Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2015, Article ID 241436, 18 pages
http://dx.doi.org/10.1155/2015/241436

Research Article

Enhancing Both Efficiency and Representational Capability of Isomap by Extensive Landmark Selection

Dong Liang, Chen Qiao, and Zongben Xu

School of Mathematics and Statistics and Institute for Information and System Science, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Zongben Xu; zbxu@mail.xjtu.edu.cn

Received 24 November 2014; Accepted 20 February 2015

Academic Editor: Wanquan Liu

The problems of improving computational efficiency and extending representational capability are the two hottest topics in approaches of global manifold learning. In this paper, a new method called extensive landmark Isomap (EL-Isomap) is presented, addressing both topics simultaneously. On one hand, originated from landmark Isomap (L-Isomap), which is known for its high computational efficiency property, EL-Isomap also possesses high computational efficiency through utilizing a small set of landmarks to embed all data points. On the other hand, EL-Isomap significantly extends the representational capability of L-Isomap and other global manifold learning approaches by utilizing only an available subset from the whole landmark set instead of all to embed each point. Particularly, compared with other manifold learning approaches, the data manifolds with intrinsic low-dimensional concave topologies and essential loops can be unwrapped by the new method more successfully, which are shown by simulation results on a series of synthetic and real-world data sets. Moreover, the accuracy, robustness, and computational complexity of EL-Isomap are analyzed in this paper, and the relation between EL-Isomap and L-Isomap is also discussed theoretically.

1. Introduction

Nonlinear dimensionality reduction (NLDR) is an attractive topic in many scientific fields [1–4]. The task of NLDR is to recover the latent low-dimensional structures hidden in high-dimensional data [5–7]. In many areas of artificial intelligence and data mining, the encountered high-dimensional data are intrinsically distributed on a smooth, low-dimensional manifold. The NLDR problem on such data is specifically called "manifold learning" problem [8, 9]. In recent years, there have emerged many manifold learning approaches [10–13] which are applied to many real-world application problems (e.g., hyperspectral imaging classification [14] and object tracking [15]), aiming at discovering the intrinsic geometric representations of the nonlinear data manifolds. Based on the intrinsic construction principles, these approaches can be divided into two categories: global and local approaches. Global approaches, such as Isomap [1] and CDA [10], attempt to preserve geometry at both local and global scales, essentially constructing entire isometric corresponding between all data pairs in the original and latent spaces. Local approaches, such as LLE [2] and Laplacian eigenmaps [11], attempt to preserve the local geometry of the data, intrinsically keeping invariance between all local areas in the original and latent spaces.

Compared with the local approaches, the global approaches are better in terms of giving more faithful global geometric representations and being more understandable on metric-preserving construction principles. Yet they mainly lose on two points [9]: (1) computational efficiency: the related algorithm of a global approach may be too expensive when the data are of large size, while, for the local approach, only sparse matrix computations are involved, yielding an acceptable polynomial speedup; (2) representational capacity: a global approach cannot consistently take effect except when the input data are uniformly distributed on the manifold with the intrinsic topology of a convex region in the latent space, while, for the local approach, a more extensive range of manifolds is available.

Corresponding to the above two points, the two topics have attracted more and more attention recently, which are

improving the computational speed and extending the available range of the global approaches, such that both performances could be comparable or in excess of those of the local approaches. Recently, both topics have been developed to a certain extent independently. The most typical work addressed to the first topic is the landmark Isomap (L-Isomap [9, 16]), which approximates the global computation on the whole data set by calculations on a much smaller subset (consisting of landmark points). The most prominent property of L-Isomap is that it significantly decreases the computational complexity of Isomap, under the condition that global geometric structures can still be well preserved. And also several extensive methods have been proposed for the second topic. Conformal Isomap extends Isomap to be applicable to the certain curved and offset data manifolds [9]; local MDS specially give an extension of CDA to let it be applicable to data set lying on the sphere manifold, through compromising the trustworthiness of the visualization and continuity of the mapping so as to split the sphere into two adjacent discs [17]. By building a neighborhood graph of the data to represent the underlying manifold in advance and then finding the maximum subgraph to tear or cut the manifold, the method presented in [18] can extend geodesic-distance-based approaches (including Isomap and CDA) to be available on some data manifolds with loops and having holes. By virtue of techniques of graph theory, several methods have also been proposed recently to let global approaches be extensively effective on multicluster manifolds [19–21].

The main purpose of this paper is to present a new method, addressed on both topics simultaneously. Similar to the L-Isomap, the new method also utilizes specific landmark set to embed the new input data, due to which it can be seen as an extension of L-Isomap, therefore called extensive L-Isomap or EL-Isomap. It is common knowledge that the EL-Isomap can have the similar high efficiency of L-Isomap by using landmark subset as the reference to embed the whole data set. However, the distinctions between L-Isomap and EL-Isomap in motivations, algorithms, and theoretical foundations lead to significantly different performance of the applications. The simulation results show that EL-Isomap considerably extends the range of manifolds, on which the original global approaches (including L-Isomap) take effect. The typical two examples are the data manifolds with loops and the ones with intrinsic topology of concave regions in the low-dimensional space. The synchronous improvement on both topics makes EL-Isomap distinguished from other global approaches, which is evidently verified by the simulations implemented on a series of synthetic and real-world data sets.

In summary, the proposed method has mainly the following threefold contributions in manifold learning. First, it essentially extends the available range of current manifold learning techniques and can be effectively utilized in data lying on loopy manifold, concave structured manifold, and others with complex manifold configurations. The new method thus possesses the advantage owned by many local manifold learning approaches. Second, by calculating and utilizing the geodesic distance across the entire manifold under mathematical deductions, the new method is capable of keeping global low-dimensional structure under the whole manifold. It thus inherits the advantage of global manifold learning approach, especially those geodesic-distance-based ones. Furthermore, the proposed method guarantees a low computational complexity in implementation, which is comparable to current most efficient manifold learning techniques. All these contributions have been theoretically evaluated or empirically substantiated through experiments.

This paper is organized as follows: Section 2 presents the new global approach, by virtue of comparing with L-Isomap in different viewpoints; Section 3 introduces specific strategies for landmark selection of EL-Isomap; the simulation results on synthetic and real-world data sets are demonstrated in Section 4; some discussions and conclusion are given finally.

2. From L-Isomap to EL-Isomap

Since being presented in [9], L-Isomap method has attracted many attention due to its high efficiency in applications. This prominent property is attributed to the utilization of the landmark subset, which is the common process in algorithms of L-Isomap and EL-Isomap. However, in essence, the two methods have significant difference in algorithm, theory, and application. In this section, EL-Isomap is presented by comparisons with L-Isomap in motivation, algorithm, reasonability, and computational complexity. The relation between two methods is also analyzed.

2.1. Motivations of L-Isomap and EL-Isomap. As mentioned in the first section, the initial motivation of L-Isomap is the first topic, that is, improving the computational efficiency of the global approaches. By approximating a large global computation through calculations of a much smaller set, L-Isomap significantly decreases the computational complexity of Isomap to almost linearly increasing with the number of input data set, which makes L-Isomap comparable to the local approaches in this point.

The main motivation of EL-Isomap is changed to the second topic, that is, enlarging the range of data manifolds, on which the global approaches can implement effective manifold learning. Furthermore, the algorithm also inherits the high efficiency property of L-Isomap.

Particularly, the construction of EL-Isomap is motivated heuristically by the following facts. The related information utilized by Isomap is the estimated geodesic distances between all data pairs, while, for the L-Isomap, those become the estimated geodesic distances between all data and the landmarks, a small subset of the original data set. Except the reduction of computational complexity, this also brings another extra advantage to L-Isomap: even when some of the geodesic distances (between nonlandmarks) are impossible, or not easy, or not faithful to be estimated (as the geodesic distance between A and B in Figure 1), L-Isomap can still take effect. So far, the first advantage is confirmed and emphasized in the applications of L-Isomap. Yet the second one is still exhibited very limitedly and almost ignored by the users of

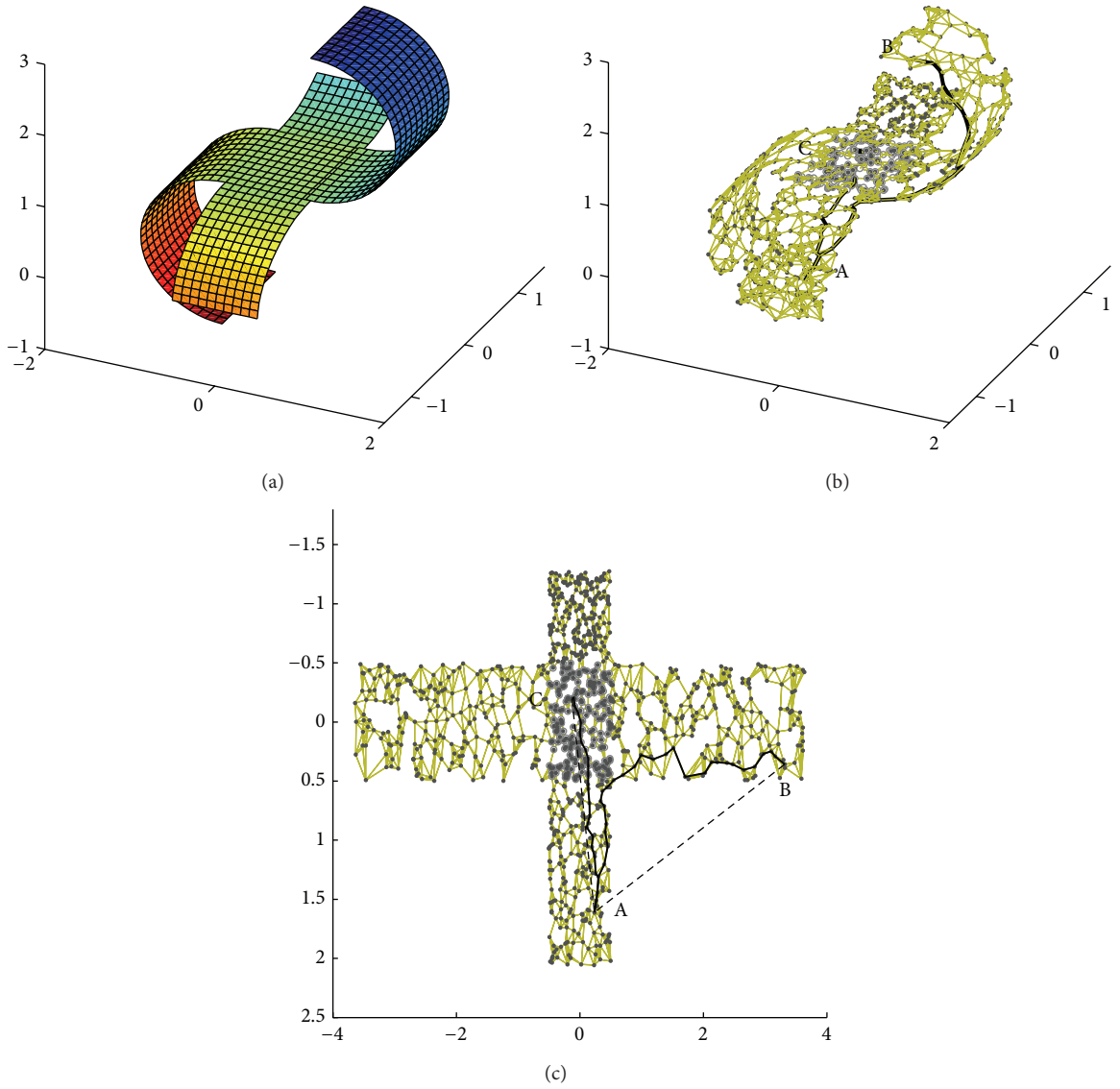

FIGURE 1: (a) is a 3D manifold combined by two crossed S-curve manifolds; (b) is a data set with 1000 vertices sampled from the manifold of (a), superimposed with 5-NN neighborhood graph; A, B, and C are three vertices lying on the manifold; the real curves are estimated geodesic curves between A and B and A and C, respectively; (c) is the unwrapped 2D corresponding points of (b); the broken lines are Euclidean lines between A and B and A and C, respectively; the big grey points are the preferred landmarks with faithful geodesic distances to other points.

the method [9, 16]. The reason is that the extensive effective range brought by L-Isomap is still not conspicuous. Through developing the EL-Isomap, which utilizes the information decreased to the estimated geodesic distances between each input datum and part of the corresponding landmarks, we aim at highlighting the second advantage to let it also be a focus of attention.

