensional embedding space. Under the condition that the data be uniformly dense in the low-dimensional embedding space, Conformal ISOMAP [24] (C-ISOMAP) uses the mean distance of each data point to its k nearest neighbors as the reasonable estimate of the corresponding scale factor, that is, C-ISOMAP replaces each edge weight in the neighborhood graph with $\frac{||X_i - X_j||}{\sqrt{\mu(i)\mu(j)}}$, where $||X_i - X_j||$ represents the Euclidean distance between the i-th and j-th data points, denoted by X_i and X_j respectively, and $\mu(i)$ represents the mean distance of X_i to its k nearest neighbors. Compared with ISOMAP, C-ISOMAP can discover the intrinsical low-dimensional manifold structure of the larger class of conformal embeddings, but requires an additional assumption that the data be uniformly dense in the low-dimensional embedding space.

III. EN-ISOMAP

ISOMAP is suitable for the data lying on a single well-sampled intrinsically flat manifold such as Swiss roll, because geodesic distances can be well approximated and the linear CMDS can also perform well. Unfortunately, the data often lie on disjoint or imperfect manifolds, where the approximation of the shortest path distances to the corresponding geodesic distances is less precise, especially for the shortest paths with many edges. Generally speaking, the distances with many edges are longer but worse approximated and then less trustworthy than those with few edges, which makes the linear CMDS unsuitable. Therefore, it's reasonable to replace the linear CMDS with a nonlinear variant of MDS, which can limit the effects of the generally worse-approximated distances with many edges and emphasize the preservation of the more trustworthy distances with few edges more. This is the thinking of our method.

The error function of the linear CMDS can be described in (1), where δ_{ij} and d_{ij} represent the distances between X_i and X_j in the high-dimensional data space and in the low-dimensional embedding space respectively. In ISOMAP, δ_{ij} is the geodesic distance between X_i and X_j, approximated by the corresponding shortest path distance in a suitable neighborhood graph, and d_{ij} is the corresponding Euclidean distance between their mappings in the low-dimensional embedding space, denoted by Y_i

and Y_j respectively, described in (2).

$$E_{CMDS} = \frac{1}{2}\sum_i \sum_{j<i} (d_{ij} - \delta_{ij})^2 \tag{1}$$

$$d_{ij} = ||Y_i - Y_j|| = \sqrt{\sum_l (Y_{il} - Y_{jl})^2} \tag{2}$$

From (1), we can see that the linear CMDS treats all the distances equally, and thus the generally worse-approximated distances with many edges often dominate the global structure of the result map, and the more trustworthy distances with few edges are often scarified. To emphasize the preservation of the more trustworthy distances with few edges more, the preservation of these distances should be given larger weight than those with many edges, for example, the error function can be described in (3), where e_{ij} represents the number of edges of the shortest path between X_i and X_j in the neighborhood graph.

$$E_{EN-MDS} = \frac{1}{2}\sum_i \sum_{j<i} \frac{(d_{ij} - \delta_{ij})^2}{e_{ij}} \tag{3}$$

E_{EN-MDS} can usually be minimized by the gradient descent method iteratively, in which the choice of the learning rate is very important. To avoid the problems generated by an unsuitable learning rate, such as oscillation, we use the variable alternation method [25] to minimize E_{EN-MDS} iteratively.

Let $\frac{\partial E_{EN-MDS}}{\partial Y_i} = 0$, then we have:

$$\begin{aligned}\frac{\partial E_{EN-MDS}}{\partial Y_i} &= \sum_{j \neq i} \frac{\partial E_{EN-MDS}}{\partial d_{ij}} \cdot \frac{\partial d_{ij}}{\partial Y_i} \\ &= \sum_{j \neq i} \frac{d_{ij} - \delta_{ij}}{e_{ij}} \cdot \frac{Y_i - Y_j}{d_{ij}} \\ &= Y_i \sum_{j \neq i} \frac{1}{e_{ij}} - \sum_{j \neq i} [\frac{1}{e_{ij}} Y_j + \frac{\delta_{ij}}{e_{ij} \cdot d_{ij}} (Y_i - Y_j)] \\ &= 0\end{aligned}$$