2.2. Algorithms of L-Isomap and EL-Isomap. The L-Isomap mainly contains two processes: calculating embeddings of the specialized landmarks in the latent space and adopting them as the reference to embed any new input. Comparatively, EL-Isomap also needs to implement two processes. The first processes of both methods are very similar, but the second are essentially different, in which L-Isomap realizes embedding for a new input via utilizing all landmarks as reference points while EL-Isomap does so through adopting only part of the landmarks as the available reference point set. We first list the main steps of L-Isomap, and then we only present the different part of EL-Isomap from L-Isomap as a comparison.

Algorithm of L-Isomap

Input. There is a given data set with size l in space R^N; desired number of landmarks n; desired number of output dimension d ($N \gg d$).

Output. There are the embeddings of the given data set in space R^d.

Step 1. Specialize a set of n landmark points $\{\vec{x}_i\}_{i=1}^{n}$ from the input data set.

Step 2. Estimate the approximate geodesic distance matrix $D_n \in R^{n\times n}$ of all landmarks by Dijkstra's or Floyd's algorithm [1]; let Δ_n be its squared distance matrix ($[\Delta_n]_{ij} = [D_n]_{ij}^2$); construct the matrix

$$B_n = -\frac{1}{2} H_n \Delta_n H_n, \tag{1}$$

where H_n is the mean-centering matrix, defined by $[H_n]_{ij} = \delta_{ij} - 1/m$; compute the d largest positive eigenvalues ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$) and the corresponding eigenvectors ($\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_d$) of B_n; and then get the required d-dimensional embedding vectors ($\vec{l}_1, \vec{l}_2, \ldots, \vec{l}_n$) of the landmarks by

$$\left[\vec{l}_1 \;\; \vec{l}_2 \;\; \cdots \;\; \vec{l}_n\right] = L_d = \begin{bmatrix} \sqrt{\lambda_1}\cdot\vec{v}_1^T \\ \sqrt{\lambda_2}\cdot\vec{v}_2^T \\ \vdots \\ \sqrt{\lambda_d}\cdot\vec{v}_d^T \end{bmatrix} \in R^{d\times n}. \tag{2}$$

Step 3. Estimate the geodesic distances d_{ai} between x_a and all landmarks $\vec{x}_i$ ($i = 1, 2, \ldots, n$); denote three vectors as

$$\begin{aligned} \vec{\delta}_a &= \left[d_{a1}^2 \;\; d_{a2}^2 \;\; \cdots \;\; d_{an}^2\right]^T, \\ \vec{\delta}_i &= \left[d_{i1}^2 \;\; d_{i2}^2 \;\; \cdots \;\; d_{in}^2\right]^T, \\ \vec{\delta}_\mu &= \frac{\left(\vec{\delta}_1 + \cdots + \vec{\delta}_n\right)}{n}, \end{aligned} \tag{3}$$

where d_{ij} ($1 \leq i, j \leq n$) is the estimated geodesic distance between x_i and x_j; calculate the low-dimensional embedding $\vec{y}_a$ of $\vec{x}_a$ as

$$\vec{y}_a = -\frac{1}{2} L_d^{\#T}\left(\vec{\delta}_a - \vec{\delta}_\mu\right), \tag{4}$$

where

$$L_d^{\#} = \left[\frac{v_1^T}{\sqrt{\lambda_1}} \;\; \frac{v_2^T}{\sqrt{\lambda_2}} \;\; \cdots \;\; \frac{v_d^T}{\sqrt{\lambda_d}}\right] \in R^{n\times d}. \tag{5}$$

Step 4. Center embeddings of all landmarks and nonlandmarks to their mean; apply PCA [22] to align the principle axes of the embedded data with the coordinate axes in order to decrease significance.

The main difference between L-Isomap and EL-Isomap is Step 3, which commonly aims at embedding a new input from the original space to the latent space. Step 3 of EL-Isomap is listed as follows in contrast to that of L-Isomap.

Step 3 of EL-Isomap

Step 3.1. For each nonlandmark $\vec{x}_a$ choose available landmark subset from the landmark set. Without influencing the process of the algorithm, we may assume that the available landmarks chosen are the first m data in the landmark set; that is, $\{\vec{x}_i\}_{i=1}^{m}$ ($m < n$); their corresponding low-dimensional embeddings are

$$\left[\vec{l}_1 \;\; \vec{l}_2 \;\; \cdots \;\; \vec{l}_m\right] = L_d' = \begin{bmatrix} \sqrt{\lambda_1}\cdot\overrightarrow{v'}_1^T \\ \sqrt{\lambda_2}\cdot\overrightarrow{v'}_2^T \\ \vdots \\ \sqrt{\lambda_d}\cdot\overrightarrow{v'}_d^T \end{bmatrix} \in R^{d\times m}, \tag{6}$$

where $\overrightarrow{v'}_i$ denotes the shortened vector of $\vec{v}_i$ by deleting its last $n - m$ elements.

Step 3.2. Estimate the geodesic distances d_{ai} between x_a and the chosen landmarks $\vec{x}_i$ ($i = 1, 2, \ldots, m$) via Dijkstra's or Floyd's algorithm; denote three vectors as

$$\begin{aligned} \vec{\rho}_a &= \left[d_{a1}^2 \;\; d_{a2}^2 \;\; \cdots \;\; d_{am}^2\right]^T, \\ \vec{\rho}_i &= \left[d_{i1}^2 \;\; d_{i2}^2 \;\; \cdots \;\; d_{im}^2\right]^T, \\ \vec{\rho}_\mu &= \frac{\left(\vec{\rho}_1 + \cdots + \vec{\rho}_m\right)}{m}, \end{aligned} \tag{7}$$

where $d_{ij} = [D_n]_{ij}$ ($1 \leq i, j \leq m$) is the estimated geodesic distance between x_i and x_j (referring to Step 2).

Step 3.3. Construct the squared distance matrix Δ_m, where $[\Delta_m]_{ij} = d_{ij}^2$ and

$$B_m = -\frac{1}{2} H_m \Delta_m H_m; \tag{8}$$

compute the d largest positive eigenvalues ($\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_d$) and the corresponding eigenvectors ($\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_d$) of B_m; construct matrix W as

$$W = \left[\frac{\vec{w}_1}{\sqrt{\gamma_1}} \;\; \frac{\vec{w}_2}{\sqrt{\gamma_2}} \;\; \cdots \;\; \frac{\vec{w}_d}{\sqrt{\gamma_d}}\right] \in R^{m\times d}. \tag{9}$$

Step 3.4. Calculate the low-dimensional embedding $\vec{y}_a$ of $\vec{x}_a$ as

$$\vec{y}_a = L_d'\left(-\frac{1}{2} W W^T\left(\vec{\rho}_a - \vec{\rho}_\mu\right) + \vec{1}\cdot\frac{1}{m}\right), \tag{10}$$

where $\vec{1} = \left[1 \;\; 1 \;\; \cdots \;\; 1\right]^T \in R^m$.

Note that Step 4 of EL-Isomap actually follows the same setting as the L-Isomap method [9], which is not for dimensionality reduction, but for easily translating the obtained embedding points into the origin and rotating them along their principal axes. That is, this stage aims to use PCA to realign the data with the coordinate axes, while the embedding dimensionality will not be changed. In practice, this stage always facilitates a better visualization.

Actually, another difference between the two algorithms is the related landmarks selecting strategies (Step 1) derived from the distinct aims of L-Isomap and EL-Isomap. Since this section intends to give only macroscopical comparisons between both methods, the detailed analysis for difference between Step 1 of both methods will be given in Section 3 specifically. Besides, the description for how to designate the available landmark subset (Step 3.1) will also be presented in that section due to its close relationship with Step 1.

Notice that the first d eigenvalues of the matrices B_n and B_m should be positive to let L_d in (2) and W in (9) make sense. If the geodesic distance matrix D_n is almost Euclidean, that is, the geodesic distances between $\{\vec{x}_i\}_{i=1}^n$ approximately correspond to the Euclidean distances between a low-dimensional data set $\{\vec{Y}_i\}_{i=1}^n$, then the largest d eigenvalues of B_n and B_m are generally positive. However, in practice, there may be certain perturbations existing in D_n, which may account for the negative eigenvalues of B_n and B_m. When the extents of perturbations are so large that the first d eigenvalues of B_n are disturbed to be negative, even the feasibility of the two algorithms can not be promised. The next section will give a theoretical comparison for how the perturbations quantitatively influence the embedding performance of L-Isomap and EL-Isomap.

2.3. Reasonabilities of L-Isomap and EL-Isomap. The reasonability of a global approach for manifold learning can be learned via validating its two capabilities: accuracy in Euclidean case and stability in non-Euclidean case. Specifically, the accuracy of a global approach means when the geodesic distance matrix of manifold data is exactly Euclidean, the approach is guaranteed to find the accurate global figure of the corresponding low-dimensional data; the stability of a global approach means when the estimated geodesic distances between high-dimensional data are with certain perturbations to render them deviated from Euclidean condition, the approach can still stably find the approximately correct embeddings.

To analyze the reasonability of L-Isomap and the reasonability of EL-Isomap, the reasonability of the calculated embeddings of the landmark set by Step 2 has to be involved due to its decisive influence on the final results of both methods. To this aim, the following two theorems are presented.

Theorem 1 (accuracy of landmark embedding [16]). *Suppose the geodesic distance matrix D_n of the landmark set $\{\vec{x}_i\}_{i=1}^n$ corresponds to the Euclidean distance matrix of a collection of data $\{\vec{Y}_i\}_{i=1}^n$. By translating if necessary we can assume that this set is mean-centered. In this case the ith row of calculated embedding $\vec{l}_j$ in (2) gives the component of $\vec{Y}_j$ when projected onto the ith principal axis of $\{\vec{Y}_i\}_{i=1}^n$; that is,*

$$\vec{l}_j = \begin{bmatrix} \vec{p}_1^T \\ \vec{p}_2^T \\ \vdots \\ \vec{p}_d^T \end{bmatrix} \cdot \vec{Y}_j, \quad j = 1, 2, \ldots, n, \tag{11}$$

where $(\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_d)$ are the first d principle directions of $\{\vec{Y}_i\}_{i=1}^n$.

Theorem 2 (stability of landmark embedding [16]). *Let Δ_n be the squared distance matrix (i.e., $[\Delta_n]_{ij} = [D_n]_{ij}^2$) of the landmarks; let $B_n = -(1/2)H_n\Delta_n H_n$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and $\lambda_d > \max(0, \lambda_{d+1})$. Consider a perturbation*

$$\widehat{\Delta}_n = \Delta_n + t\phi + O\left(t^2\right). \tag{12}$$

Then there is

$$\widehat{L}_d = L_d + t\chi + O\left(t^2\right), \tag{13}$$

where the columns of $\widehat{L}_d$ and L_d are the calculated embedding vectors by Step 2 on squared distance matrixes with and without perturbations, respectively. Moreover, there are the bounds

$$\|\chi\| \leq \left(\frac{1}{4\lambda_d^{1/2}} + \frac{\lambda_1^{1/2}}{2\left(\lambda_d - \lambda_{d+1}\right)} \right) \|\phi\| \tag{14}$$

in the Frobenius norm.

The above theorems can conveniently deduce the reasonability of L-Isomap. The related results are listed in the following as comparisons with those of EL-Isomap.

Theorem 3 (accuracy of L-Isomap [16]). *Under the situation of Theorem 1, suppose that there exists $\vec{Y}_a$ such that the geodesic distance d_{ai} (as defined in (3)) accords with the Euclidean distance between $\vec{Y}_a$ and each $\vec{Y}_i$ ($i = 1, 2, \ldots, n$), and then the vector $\vec{y}_a$ calculated in (4) gives the components of $\vec{Y}_a$ when projected onto the first d principal axes of $\{\vec{Y}_i\}_{i=1}^n$; that is,*

$$\vec{y}_a = \begin{bmatrix} \vec{p}_1^T \\ \vec{p}_2^T \\ \vdots \\ \vec{p}_d^T \end{bmatrix} \cdot \vec{Y}_a. \tag{15}$$

Theorem 4 (stability of L-Isomap [16]). *Under the situation of Theorem 2, let $\widehat{\vec{\delta}}_a = \vec{\delta}_a + t\vec{\zeta}_a + O(t^2)$ be a perturbation family for the vector of squared geodesic distances between the point x_a and the landmarks. Then the perturbation family*

$$\widehat{y}_a = \vec{y}_a - \frac{t\left(L_d^{\#T}\vec{\zeta}_a + \psi\vec{\delta}_a\right)}{2} + O\left(t^2\right) \tag{16}$$

agrees with the family of embedding vectors by applying L-Isomap, up to a rotation and translation which depend on t but are independent of x_a, where ψ has the bound

$$\|\psi\| \leq \left(\frac{1}{4\lambda_d^{3/2}} + \frac{1}{2\lambda_d^{1/2}\left(\lambda_d - \lambda_{d+1}\right)} \right) \|\phi\|. \tag{17}$$

Like L-Isomap, EL-Isomap also aims at consistently finding the embedding of any new input, that is, finding the components of its low-dimensional corresponding when projected onto the uniform coordinate system $(p_1, p_2, \ldots, p_d)$ (as defined in Theorem 1). However, different from L-Isomap, EL-Isomap utilizes dissimilar available landmark subsets as references to embed different inputs, which brings difficulty to keep the consistency of the whole data embedding by EL-Isomap. If we denote $E^{n,d}$ as the space affinely spanned by top d principle axes of $\{\vec{Y}_i\}_{i=1}^n$ (as defined in Theorem 1) and $E_a^{m,d}$ as that by top d principle axes of $\{\vec{Y}_i\}_{i=1}^m$, which indicates that the related available landmark subset corresponds to the new input x_a, a necessary condition to make the consistency possible is that $E^{n,d}$ is close to each $E_a^{m,d}$; that is, $E_a^{n,d}$ can be linearly expressed by $E_a^{m,d}$ approximately. This condition is natural since if $E_a^{m,d}$ is heavily deviated from the space $E^{n,d}$, it is impossible to approximate projection of the low-dimensional corresponding of x_a to $(p_1, p_2, \ldots, p_d)$ by virtue of $\{\vec{x}_i\}_{i=1}^m$ (corresponding to $E_a^{m,d}$) instead of $\{\vec{x}_i\}_{i=1}^n$ (corresponding to $E^{n,d}$). Using this condition as the only supplement, the reasonability of EL-Isomap can be proved, which is listed in the following.

Theorem 5 (accuracy of EL-Isomap). *Under the situation of Theorem 1, suppose the affine space $E_a^{m,d}$ is similar to $E^{n,d}$ and there exists a low-dimensional vector $\vec{Y}_a$ such that the geodesic distance d_{ai} accords with the Euclidean distance between $\vec{Y}_a$ and $\vec{Y}_i$ $(i = 1, 2, \ldots, n)$; then the vector $\vec{y}_a$ calculated in (10) gives the components of $\vec{Y}_a$ when projected onto the first d principal axes of $\{\vec{Y}_i\}_{i=1}^n$; that is,*

$$\vec{y}_a = \begin{bmatrix} \vec{p}_1^T \\ \vec{p}_2^T \\ \vdots \\ \vec{p}_d^T \end{bmatrix} \cdot \vec{Y}_a. \tag{18}$$

Proof. (i) Center $\{\vec{Y}_i\}_{i=1}^m$ to get new data as

$$\vec{z}_i = \vec{Y}_i - \frac{1}{m}\sum_{k=1}^m \vec{Y}_k, \quad i = 1, 2, \ldots, m \tag{19}$$

and, correspondingly, transform $\vec{Y}_a$ as

$$\vec{z}_a = \vec{Y}_a - \frac{1}{m}\sum_{k=1}^m \vec{Y}_k. \tag{20}$$

Then the m-dimensional vectors $\vec{\rho}_a$ and $\vec{\rho}_\mu$ defined in (7) are equivalent to

$$\begin{aligned} \vec{\rho}_a &= \begin{pmatrix} \vdots \\ \left|\vec{z}_a - \vec{z}_j\right|^2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \left|\vec{z}_a\right|^2 - 2\vec{z}_a^T\vec{z}_j + \left|\vec{z}_j\right|^2 \\ \vdots \end{pmatrix}, \\ \vec{\rho}_\mu &= \begin{pmatrix} \vdots \\ \dfrac{1}{m}\displaystyle\sum_{k=1}^m \left|\vec{z}_j - \vec{z}_k\right|^2 \\ \vdots \end{pmatrix} \\ &= \begin{pmatrix} \vdots \\ \dfrac{1}{m}\displaystyle\sum_{k=1}^m \left|\vec{z}_k\right|^2 - \dfrac{1}{m}\displaystyle\sum_{k=1}^m \vec{z}_k^T\vec{z}_j + \left|\vec{z}_j\right|^2 \\ \vdots \end{pmatrix}. \end{aligned} \tag{21}$$

Hence

$$\begin{aligned} \vec{\rho}_a - \vec{\rho}_\mu &= \begin{pmatrix} \vdots \\ \left|\vec{z}_a\right|^2 - \dfrac{1}{m}\displaystyle\sum_{k=1}^m \left|\vec{z}_k\right|^2 - 2\vec{z}_a^T\vec{z}_j \\ \vdots \end{pmatrix} \\ &= c\vec{1} - 2Z^T\vec{z}_a, \end{aligned} \tag{22}$$

where $c = |\vec{z}_a|^2 - (1/m)\sum_{k=1}^m |\vec{z}_k|^2$ is a scalar and

$$Z = \begin{bmatrix} \vec{z}_1 & \vec{z}_2 & \cdots & \vec{z}_m \end{bmatrix} \in R^{d\times m}. \tag{23}$$

As defined in (8) and (9), $\vec{w}_i$ is the ith eigenvector of $B_m = -(1/2)H_m\Delta_m H_m$; thus $\vec{w}_i^T\vec{1} = 0$. Then we have

$$-\frac{1}{2}W^T\left(\vec{\rho}_a - \vec{\rho}_\mu\right) = W^TZ^T\vec{z}_a, \tag{24}$$

where W is defined as formula (9).

Denote the first d principle directions of $\{\vec{z}_i\}_{i=1}^m$ as $\vec{q}_1 \ \vec{q}_2 \ \cdots \ \vec{q}_d$ and construct orthogonal matrix

$$Q = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_d \end{bmatrix}. \tag{25}$$

Since $\vec{w}_i$ is the eigenvector of $B_m = Z^TZ$ and intrinsically $\vec{q}_i$ is the eigenvector of ZZ^T, there exist relations between them as [23]

$$\vec{q}_i = \frac{Z\vec{w}_i}{\sqrt{\gamma_i}}, \tag{26}$$

where γ_i is defined as (9). Then it naturally follows that

$$Q = ZW. \tag{27}$$

According to (24) and (27), there holds

$$-\frac{1}{2}W^T\left(\rho_a-\rho_\mu\right)=Q^T\vec{z}_a=\begin{bmatrix}\vec{q}_1^T\\ \vec{q}_2^T\\ \vdots\\ \vec{q}_d^T\end{bmatrix}\cdot\vec{z}_a. \tag{28}$$

(ii) Calculate the projection of $\vec{p}_i$ when projected onto the principle directions of $\{\vec{Y}_i\}_{i=1}^m$ as

$$\begin{aligned}&\begin{bmatrix}\vec{q}_1^T\\ \vec{q}_2^T\\ \vdots\\ \vec{q}_d^T\end{bmatrix}\cdot\left[\vec{p}_1\ \ \vec{p}_2\ \ \cdots\ \ \vec{p}_d\right]\\ &=Q^T\left[\vec{p}_1\ \ \vec{p}_2\ \ \cdots\ \ \vec{p}_d\right]\\ &=W^TZ^T\left[\vec{p}_1\ \ \vec{p}_2\ \ \cdots\ \ \vec{p}_d\right]\\ &=W^TH_d\left[\vec{Y}_1\ \ \vec{Y}_2\ \ \cdots\ \ \vec{Y}_m\right]^T\cdot\left[\vec{p}_1\ \ \vec{p}_2\ \ \cdots\ \ \vec{p}_d\right]\\ &=W^TH_dL_d'^T=W^TL_d'^T,\end{aligned} \tag{29}$$

where H_d is the mean-centering matrix and L_d' is defined as formula (6). The omitting of the matrix H_d in the last step is due to $W^TH_d=W^T$.

(iii) According to the supplemental condition of the theorem, each p_i can be linearly expressed by $\{q_i\}_{i=1}^d$. Then, based on (11), (28), and (29), it can be deduced that

$$\begin{aligned}\begin{bmatrix}\vec{p}_1^T\\ \vec{p}_2^T\\ \vdots\\ \vec{p}_d^T\end{bmatrix}\cdot\vec{Y}_a&=\begin{bmatrix}\left(\vec{p}_1^T\vec{q}_1\right)\vec{q}_1^T+\cdots+\left(\vec{p}_1^T\vec{q}_d\right)\vec{q}_d^T\\ \left(\vec{p}_2^T\vec{q}_1\right)\vec{q}_1^T+\cdots+\left(\vec{p}_2^T\vec{q}_d\right)\vec{q}_d^T\\ \vdots\\ \left(\vec{p}_n^T\vec{q}_1\right)\vec{q}_1^T+\cdots+\left(\vec{p}_n^T\vec{q}_d\right)\vec{q}_d^T\end{bmatrix}z_a\\ &\quad+\frac{1}{m}\sum_{k=1}^m\begin{bmatrix}\vec{p}_1^T\\ \vec{p}_2^T\\ \vdots\\ \vec{p}_d^T\end{bmatrix}\vec{Y}_k\\ &=\begin{bmatrix}\vec{p}_1^T\\ \vec{p}_2^T\\ \vdots\\ \vec{p}_d^T\end{bmatrix}\cdot\left[\vec{q}_1\ \ \vec{q}_2\ \ \cdots\ \ \vec{q}_d\right]\cdot\begin{bmatrix}\vec{q}_1^T\\ \vec{q}_2^T\\ \vdots\\ \vec{q}_d^T\end{bmatrix}z_a\\ &\quad+\frac{1}{m}\sum_{k=1}^m\vec{l}_k\\ &=-\frac{1}{2}L_d'WW^T\left(\vec{\rho}_a-\vec{\rho}_\mu\right)+\frac{1}{m}L_d'\vec{1}\\ &=L_d'\left(-\frac{1}{2}WW^T\left(\vec{\rho}_a-\vec{\rho}_\mu\right)+\vec{1}\cdot\frac{1}{m}\right),\end{aligned} \tag{30}$$

which is formula (10) in the algorithm of EL-Isomap.

The proof is completed. □

The stability of EL-Isomap can then be given by virtue of the above results.