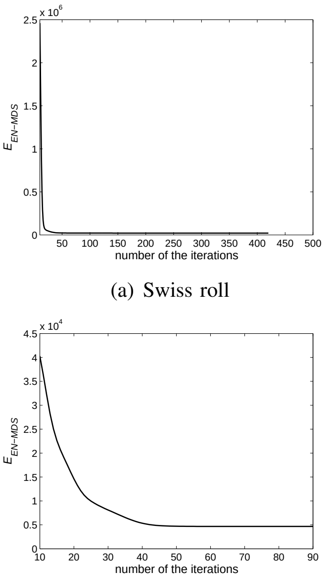
So the adjustment rule of Y_i, i.e. the mapping of X_i in the low-dimensional embedding space, can be described in (4) according to the variable alternation method, which we call EN-MDS.

$$Y_i = \frac{\sum_{j \neq i} [\frac{1}{e_{ij}} Y_j + \frac{\delta_{ij}}{e_{ij} \cdot d_{ij}} (Y_i - Y_j)]}{\sum_{j \neq i} \frac{1}{e_{ij}}} \tag{4}$$

The minimization of E_{EN-MDS} using (4) doesn't require the learning rate, while converging to a local minimum like the gradient descent method. The feasibility of the adjustment rule in (4) can be verified by experimental results on Swiss roll and fishbowl, seen in Fig. 1.

So our EN-ISOMAP can be described as follows:

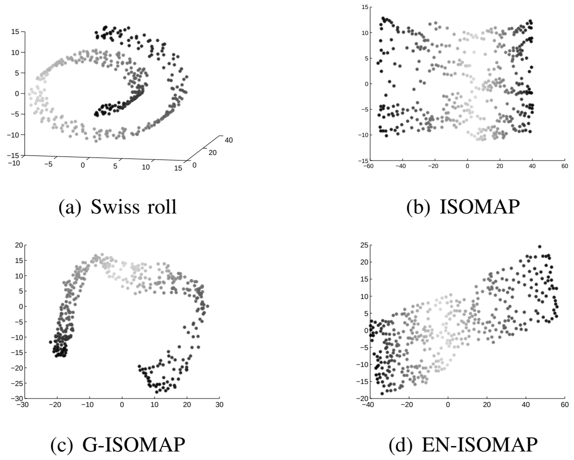
1) Select *n* representative data points using vector quantization for very large data sets to keep subsequent computation tractable;
2) Construct a suitable neighborhood graph (should be connected for data visualization) using the *k*-nearest neighbors method with a suitable neighborhood size *k*;
3) Compute all the shortest path distances in this neighborhood graph;
4) Apply EN-MDS to map the data into the low-dimensional embedding space.



(a) Swiss roll

(b) fishbowl

Figure 1. Convergence processes of the variable alternation method



(a) Swiss roll

(b) ISOMAP

(c) G-ISOMAP

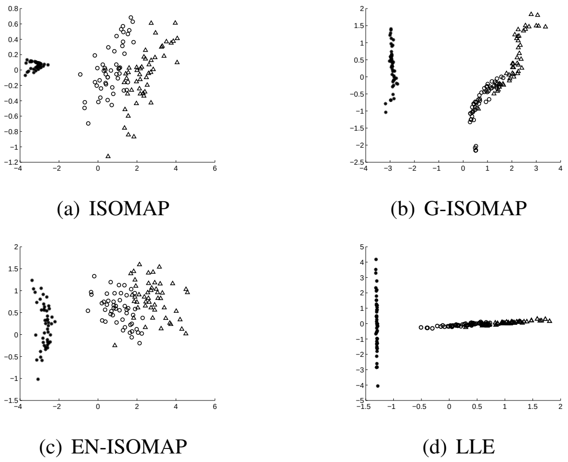
(d) EN-ISOMAP

Figure 2. The results of different algorithms with *k*=5 on Swiss roll

IV. Experimental Results

As a variant of ISOMAP, EN-ISOMAP can visualize the well-sampled intrinsically flat manifold such as Swiss roll (with 400 representatives selected from 2000 data points [13], seen in Fig. 2(a)) as nicely as ISOMAP (seen in Fig. 2(d) and Fig. 2(b) respectively), because both of them can preserve the neighborhood structure of Swiss roll well; however, G-ISOMAP can't visualize Swiss roll well (seen in Fig. 2(c)), because it uses some unsuitable or untrustworthy long Euclidean distances directly.