Theorem 6 (stability of EL-Isomap). *Consider perturbations*

$$\begin{aligned}\widehat{\Delta}_n&=\Delta_n+t\phi+O\left(t^2\right),\\ \widehat{\Delta}_m&=\Delta_m+t\xi+O\left(t^2\right),\\ \widehat{\rho}_a&=\vec{\rho}_a+t\vec{\varsigma}_a+O\left(t^2\right).\end{aligned} \tag{31}$$

Let B_n (as defined in Section 2.2) with eigenvalues $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$ and $\lambda_d>\max(0,\lambda_{d+1})$; B_m (as defined in (8)) with eigenvalues $\gamma_1\geq\gamma_2\geq\cdots\geq\gamma_m$ and $\gamma_d>\max(0,\gamma_{d+1})$; then the perturbation family

$$\widehat{y}_a=\vec{y}_a-\frac{t\Upsilon}{2}+O\left(t^2\right) \tag{32}$$

agrees with the family of embedding vectors by applying EL-Isomap, up to a rotation and translation which depend on t but are independent of x_a, where

$$\Upsilon=\kappa WW^T\vec{\rho}_a+L_d'\vartheta W^T\vec{\rho}_a+L_d'W\vartheta^T\vec{\rho}_a+L_d'WW^T\overrightarrow{\varsigma_a}. \tag{33}$$

Besides, there are bounds

$$\begin{aligned}\|\kappa\|&\leq\left(\frac{1}{4\lambda_d^{1/2}}+\frac{\lambda_1^{1/2}}{2\left(\lambda_d-\lambda_{d+1}\right)}\right)\|\phi\|,\\ \|\vartheta\|&\leq\left(\frac{1}{4\gamma_d^{1/2}}+\frac{\gamma_1^{1/2}}{2\left(\gamma_d-\gamma_{d+1}\right)}\right)\|\xi\|.\end{aligned} \tag{34}$$

Proof. (i) Since

$$\widehat{\Delta}_m=\Delta_m+t\xi+O\left(t^2\right) \tag{35}$$

according to Theorem 2, we have

$$\widehat{W}=W+t\vartheta+O\left(t^2\right). \tag{36}$$

Besides, the following bound holds:

$$\|\vartheta\|\leq\left(\frac{1}{4\gamma_d^{1/2}}+\frac{\gamma_1^{1/2}}{2\left(\gamma_d-\gamma_{d+1}\right)}\right)\|\xi\|. \tag{37}$$

(ii) According to (2) and (6), we know that $\widehat{L'_d}$ and L'_d are subvectors of $\widehat{L_d}$ and L_d, respectively. Based on formula (12), it is clear that

$$\widehat{L'_d} = L'_d + t\kappa + O\left(t^2\right), \tag{38}$$

where κ is the subvector of χ defined as (14). Subsequently

$$\|\kappa\| \le \|\chi\| \le \left(\frac{1}{4\lambda_d^{1/2}} + \frac{\lambda_1^{1/2}}{2\left(\lambda_d - \lambda_{d+1}\right)}\right)\|\phi\|. \tag{39}$$

(iii) Since

$$\vec{y}_a = L'_d\left(-\frac{1}{2}WW^T\left(\vec{\rho}_a - \vec{\rho}_\mu\right) + \vec{1}\cdot\frac{1}{m}\right), \tag{40}$$

based on (37) and (39), it is easy to deduce the conclusion. □

Based on the above theorems we can give theoretical analysis for both methods comparatively. Theorems 3 and 5 show that accuracies of both L-Isomap and EL-Isomap can be guaranteed when all required theoretical preconditions are satisfied and the related quantities can be estimated without errors. Yet there exists unignorable difference between the stability results of the two methods. Except the common perturbations on distance matrix of landmarks, Theorem 4 implies that the stability of L-Isomap depends on the perturbations of the estimated geodesic distances between the input $\vec{x}_a$ and all landmarks (i.e., $\vec{\zeta}_a$), while Theorem 6 indicates that EL-Isomap's stability relies on less information: the perturbations of geodesic distances between the input and part of landmarks (i.e., $\vec{\varsigma}_a$, a subvector of $\vec{\zeta}_a$). When the geodesic distances between $\vec{x}_a$ and some landmarks are estimated with inevitable heavy perturbations, the stable performance of L-Isomap can not be guaranteed in theory due to the large value of $\|\vec{\zeta}_a\|$; however, EL-Isomap would possibly avoid this by only utilizing landmarks, between which and x_a the geodesic distances can be estimated with less perturbations; that is, the value of $\|\vec{\varsigma}_a\|$ is small. It is obvious in the above cases that EL-Isomap tends to have more stable performance than L-Isomap, which exactly accords with the motivation of constructing EL-Isomap mentioned above, via carefully selecting the available landmark subset corresponding to the new input.

2.4. Computational Complexities of Isomap, L-Isomap, and EL-Isomap. As mentioned above, EL-Isomap is closely related to L-Isomap and Isomap, because L-Isomap is derived from Isomap, whose main aim is to improve the computational complexity of Isomap; on the other hand EL-Isomap is originated from L-Isomap, whose main purpose is to extend the representational capacity of L-Isomap and further the global manifold learning approaches. Another consideration of EL-Isomap is to inherit high efficiency property of L-Isomap, so as to make EL-Isomap have improvement in both hot topics, which are addressed on global approaches for manifold learning. To this aim, the computational complexities (time and space complexities) for three methods are analyzed as follows, aiming at showing efficiency property of EL-Isomap through comparisons.

The Isomap algorithm costs time mainly on two steps: geodesic distance estimation, which estimates an approximate geodesic distance matrix D_l ($l \times l$) through calculating the shortest paths between all data pairs, and the MDS eigenvalue calculation on this matrix. The time complexity of the first step is about $O(kl^2 \log l)$ or $O(l^3)$ (k is the neighborhood size), by utilizing Dijkstra or Floyd algorithm, and that of the second step is around $O(l^3)$. Hence the total time complexity of Isomap is about $O(kl^2 \log l + l^3)$. As to the space complexity, we only consider the storage of the geodesic distance matrix D_l since it mainly determines the space complexity of the method. It is easy to see that $O(l^2)$ storage is needed for Isomap to store the estimated geodesic distances between all data.

L-Isomap reduces both the time and space complexities of Isomap significantly. For geodesic distance estimation, instead of calculating D_l, only the $n \times l$ matrix $D_{n,l}$ needs to be computed, which contains the geodesic distances estimated from each data point to all landmark points. This only needs $O(knl \log l)$ time by Dijkstra algorithm. And MDS calculation on the whole data set of Isomap is simplified to two simple processes of L-Isomap: implementing MDS on the landmark set, which costs $O(n^3)$ time, and directly computing an Euclidean embedding of the data, as shown in (3) and (4), which costs $O(n^2 l)$ time. The total time complexity of L-Isomap is hence about $O(knl \log l + n^2 l)$, which is a considerable decrease compared with Isomap, especially when the data size l is large. Besides, the required space to store geodesic distance matrix $D_{n,l}$, which is a submatrix of D_l, is $O(nl)$ evidently. This shows L-Isomap also decreases space complexity of Isomap considerably.

Like Isomap and L-Isomap, the time complexity of EL-Isomap is also mainly determined by geodesic distance estimation and MDS eigenvalue calculation. Therein, geodesic distances between landmarks and between nonlandmarks and the corresponding available landmarks (with size m) need to be estimated, and the time cost is about $O(kn^2 \log l + km(l - n) \log l)$. Yet different from the other two methods, EL-Isomap requires implementing MDS in two steps (Steps 2 and 3.3), all time cost of which is about $O(n^3 + m^3 l)$. Accordingly, the total time complexity of EL-Isomap is around $O(kn^2 \log l + km(l - n) \log l + n^3 + m^3 l)$. Since $kn^2 \log l + km(l - n) \log l < knl \log l$ holds, if we set $m^3 \le n^2$, then the time complexity costed by EL-Isomap is evidently comparable to that of L-Isomap, which both improve the time complexity of the original Isomap considerably.

Notice that the geodesic distance the matrix EL-Isomap requires storing is also an $n \times N$ matrix $D_{m,n,l}$. Yet, different from $D_{n,l}$ of L-Isomap, $D_{m,n,l}$ is comparatively sparse. In particular, in $D_{m,n,l}$, the elements corresponding to the geodesic distances between nonlandmarks and all unavailable landmarks are all 0. Hence the actually required storing space is $O(n^2 + m(l-n))$, which is smaller than L-Isomap, and of course, than Isomap.

To sum up, both time and space complexities of EL-Isomap are close to or even in excess of those of L-Isomap,

and both have a considerable improvement compared with Isomap. That is to say, inherited from L-Isomap, EL-Isomap also owns low computational complexity, which can make the method also be comparable with the local approaches for manifold learning in this point.

2.5. Relation between EL-Isomap and L-Isomap. Notice that as the size of the selected available landmark subset by EL-Isomap increases, the utilized information by EL-Isomap intends to be similar to those by L-Isomap. Then a problem arises: when the landmark subset utilized by EL-Isomap is selected degradedly to be the whole landmark set, how about the relationship between L-Isomap and EL-Isomap? The following theorem proves the equivalence of both methods in this degradation case.

Theorem 7 (degradation case of EL-Isomap). *In the case that the available landmark subset is selected as the whole landmark set, EL-Isomap is equivalent to L-Isomap.*

Proof. In this degradation case, it is easy to deduce that L_d', W, $\vec{\rho}_a$, and $\vec{\rho}_\mu$ (referring to (6), (9), and (7)) are translated to L_d, $L_d^{\#}$, $\vec{\delta}_a$, and $\vec{\delta}_\mu$ (referring to (2), (5), and (3)), respectively. Then the result (10) got by EL-Isomap is changed as

$$\begin{aligned}
\vec{y}_a &= L_d'\left(-\frac{1}{2}WW^T\left(\vec{\rho}_a-\vec{\rho}_\mu\right)+\vec{1}\cdot\frac{1}{m}\right)\\
&= -\frac{1}{2}\begin{bmatrix}\sqrt{\lambda_1}v_1^T\\ \sqrt{\lambda_2}v_2^T\\ \vdots\\ \sqrt{\lambda_d}v_d^T\end{bmatrix}\begin{bmatrix}\frac{v_1}{\sqrt{\lambda_1}} & \frac{v_2}{\sqrt{\lambda_2}} & \cdots & \frac{v_d}{\sqrt{\lambda_d}}\end{bmatrix}\\
&\quad\cdot\begin{bmatrix}\frac{v_1^T}{\sqrt{\lambda_1}}\\ \frac{v_2^T}{\sqrt{\lambda_2}}\\ \vdots\\ \frac{v_d^T}{\sqrt{\lambda_d}}\end{bmatrix}\left(\vec{\rho}_a-\vec{\rho}_\mu\right)\\
&= -\frac{1}{2}\begin{bmatrix}\frac{v_1^T}{\sqrt{\lambda_1}}\\ \frac{v_2^T}{\sqrt{\lambda_2}}\\ \vdots\\ \frac{v_d^T}{\sqrt{\lambda_d}}\end{bmatrix}\left(\vec{\rho}_a-\vec{\rho}_\mu\right)\\
&= -\frac{1}{2}L_d^{\#T}\left(\vec{\delta}_a-\vec{\delta}_\mu\right),
\end{aligned} \tag{41}$$

which is similar to the result calculated by L-Isomap according to (4).

The proof is completed. □

So far, we have presented EL-Isomap through comparing with L-Isomap (and Isomap) in various viewpoints. Yet as mentioned above, the algorithm of the new method is incomplete since the landmark selection strategies (Steps 2 and 3.1) are still not introduced. To complete the algorithm, the next section is specialized to construct the reasonable strategies for landmark selection of EL-Isomap.

3. The Strategies for Landmark Selection

Theorems 3–6 imply that, to best guarantee the reasonability of L-Isomap or EL-Isomap, the selected landmarks in Step 1 (and Step 3.1 for EL-Isomap) should possibly have faithfully estimated geodesic distances from other related data points; that is, the estimated distances should accord with the Euclidean distances between their low-dimensional correspondings approximately. The above implication forms the basic principle to formulate the strategies for landmark selection in Steps 1 and 3.1 of EL-Isomap algorithm. First, we introduce the landmark selection strategy for Step 1.

3.1. Selecting the Whole Landmark Set in Step 1. Several strategies for landmark selection of Step 1 for L-Isomap have been presented. Two of the most typical ones are random choice method [9], which chooses landmarks randomly from the given data, and LASSO regression method [24], which uses most important points for preserving the geometric curvature of the underlying manifold as landmarks. The principle underlying both methods is common: to find landmarks which can typically represent the whole data manifold. Hence we call this principle "typicality principle."

As mentioned in the beginning of this section, another consideration for EL-Isomap to designate landmark selection strategy is that the geodesic distances between these landmarks and other points should be estimated possibly faithfully. When the data manifold is globally isometric to a convex region of the low-dimensional Euclidean space, the geodesic distance between any data pair can approximately meet this requirement [25] and hence only typicality principle needs to be considered. However, in many applications of manifold learning, the data manifolds are of more complex geometries, like the manifold with intrinsic topology of a low-dimensional concave region on which the geodesic distances between faraway data pairs would tend to be mistakenly estimated (as A and B in Figure 1). In these cases, the landmarks should be selected preferably apart from manifold edges (as C or other big grey points shown in Figure 1) to mostly avoid the unfaithfully estimated geodesic distances. This principle can be called "faithfulness principle" for convenience.

Then a confliction occurs: typicality principle tends to select landmarks possibly scattered all over the manifold, while faithfulness principle inclines to construct landmark set assembled to the central part of the manifold. To compromise both principles, the following algorithm for landmark selection provides a tradeoff: first to construct a subset in

the approximate central part of the data manifold and then to choose the landmarks from the subset using random choice method or LASSO regression method. The details are given as follows.

Subalgorithm of EL-Isomap for Landmark Selection (supplement to Step 1)

Input. There is a given data set with size l in space R^N; desired number of landmarks n; the convexity coefficient $\mu \geq 1$.

Output. There is the landmark set.

Step 1. Choose $s = \lfloor \mu n \rfloor$ data points randomly from the input data set, denoted as $\Lambda' = \{\vec{x}_i'\}_{i=1}^{s}$.

Step 2. Estimate the geodesic distance d_{ij}' $(1 \leq i, j \leq s)$ between each data pair $\vec{x}_i'$ and $\vec{x}_j'$; calculate the circumcenter [26] of the chosen data

$$cx = \arg\min_{\vec{x}_j' \in \Lambda'} \left(\max_{\vec{x}_i' \in \Lambda'} \left(d_{ij}' \right) \right). \tag{42}$$

Let $\Xi = \{\overrightarrow{cx}\}$, $\Lambda = \{\overrightarrow{cx}\}$.

Step 3. Find the neighborhood data set Θ for all points of Ξ; let $\Xi = \Theta$ and $\Lambda = \Lambda \cup \Theta$.

Step 4. If more than n points in Λ' have been contained in Λ, go to *Step 5*; else, return to *Step 3*.

Step 5. Use random choice method or LASSO regression method to choose n landmarks from Λ, and output them as the landmark set.

In the above algorithm, the function of the convexity coefficient μ is to adjust the location of the final chosen central part Λ. It is evident that the more largely μ is valued, the smaller the calculated central part that Λ gets. Hence, when the data manifold is of the topology of a low-dimensional convex region, μ should be valued as the smallest value 1, while, in the opposite case, its value should be bigger. Practically, we select μ by experience or the utilizable prior information of the data manifold.

Notice that in Step 2 of the algorithm only s points are utilized to approximate the circumcenter of the whole data set. The reason to do so is avoiding the large computational complexity of geodesic distance estimation on the whole data set. Through only adopting $s = \lfloor \mu n \rfloor$ points, the approximate circumcenter is found within time cost $O(kn^2 \log l)$ and space cost $O(n^2)$, which have no influence on the computational complexity of the whole EL-Isomap algorithm.

3.2. Selecting an Available Landmark Subset in Step 3.1. The main aim of Step 3.1 of EL-Isomap algorithm is to specify the available landmarks, between which and the new input $\vec{x}_a$ the geodesic distances can be estimated faithfully (i.e., according with faithfulness principle). That is to say, the faraway landmarks from $\vec{x}_a$ (as C, D from A shown in Figure 2) should be possibly avoided. Thus we suggest a simple strategy for landmark selection in Step 3.1: choosing possibly near landmarks to the new input as the utilized landmarks. The details are listed as follows.

Subalgorithm of EL-Isomap for Available Landmark Selection (supplement to Step 3.1)

Input. There are the new input $\vec{x}_a$, the landmark set $\{\vec{x}_i\}_{i=1}^{n}$, and the desired number of available landmarks m.

Output. There is the available landmark subset corresponding to $\vec{x}_a$.

Step 1. Choose m data points randomly from the landmark set, denoted as $\Lambda' = \{\vec{x}_i'\}_{i=1}^{m}$.

Step 2. Estimate the geodesic distance d_{ai}' $(1 \leq i \leq m)$ between x_a and each landmark $\vec{x}_i'$; search the nearest point from the selected data:

$$nx = \arg\min_{\vec{x}_i' \in \Lambda'} \left(d_{ai}' \right). \tag{43}$$

Let $\Xi = \{\overrightarrow{nx}\}$, $\Lambda = \{\overrightarrow{nx}\}$.

Step 3. Find the neighborhood data set Θ for all points of Ξ; let $\Xi = \Theta$ and $\Lambda = \Lambda \cup \Theta$.

Step 4. If there are at least m points in Λ' that have been contained in Λ, go to *Step 5*; else, return to *Step 3*.

Step 5. Randomly select m points from Λ and return the selected set as the available landmark set corresponding to x_a.

In the algorithm, the reason why only m randomly selected landmarks are considered in Steps 1 and 2 is because we do not want to increase the computational complexity of EL-Isomap. Since the algorithm only needs to extra estimate and store geodesic distances between each new input and the corresponding available landmarks with size m (with time complexity $O(kml \log l)$ and space complexity $O(ml)$), the essential time and space complexities of EL-Isomap have not changed.

4. Experiment Results

This section mainly aims at demonstrating performance of EL-Isomap on representational capacity extending, by comparisons with Isomap, L-Isomap, LLE, and Laplacian eigenmap. In particular, two series of simulations are designated: one is on data manifolds with intrinsic topologies of concave regions (concave case) and the other is on data manifolds with intrinsic loops (loop case).

4.1. Concave Case. The first applied data set in this series of simulations is composed of 3000 points sampled from a low-dimensional manifold with the shape shown as Figure 1(a). The manifold is formed by combination of two crossed S-curves, which are wrapped from two 2D perpendicularly

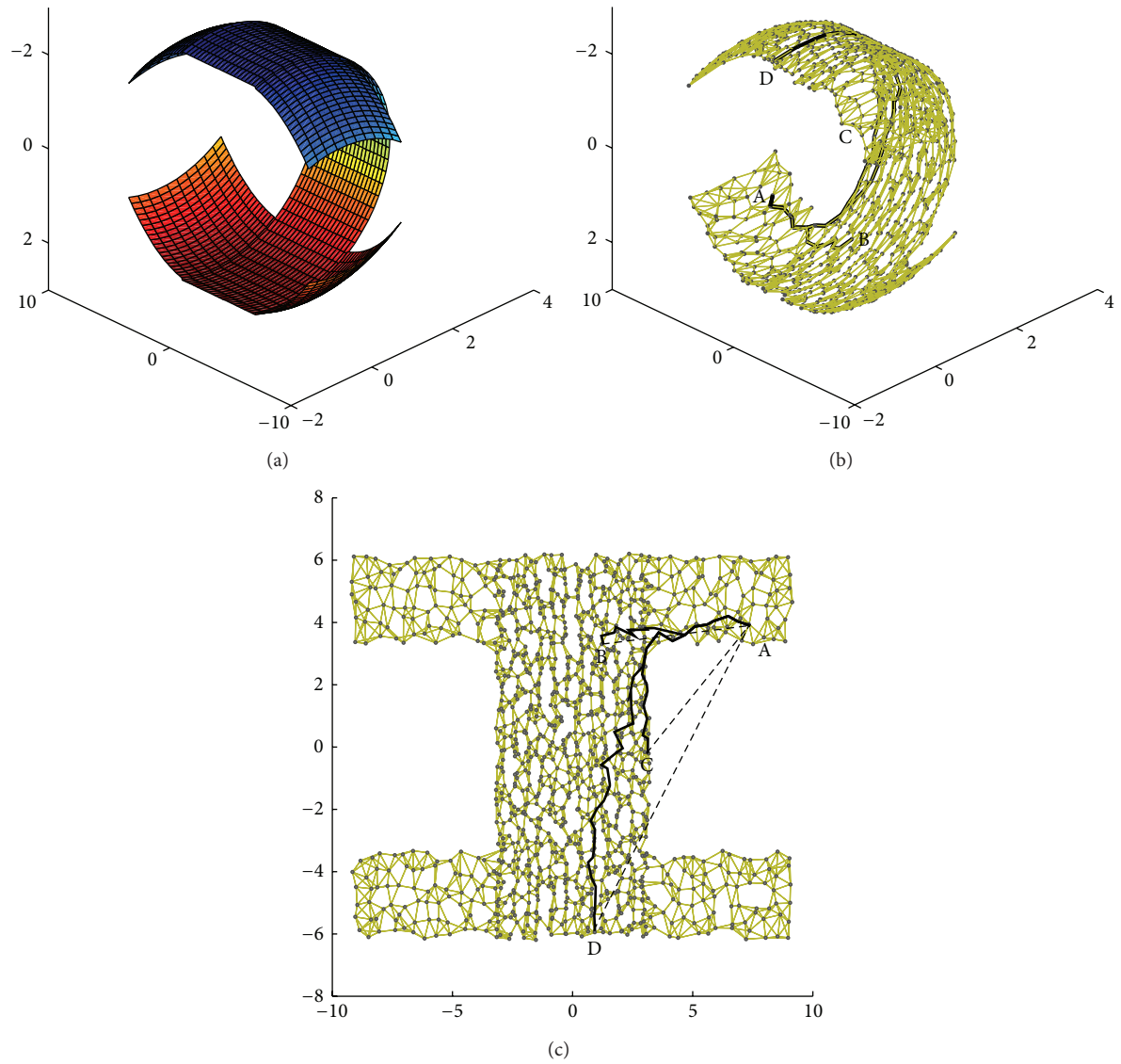

Figure 2: (a) is a 3D manifold with "H" form; (b) is a data set with 1000 vertices sampled from the manifold of (a), superimposed with 5-NN neighborhood graph; A, B, and C are three vertices lying on the manifolds; the real curves are estimated geodesic curves between A and B, A and C, and A and D, respectively; (c) is the unwrapped 2D corresponding points of (b); the broken lines are Euclidean lines between A and B, A and C, and A and D, respectively.

crossed rectangles. The unwrapped 2D points corresponding to the adopted 3D data are showed in Figure 3. The 2D embeddings of the data set were calculated by LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap. Each method was executed multiple times under different presetting parameters and the best performance is demonstrated in Figure 3. The neighborhood sizes of LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap are set as 16, 12, 8, 8, and 8, respectively. The numbers of landmarks of L-Isomap and EL-Isomap are both set as 400. The convex coefficient μ is set as 4 (in subalgorithm for Step 1) and the number of available landmarks for each new input as 20 (in subalgorithm for Step 3.1).

From Figure 3, it can be seen that both LLE and Laplacian eigenmap preserve the local neighborhood configurations of the original data set. However, neither of them represents the intrinsic global geometry of the data manifold very well. Particularly, one of the crossed parts is embedded flexurally by LLE and the whole embeddings calculated by Laplacian eigenmap highly shrink. Isomap, L-Isomap, and EL-Isomap better maintain the global figures of the whole data embedding. Yet L-Isomap and EL-Isomap outperform Isomap in two points: firstly, the data embeddings in the vertical rectangle calculated by Isomap are pinched, while the approximate correct shape of this rectangle can be clearly recognized from the embeddings got by L-Isomap and EL-Isomap; secondly, the embeddings of the crossed two parts got by Isomap are not perpendicular, while the approximate perpendicularity of two parts can be easily observed from embeddings obtained by L-Isomap and EL-Isomap. This can