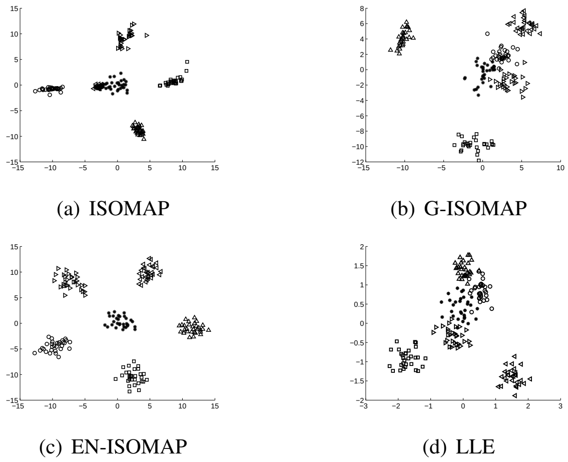
To test if EN-ISOMAP can effectively alleviate the overlapping problem presented in [14], [17], [20] and



(a) ISOMAP (b) G-ISOMAP

(c) EN-ISOMAP (d) LLE

Figure 3. The results of different algorithms with *k*=28 on IRIS



(a) ISOMAP (b) G-ISOMAP

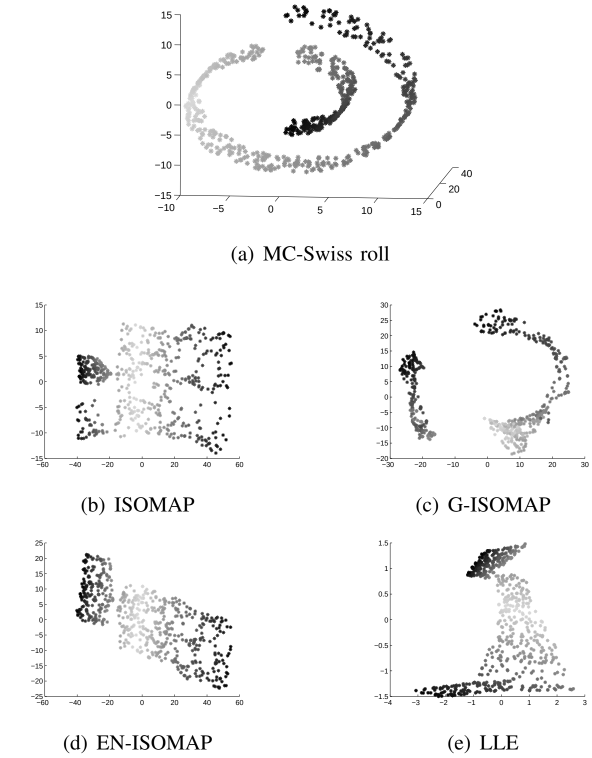
(c) EN-ISOMAP (d) LLE

Figure 4. The results of different algorithms with *k*=36 on Gaussian5d

then obtain enhanced visualization results for the data lying on imperfect manifolds, we also run ISOMAP, G-ISOMAP, EN-ISOMAP and another typical manifold learning algorithm–LLE on the following several data sets, the results are given in Fig. 3-6 respectively.