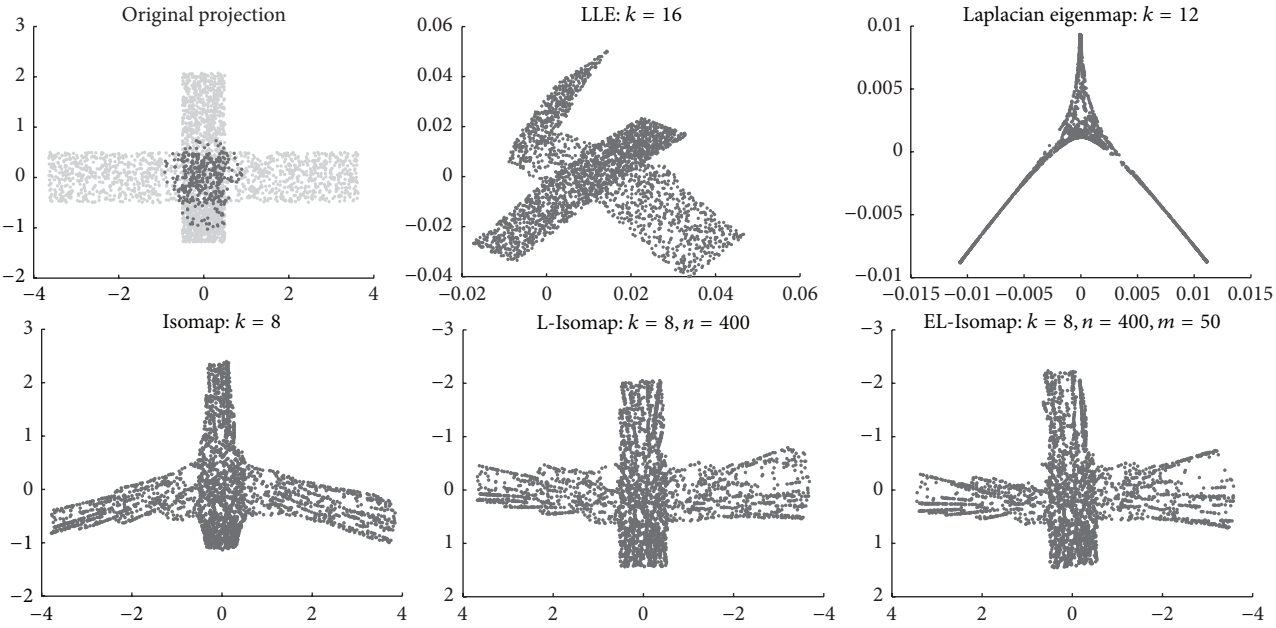

Figure 3: 2D embeddings of the crossed S-curve data calculated by LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap, respectively. The fuscous points shown in the top left figure are the adopted landmarks by L-Isomap and EL-Isomap.

be explained as follows: Isomap utilizes geodesic distances estimated between all data pairs, many of which are inevitably unfaithful. Yet since the landmark set is located on the central part of the whole manifold (as shown in Figure 3), the geodesic distances used by L-Isomap and EL-Isomap are much more faithful. Based on the theories presented in Section 2.3, L-Isomap and EL-Isomap reasonably perform better.

The second utilized data set contains 3000 points generated from a manifold with intrinsic 2D concave H-like region (as shown in Figure 2(a)). LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap were implemented, respectively, on the data set multiple times. The best performance is demonstrated in Figure 4. The adopted neighborhood sizes of LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap are 12, 10, 8, 8, and 8, respectively. The number of landmarks in L-Isomap and EL-Isomap is 800 and that of the available landmarks in Step 3.1 of EL-Isomap is set as 50. μ in subalgorithm of Step 1 is set as 2.

It is shown clearly in the figure that LLE and Laplacian eigenmap commonly preserve the local topologies of neighborhood regions of the original data. However, the global figures of the whole data set are not well revealed by both methods. LLE distorts the configuration of the left part of the manifold and Laplacian eigenmap pinches the global frame to the center of the manifold. The performances of Isomap and L-Isomap are very similar, both of which approximately keep the global structure of the whole manifold, except that the embeddings in four corner parts outspread abnormally. EL-Isomap alleviates this problem to a certain extent. In particular, the perpendicular relations between each of the three corner parts (except top right corner) and the central part can be observed from the embeddings calculated by EL-Isomap. This can be explained by the fact that less unfaithful estimated geodesic distances are applied by EL-Isomap than those by Isomap and L-Isomap, which guarantees that EL-Isomap more robustly finds reasonable embeddings of the data set.

The value preset for the desired number of the available landmarks corresponding to each new input has significant influence on the final performance of EL-Isomap. To intuitively exhibit this influence, the performance of EL-Isomap under different values of this parameter (from 5 to 800) on the above data set is listed in Figure 5. The other parameters are set similar to the above simulation.

From Figure 5, it is clear that, to promise the good performance of EL-Isomap, the value for the number of available landmarks can not be set too large or too small. When it is valued as a big number, EL-Isomap tends to degrade as L-Isomap according to Theorem 7. This can be clearly observed from the embeddings as $m = 800$ in Figure 5, which are the same as the embeddings shown in Figure 4 got by L-Isomap. Since in this case more unfaithful geodesic distances will be possibly involved, EL-Isomap will perform abnormally especially for the data manifolds with complex topologies, while if the parameter is set too small, EL-Isomap also inclines to perform badly, as shown by the embeddings of $m = 5$ in Figure 5. This can be explained by the presumption for the reasonability of EL-Isomap in Theorems 5 and 6. When m is set too small, the affine space $E_a^{m,d}$ related to the available landmark subset (as defined in Section 2.3) tends to be degenerated, so that this space can not be promised to be

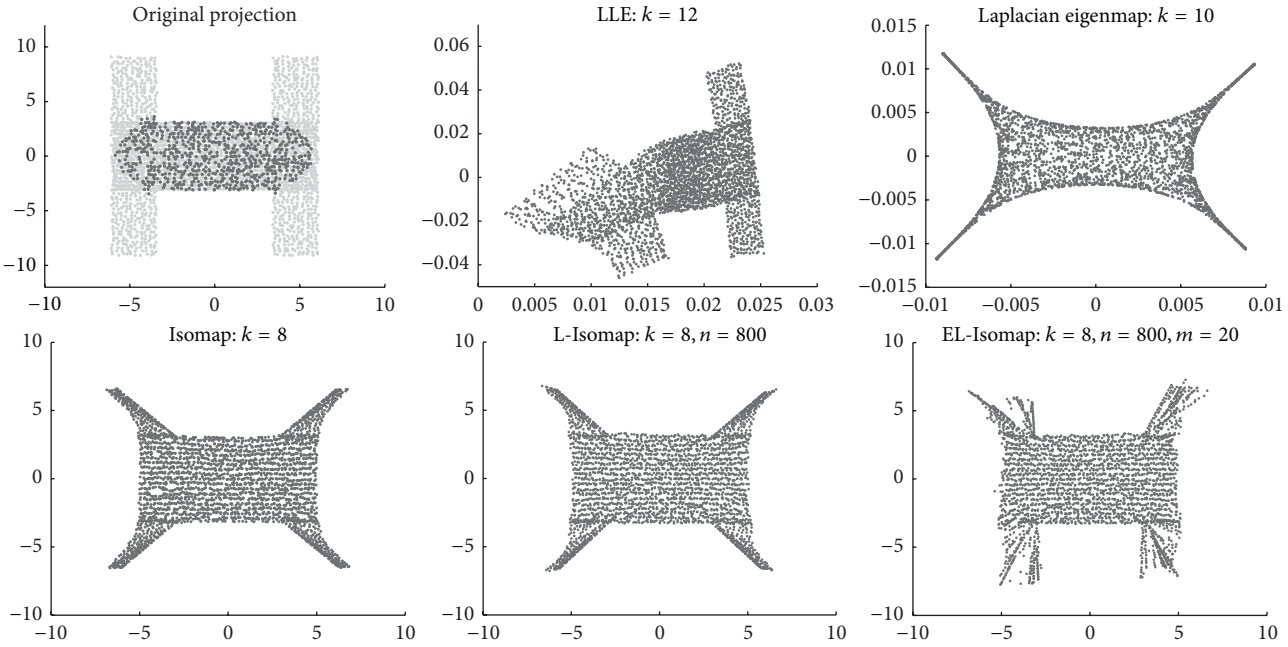

Figure 4: 2D embeddings of the H-form manifold data calculated by LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap, respectively. The fuscous points shown in the top left figure are the adopted landmarks by L-Isomap and EL-Isomap.

consistent with the space $E_a^{n,d}$ (also defined in Section 2.3). Then the accuracy and robustness of EL-Isomap are not guaranteed to be satisfied, which leads to the abnormal performance.

From the above simulations, it can be observed that both L-Isomap and EL-Isomap expend the representational capability of the original Isomap to the manifolds with concave topologies, and EL-Isomap can be implemented effectively in a more extensive range of data manifolds. Other two cases on which L-Isomap and EL-Isomap are effective are the data manifolds as shown in Figures 6(a) and 6(b). EL-Isomap also takes effect on manifolds like Figures 6(c) and 6(d). These four extra cases are all demonstrated to let the capabilities of the two methods in concave cases be more learned.

4.2. Loop Case. This section mainly demonstrates the representational capability of the new method on manifolds with intrinsic loops. The first two applied manifolds are cylinder and sphere manifolds (as shown in Figure 7), which are two typical loopy manifolds. Two data sets with 1000 and 3000 points were generated from the cylinder and sphere manifolds, respectively, as the test sets. LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap were implemented on the data sets multiple times, respectively. Figures 8 and 9 show the best performance of each method, where the numbers of neighborhood sizes are set as 10, 8, 7, 7, and 7 (for cylinder data) and 12, 8, 8, 8, and 8 (for sphere data); correspondingly, the numbers of landmarks are set as 200 (for cylinder data) and 300 (for sphere data), and the desired numbers of available landmarks are set as 40 (for cylinder data) and 30 (for sphere data). Both methods set convex coefficients μ as 2.

From Figure 8, it can be easily observed that all LLE, Laplacian eigenmap, Isomap, and L-Isomap are disabled to find the correct embeddings of the cylinder data. To the opposite, EL-Isomap has a more prominent performance on this data set: on one side, the calculated embeddings by the new method successfully preserve the local geometry of the original data set, which can be seen by the continuous changing colors of the neighborhood points in the bottom right figure of Figure 8. On the other side, the global figure of the embeddings is perfect, which exhibits as an approximate $2 \times 2\pi$ rectangle, just similar to the exact unfolded mapping of the original data in 2D space.

The performances of five methods on sphere data are similar to those on cylinder data (as shown in Figure 9). EL-Isomap also outperforms the other four methods both in local geometry preservation and global configuration maintenance.

A real-world data set intrinsically located on the manifold with loops is also adopted, which contains 390 images of a cartoon pig, each of which is a 140×120 pixel (16800-dimensional) grayscale picture, as shown in the top left figure in Figure 10. The image set contains 13 sequences of pig images, and each sequence is pictured by rotating the camara in isometric 30 locations on a circle, which certainly leads to loops in the underlying data manifold. Like the above simulations, we execute LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap on the data set multiple times and select ones with the best performance as demonstrations (as shown in Figure 10). The neighborhood sizes of the five methods are set as 9, 10, 5, 5, and 6, respectively. The desired number of landmark set of L-Isomap and EL-Isomap is set as 150 and that of the available landmark subset of EL-Isomap is set as 20. The convex coefficient μ is set as 2.

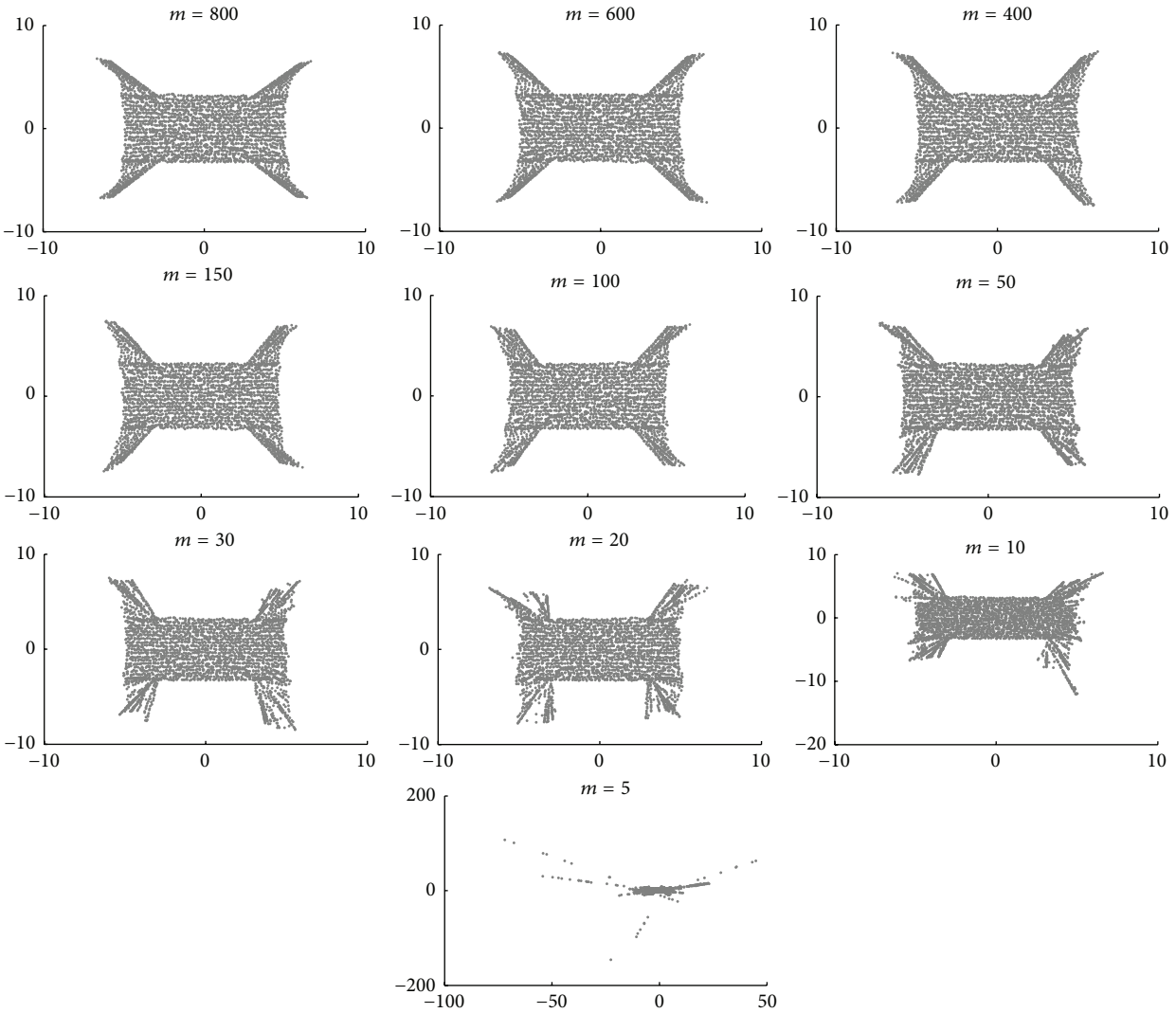

Figure 5: 2D embeddings of the H-form manifold data calculated by EL-Isomap as the numbers of available landmarks are set as 800, 600, 400, 150, 100, 50, 30, 20, 10, and 5, respectively.

The locations of a sequence of images, which forms a loop, are denoted as circled points in 2D embeddings got by five methods, respectively. The corresponding images are shown on top of each embedding, with increasing order of the first coordinates of their corresponding embeddings.

Figure 10 shows clearly that all LLE, Laplacian eigenmap, and Isomap do not unwrap the loops existing in the data manifold, and hence the rotation feature is not explored by any of the above methods. L-Isomap alleviates this problem to a certain extent: one cannot find obvious loops from the embeddings got by the method. However, essentially the rotation feature is still not discovered, which can be observed from the image sequence shown in the figure. EL-Isomap outperforms the other four methods, which correctly finds the rotation feature, corresponding to the first coordinates of the embeddings. This further verifies the effectiveness of EL-Isomap on manifold with essential loops.

Then why is EL-Isomap successful on manifolds with loops? This problem can be explained as follows. EL-Isomap only considers two kinds of interpoint connections: the connections between data pairs located in the approximately central part of the manifold and those between the points in the noncentral part and some nearest ones in the central part, which intrinsically construct a connected component based on the process of Step 3.1. These connections implicitly construct a graph superimposed on the whole data set, which is intrinsically utilized by EL-Isomap. A nice character of this graph is that it implicitly breaks the essential loops existing in the underlying manifold, which intrinsically leads to the effectiveness of EL-Isomap on manifold with loops.

5. Discussion

Based on its reasonability theory, the performance of EL-Isomap highly depends on the selection of the landmark

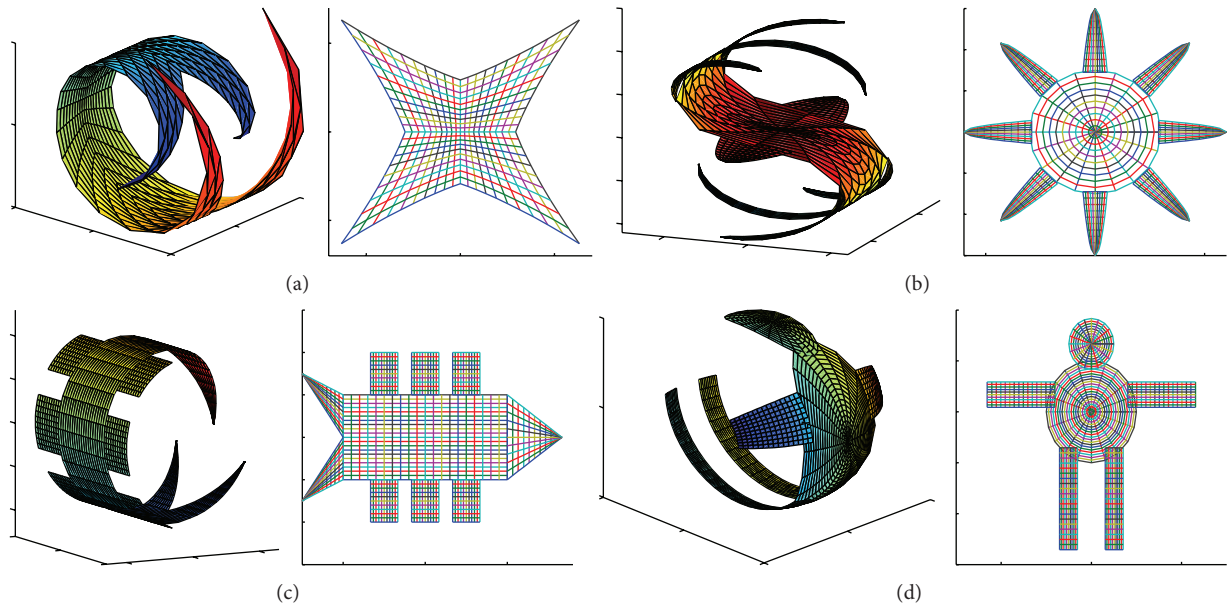

Figure 6: (a), (b), (c), and (d) demonstrate 4 manifolds with intrinsic topologies of low-dimensional concave regions. Each left figure shows the shape of the 3D manifold, and the right corresponding figure is its unwrapped projection in 2D space.

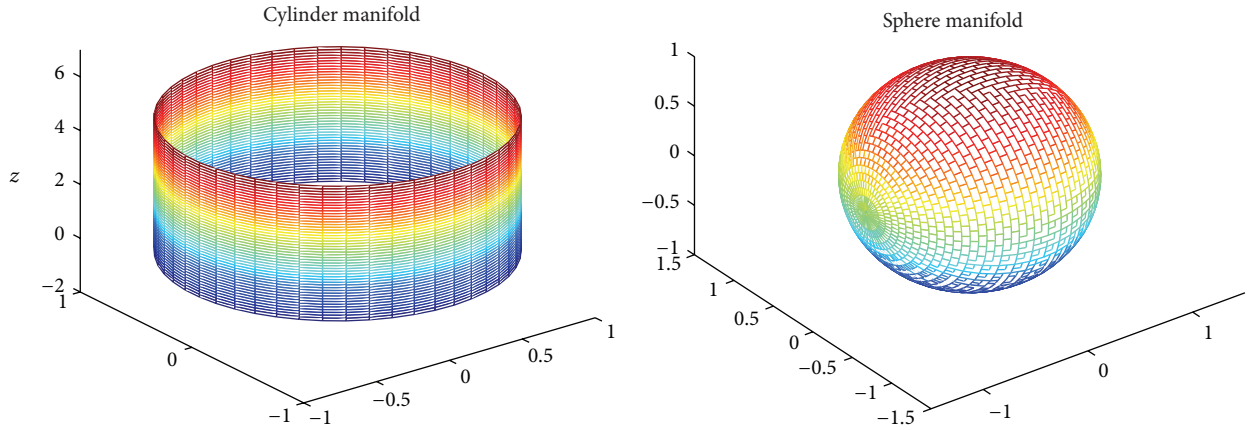

Figure 7: Demonstrations of cylinder and sphere manifolds.

set (Step 1) and the available landmark subset to the new input (Step 3.1). Essentially speaking, the aims of two steps are to select landmarks so as to make the geodesic distance matrix of the landmark set and the geodesic distance vector between the available landmarks and the new input comply with Euclidean condition possibly, that is, according to the "faithfulness principle" mentioned in Section 3. Through selecting landmarks from central part of the manifold and from the nearest points of all landmarks, we approximately realize the faithfulness principle for Steps 1 and 3.1. However, in some cases, the above strategies might not guarantee the corresponding geodesic distances to accord with Euclidean condition. Take Figure 4 as an example: although the global configuration of the original manifold has been represented by EL-Isomap, one can easily find that there still exist some abnormal embeddings. This is because geodesic distances between some nonlandmarks and the nearest landmarks to them still deviated from Euclidean condition more or less.

A direct way to solve this problem is to calculate the maximal submatrix, to let it be an approximate Euclidean matrix, from the geodesic distance matrix of the whole data set. In fact, some methods have been presented to solve this problem theoretically [27]. Adopting these methods will improve the accuracy and robustness of EL-Isomap due to Theorems 5 and 6. However, applying these methods will increase the computational complexity of the algorithm of EL-Isomap significantly. Firstly, the geodesic distance matrix of all data has to be calculated, which increases the time and space complexities of EL-Isomap according to the analysis in Section 2.4. Secondly, the model to find the approximate

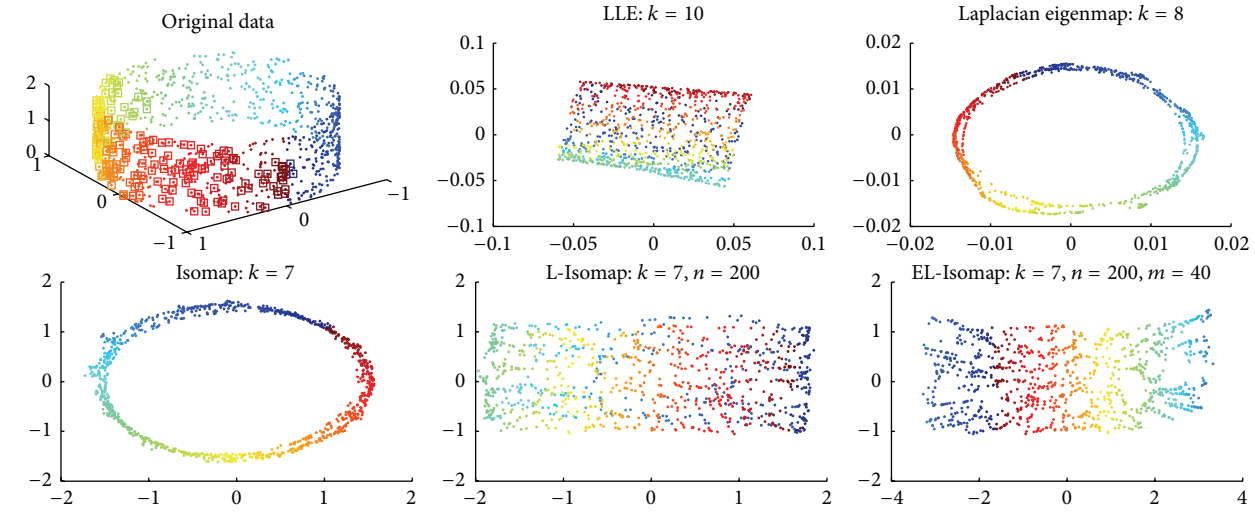

FIGURE 8: 2D embeddings of the cylinder manifold data calculated by LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap, respectively. The squared points shown in the top left figure are the adopted landmarks by L-Isomap and EL-Isomap.