1) IRIS: a well-known four-dimensional data set with 150 data points, which are divided into three groups equally but two of them are overlapping.
2) Gaussian5d: a five-dimensional data set with 180 data points, which are divided into six groups equally and each of them follows the normal distribution with the covariance matrix of I (I is a 5×5 identity matrix). These normal distributions are independent of one another and the means are specified as (0,0,0,0,0), (10,0,0,0,0), (0,10,0,0,0), (0,0,10,0,0), (0,0,0,10,0) and (0,0,0,0,10) respectively.
3) MC-Swiss roll: Swiss roll with a small gap, seen in Fig. 5(a).
4) O-Swiss roll: the noisy Swiss roll [26] with one hole in it, seen in Fig. 6(a).

From Fig. 3-6, we can see that the data lying on imperfect manifolds such as IRIS, Gaussian5d, MC-Swiss roll and O-Swiss roll can be visualized much better by

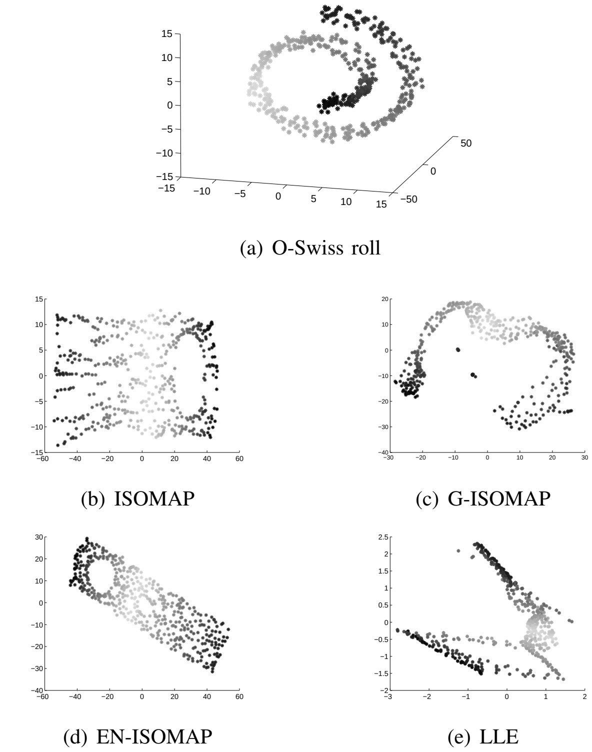


(a) MC-Swiss roll

(b) ISOMAP (c) G-ISOMAP

(d) EN-ISOMAP (e) LLE

Figure 5. The results of different algorithms with *k*=7 on MC-Swiss roll

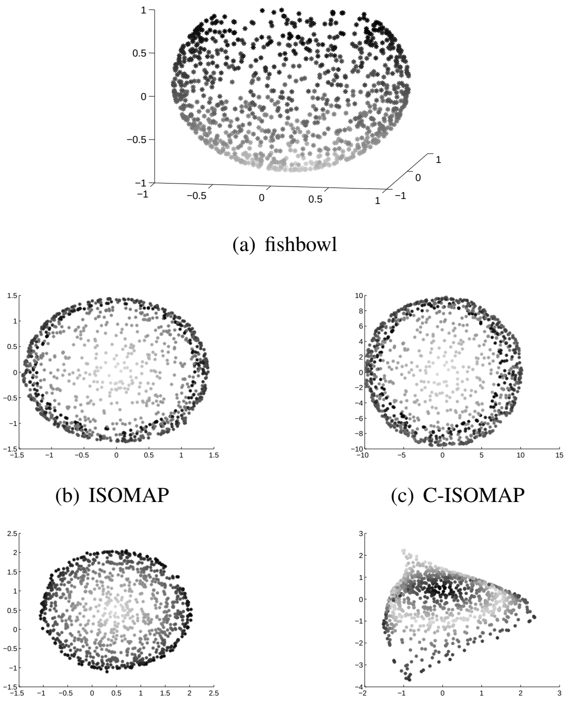


(a) O-Swiss roll

(b) ISOMAP (c) G-ISOMAP

(d) EN-ISOMAP (e) LLE

Figure 6. The results of different algorithms with *k*=5 on O-Swiss roll



(a) fishbowl

(b) ISOMAP

(c) C-ISOMAP

(d) EN-ISOMAP

(e) LLE

Figure 7. The results of different algorithms with k=10 on fishbowl

EN-ISOMAP (seen in Fig. 3(c), 4(c), 5(d), 6(d)) than ISOMAP (seen in Fig. 3(a), 4(a), 5(b), 6(b)), G-ISOMAP (seen in Fig. 3(b), 4(b), 5(c), 6(c)) and LLE (seen in Fig. 3(d), 4(d), 5(e), 6(e)). EN-ISOMAP emphasizes the preservation of the distances with few edges more, and thus can greatly alleviate the overlapping problem existed in Fig. 3(a), 4(a), 5(b), 6(b) and Fig. 3(d), 4(d), 5(e), 6(e). In addition, as we expected, LLE can preserve the global structure of these data sets less precisely than ISOMAP and EN-ISOMAP. G-ISOMAP can visualize IRIS nicely (seen in 3(b)), but can't visualize Gaussian5d, MC-Swiss roll and O-Swiss roll well (seen in 4(b), 5(c), 6(c)), because G-ISOMAP is unsuitable for highly nonlinear data sets.

By using the powerful nonlinear EN-MDS instead of the linear CMDS, EN-ISOMAP is expected to visualize more manifolds nicely. To validate this point, we also run ISOMAP, C-ISOMAP, EN-ISOMAP and LLE on an intrinsically curved manifold–uniform fishbowl [24] (seen in Fig. 7(a)), the results are given in Fig. 7. As we expected, ISOMAP and LLE can't visualize uniform fishbowl well (seen in Fig. 7(b), 7(e)), C-ISOMAP can't visualize uniform fishbowl well too (seen in Fig. 7(c)), because uniform fishbowl isn't uniformly dense in the low-dimensional embedding space. Unlike ISOMAP and C-ISOMAP, EN-ISOMAP can visualize uniform fishbowl well (seen in Fig. 7(d)), because EN-ISOMAP can better preserve the more trustworthy distances with few edges.

V. CONCLUSION

In this paper, we improve data visualization based on ISOMAP especially for the data lying on imperfect manifolds by replacing the linear CMDS with the powerful nonlinear EN-MDS, which can better preserve the more trustworthy distances with few edges and thus can alleviate the overlapping problem greatly. To avoid the problems generated by an unsuitable learning rate, we minimize the error function of EN-MDS iteratively by using the variable alternation method instead of the gradient descent method.

Similar to ISOMAP, the success of EN-ISOMAP depends greatly on selecting a suitable neighborhood size, and EN-ISOMAP is sensitive to the noise, which is the future subject.

REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification, Second Edition*. New York, NY, USA: John Wiley & Sons, Inc., 2000.

[2] D. A. Keim, "Designing pixel-oriented visualization techniques: Theory and applications," *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, no. 1, pp. 59–78, 2000.

[3] A. Inselberg and B. Dimsdale, "Parallel coordinates: A tool for visualizing multi-dimensional geometry," in *Proceedings of the 1st IEEE Conference on Visualization*, San Francisco, CA, USA, 1990, pp. 361–378.

[4] E. Kandogan, "Visualizing multi-dimensional clusters, trends, and outliers using star coordinates," in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2001, pp. 107–116.

[5] J. LeBlanc, M. O. Ward, and N. Wittels, "Exploring n-dimensional databases," in *Proceedings of the 1st IEEE Conference on Visualization*, San Francisco, CA, USA, 1990, pp. 230–237.

[6] B. Shneiderman, "Tree visualization with treemaps: A 2d space-filling approach," *ACM Transactions on Graphics*, vol. 11, no. 1, pp. 92–99, 1992.

[7] R. M. Pickett and G. G. Grinstein, "Iconographic displays for visualizing multidimensional data," in *Proceedings of the 1988 IEEE Conference on Systems, Man and Cybernetics*, Beijing and Shenyang, China, 1998, pp. 514–519.