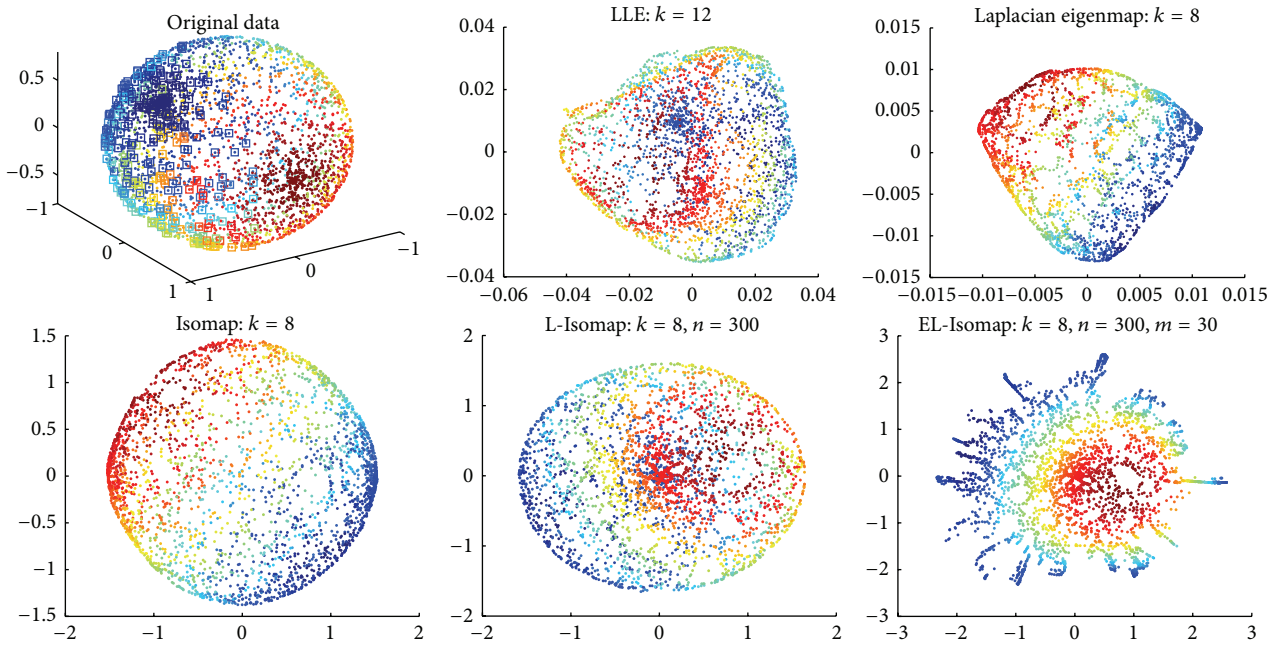

FIGURE 9: 2D embeddings of the sphere manifold data calculated by LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap, respectively. The squared points shown in the top left figure are the adopted landmarks by L-Isomap and EL-Isomap.

Euclidean distance submatrix will cost polynomial time complexity [27], which is also a considerable increase to EL-Isomap algorithm. Since our main aim is to let EL-Isomap have improvement in both issues (representational capability and computational efficiency), the simpler strategies adopted in the research are still preferred. However, constructing fast and effective heuristic method to pick up landmarks which approximately comply with Euclidean condition is still an important issue of our future research to further improve computational capability of EL-Isomap.

Note that how to select a proper neighborhood size k is also a very important problem. In practice, especially when the data set is large, selecting parameters is generally based on experience because of its high efficiency. Actually, in all of our experiments, the neighborhood sizes k of the proposed algorithm on all our experiments were easily set around 8 by experience. Nevertheless, to completely automate the proposed algorithm, constructing an efficient parameter selection strategy is still necessary. Currently, methods such as the "trial-and-error" method [28] and the neighborhood contraction and expansion method [29] have been developed to adaptively determine a reasonable neighborhood size in local-to-global context. These methods should be examined in further research. Besides, further investigation still needs

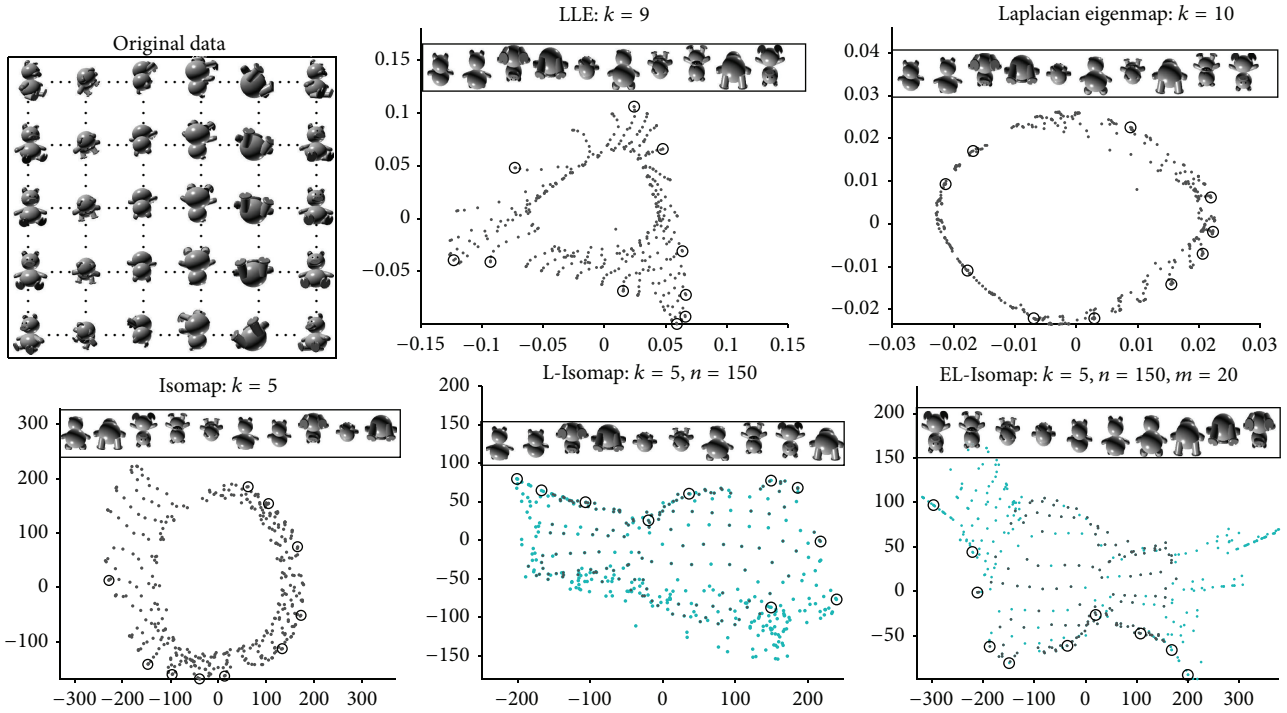

FIGURE 10: 2D embeddings of the pig image data calculated by LLE, Laplacian eigenmap, Isomap, L-Isomap, and EL-Isomap, respectively. The images shown on top of each embeddings are images corresponding to the circled points with increasing order of their first coordinates. The fuscous points shown in the embeddings got by L-Isomap and EL-Isomap are the adopted landmarks by the methods.

to be made to extend our loop detection theory to a wide range of manifold types, such as the nonmetric or the holed manifolds.

6. Conclusion

In this paper, we have proposed a new manifold learning method called EL-Isomap. The proposed method mainly has twofold contributions. On one hand, it possesses the advantage of the previous local approaches on computational efficiency. On the other hand, it inherited the advantage of the current global approaches on representation capability. Particularly, originated from L-Isomap, EL-Isomap naturally has a similar computational efficiency property to L-Isomap, and, through constructing reasonable strategies for landmark selection, EL-Isomap significantly extends the effective range of data manifolds. Another contribution of this work is to give the reasonability theory, that is, accuracy and robustness of EL-Isomap, which provides theoretical foundation for the new method. The computational complexity of the new method and the relation between it and L-Isomap have also been analyzed. Through extensive experiments implemented on synthetic and real-world data sets, the capability of EL-Isomap has been verified to outperform the state-of-the-art approaches along this line, especially for manifolds with complex shapes.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant nos. 61373114, 11131006, 91330204, and 11471006), National Basic Research Program of China (973) (Grant no. 2013CB329404), and Industrial Project of Shaanxi Province (Grant no. 2013K06-03).

References

[1] 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.

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

[3] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[4] Y. Leung, D. Meng, and Z. Xu, "Evaluation of a spatial relationship by the concept of intrinsic spatial distance," *Geographical Analysis*, vol. 45, no. 4, pp. 380–400, 2013.

[5] Z. Han, D. Y. Meng, Z. B. Xu, and N. N. Gu, "Incremental alignment manifold learning," *Journal of Computer Science and Technology*, vol. 26, no. 1, pp. 153–165, 2010.

[6] D. Y. Meng, Y. Leung, Z. B. Xu, T. Fung, and Q. F. Zhang, "Improving geodesic distance estimation based on locally linear assumption," *Pattern Recognition Letters*, vol. 29, no. 7, pp. 862–870, 2008.

[7] D. Y. Meng, Y. Leung, and Z. B. Xu, "Detecting intrinsic loops underlying data manifold," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 2, pp. 337–347, 2013.

[8] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *Journal of Machine Learning Research*, vol. 4, no. 2, pp. 119–155, 2004.

[9] V. de Silva and J. B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reductionction," in *Proceedings of the Neural Information Processing Systems (NIPS '02)*, vol. 15, pp. 705–712, 2002.

[10] J. A. Lee, A. Lendasse, and M. Verleysen, "Nonlinear projection with curvilinear distances: isomap versus curvilinear distance analysis," *Neurocomputing*, vol. 57, no. 1–4, pp. 49–76, 2004.

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

[12] D. Y. Meng, Y. Leung, and Z. B. Xu, "Evaluating nonlinear dimensionality reduction based on its local and global quality assessments," *Neurocomputing*, vol. 74, no. 6, pp. 941–948, 2011.

[13] Z. Y. Zhang and H. Y. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *Journal of Shanghai University. English Edition*, vol. 8, no. 4, pp. 406–424, 2004.

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

[15] J. C. Nascimento, J. G. Silva, J. S. Marques, and J. M. Lemos, "Manifold learning for object tracking with multiple nonlinear models," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1593–1605, 2014.

[16] V. de Silva and J. B. Tenenbaum, "Sparse multidimensional scaling using landmark points," Tech. Rep., Stanford University, 2004.

[17] J. Venna and S. Kaski, "Local multidimensional scaling," *Neural Networks*, vol. 19, no. 6-7, pp. 889–899, 2006.

[18] J. A. Lee and M. Verleysen, "Nonlinear dimensionality reduction of data manifolds with essential loops," *Neurocomputing*, vol. 67, no. 1–4, pp. 29–53, 2005.

[19] D. Y. Meng, Y. Leung, T. Fung, and Z. B. Xu, "Nonlinear dimensionality reduction of data lying on the multicluster manifold," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol. 38, no. 4, pp. 1111–1122, 2008.

[20] D. Meng, Y. Leung, and Z. Xu, "Passage method for nonlinear dimensionality reduction of data on multi-cluster manifolds," *Pattern Recognition*, vol. 46, no. 8, pp. 2175–2186, 2013.

[21] L. Yang, "Building k-connected neighborhood graphs for isometric data embedding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 827–831, 2006.

[22] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 1989.

[23] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*, Chapman & Hall, 2nd edition, 2001.

[24] J. G. Silva, J. S. Marques, and J. M. Lemos, "Selecting landmark points for sparse manifold learning," in *Proceedings of the Neural Information Processing Systems (NIPS '05)*, 2005.

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

[26] M. Brückner, *Large margin Kernel machines for binary classification [Diplom thesis]*, 2005.

[27] A. Y. Alfakih, A. Khandani, and H. Wolkowicz, "Solving euclidean distance matrix completion problems via semidefinite programming," *Computational Optimization and Applications*, vol. 12, no. 1–3, pp. 13–30, 1999.

[28] M. Balasubramanian and E. L. Schwartz, "The isomap algorithm and topological stability," *Science*, vol. 295, no. 5552, p. 7, 2002.

[29] J. Wang, Z. Zhang, and H. Zha, "Adaptive manifold learning," in *Advances in Neural Information Processing Systems: Proceedings of the NIPs*, vol. 17, pp. 1473–1480, MIT Press, 2005.