[8] M. O. Ward, "Xmdvtool: Integrating multiple methods for visualizing multivariate data," in *Proceedings of the 5th IEEE Conference on Visualization*, Washington, DC, USA, 1994, pp. 326–336.

[9] A. Naud and W. Duch, "Interactive data exploration using mds mapping," in *Proceedings of the 5th Conference on Neural Networks and Soft Computing*, Zakopane, Poland, Jun. 2000, pp. 255–260.

[10] M. Quist and G. Yona, "Distributional scaling: An algorithm for structure-preserving embedding of metric and nonmetric spaces," *Journal of Machine Learning Research*, vol. 5, pp. 399–420, 2004.

[11] H. Yin, "Nonlinear multidimensional data projection and visualisation," in *Proceedings of the 4th International Conference on Intelligent Data Engineering and Automated Learning*, Hong Kong, China, Mar. 2003, pp. 377–388.

[12] C. Shao, X. Zhang, C. Wan, and W. Shang, "A som-based method for manifold learning and visualization," in *Proceedings of the 2nd International Joint Conference on Computational Sciences and Optimization*, Sanya, Hainan, China, 2009, pp. 312–316.

[13] 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.

[14] M. Vlachos, C. Domeniconi, D. Gunopulos, G. Kollios, and N. Koudas, "Non-linear dimensionality reduction techniques for classification and visualization," in *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, Jul. 2002, pp. 645–651.

[15] C. Shao, H. Huang, and L. Zhao, "A more topologically stable isomap algorithm," *Journal of Software (in Chinese)*, vol. 18, no. 4, pp. 869–877, 2007.

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

[17] A. Hadid and M. Pietikäinen, "Efficient locally linear embeddings of imperfect manifolds," in *Proceedings of the 3rd International Conference on Machine Learning and Data Mining*, Leipzig, Germany, Jul. 2003, pp. 188–201.

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

[19] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high dimensional data," *Proceedings of the National Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, 2003.

[20] Y. Li, "Distance-preserving projection of high-dimensional data for nonlinear dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1243–1246, 2004.

[21] M. Bernstein, V. de Silva, J. C. Langford, and J. B. Tenenbaum, "Graph approximations to geodesics on embedded manifolds," Department of Psychology, Stanford University, Tech. Rep., 2000.

[22] J. Lee, A. Lendasse, and M. Verleysen, "Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis," *Neurocomputing*, vol. 57, no. 1, pp. 49–76, 2004.

[23] O. Kouropteva, O. Okun, A. Hadid, M. Soriano, S. Marcos, and M. Pietikäinen, "Beyond locally linear embedding algorithm," Machine Vision Group, University of Oulu, Tech. Rep. MVG-01-2002, 2002.

[24] V. de Silva and J. B. Tenenbaum, "Unsupervised learning of curved manifolds," in *Proceedings of the MSRI workshop on nonlinear estimation and classification*, Berkeley, CA, USA, 2002, pp. 453–466.

[25] M. W. Trosset, "Extensions of classical multidimensional scaling: Computational theory," *Computational Statistics*, vol. 17, pp. 147–162, 2002.

[26] M. Balasubramanian, E. Shwartz, J. Tenenbaum, V. de Silva, and J. Langford, "The isomap algorithm and topological stability," *Science*, vol. 295, no. 5552, pp. 7a–7, 2002.

Chao Shao received his PH.D. and BS degrees in computer application technology from Beijing Jiaotong University of Beijing City and Lanzhou Jiaotong University of Lanzhou City, China, in 2006 and 1999, respectively.

He is currently an associate professor at Henan University of Economics and Law of Zhengzhou City, China. His research interests include artificial neural network, machine learning, data mining and data visualization.

Haitao Hu received his PH.D. and MS degree in electromagnetic field and microwave technology from Beijing University of Post and Telecommunication of Beijing City and Zhengzhou University of Zhengzhou City, China, in 2010 and 2007, respectively.

He is currently an associate professor at Henan University of Economics and Law of Zhengzhou City, China. His research interests include pattern recognition, image processing and digital watermarking.