

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Manifold topological multi-resolution analysis method

Shaodi You, Huimin Ma*

3D Image System Simulation Lab, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:

Received 23 March 2010

Received in revised form

15 December 2010

Accepted 31 December 2010

Available online 12 January 2011

Keywords:

Manifold topological analysis

Manifold multi-resolution analysis

Locally linear embedding

ABSTRACT

In this paper, two significant weaknesses of locally linear embedding (LLE) applied to computer vision are addressed: “intrinsic dimension” and “eigenvector meanings”. “Topological embedding” and “multi-resolution nonlinearity capture” are introduced based on mathematical analysis of topological manifolds and LLE. The manifold topological analysis (MTA) method is described and is based on “topological embedding”. MTA is a more robust method to determine the “intrinsic dimension” of a manifold with typical topology, which is important for tracking and perception understanding. The manifold multi-resolution analysis (MMA) method is based on “multi-resolution nonlinearity capture”. MMA defines LLE eigenvectors as features for pattern recognition and dimension reduction. Both MTA and MMA are proved mathematically, and several examples are provided. Applications in 3D object recognition and 3D object viewpoint space partitioning are also described.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Manifold learning has attracted attention as a nonlinear learning method in the past decade. As illustrated in Fig. 1(a), it hypothesizes that high dimensional data samples are distributed on an underlying manifold, which is controlled by a few parameters [3,5,51]. Manifold learning techniques try to simplify the high dimensional data by recovering the underlying low dimensional manifold. Local methods are a subset of manifold learning techniques that try to recover global manifold structure from local linear space information. This idea is summarized as “Think globally, fit locally” by L.K. Saul et al., who proposed locally linear embedding (LLE) [1,2]. After LLE, several new methods have been proposed to improve geometrical reconstruction performance, such as Laplacian eigenmaps (LEM) [6], Hessian LLE (HLE) [7], diffusion maps (DFM) [9] and local tangent space alignment (LSTA) [10]. All these algorithms output a set of eigenvectors and eigenvalues. The basic idea of these local methods is summarized as Fig. 1(b). As illustrated in Fig. 1(c), there are two main fields of applications based on local methods, and LLE in particular: (1) choosing the right eigenvectors, underlying manifold structure can be reconstructed for tracking and perception understanding [23,27–34,36,37,39]; (2) eigenvectors can also be considered as low-dimensional data features for pattern recognition [10,13–16,18,24, 26,35,45,52].

Local methods are popular because they have good computational efficiency than other manifold learning techniques [22]. However, researchers found significant weaknesses in applications: (1) as illustrated in Fig. 1(c), local methods output a set of eigenvectors and eigenvalues. To recover the manifold, right eigenvectors should be chosen. Existing local methods choose the eigenvectors based on the “intrinsic dimension” of manifold. However, they fail to give a robust method to determine the “intrinsic dimension”. (2) Local methods output many eigenvectors, but only few of them contribute to recovering the manifold. The meanings of the rest of the eigenvectors are unclear. Some research [2,14–16,52] consider the local methods as dimension reduction techniques. The eigenvectors are considered to have similar meanings as those eigenvectors from linear dimension reduction techniques (such as PCA). However, the results from pattern recognition experiments are negative.

LLE is the oldest local method. LEM [6], HLE [7], DFM [9] and LSTA [10] have attempted to achieve improved geometrical reconstruction performance. However, all these methods are limited to local linear relation capture and global optimization. None of them tried to solve the “intrinsic dimension” problem or interpret the eigenvectors. Referring to Fig. 1, their works are limited in part (b), they cannot solve the application problems in part (c).

The “intrinsic dimension” is derived from an analogous feature from global manifold learning methods, such as multidimensional scaling (MDS) [46], Isomap [4,19] and maximum variance unfolding (MVU) [16]; or from linear dimension reduction techniques, such as principle component analysis (PCA). Current research suggests that local methods cannot determine the “intrinsic dimension” by examining eigenvalues like global methods [1,2]. Kégl [20] and Levina and Bickel [21] proposed “intrinsic dimension” determining

* Corresponding author. Tel.: +86 10 62781432.

E-mail addresses: youshaodi@gmail.com (S. You), mhmpub@tsinghua.edu.cn (H. Ma).

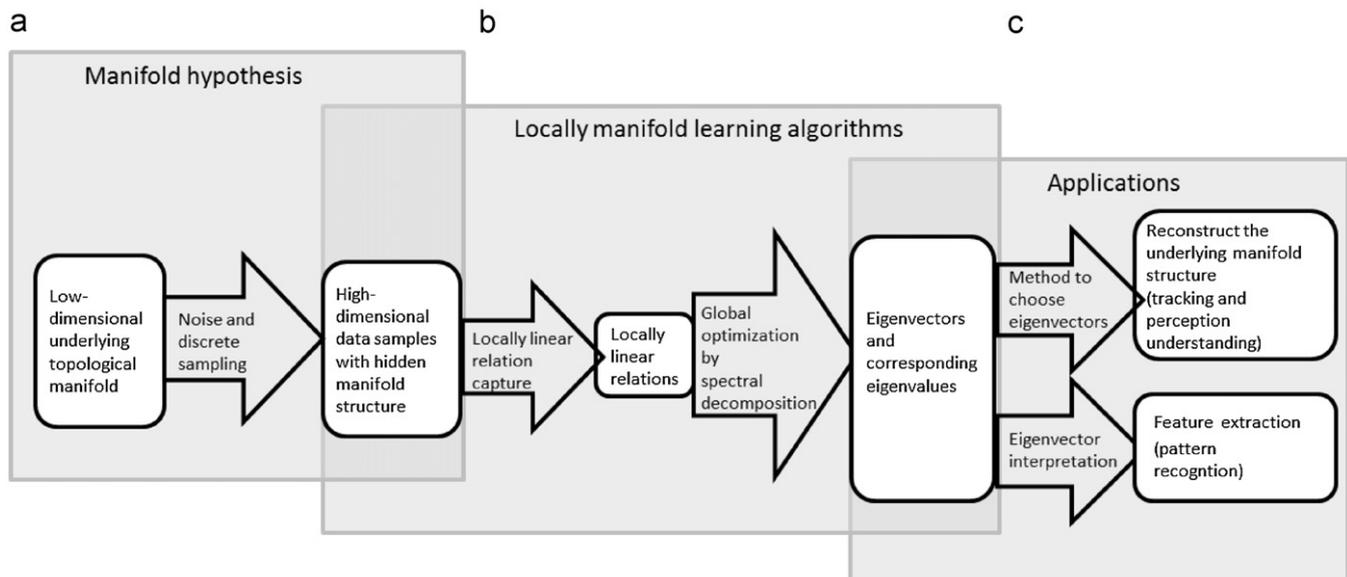


Fig. 1. Framework of locally manifold learning. (a) Manifold hypothesis: high dimensional data samples are distributed on an underlying manifold, which is controlled by a few parameters [3,5]. (b) Locally manifold learning algorithms: locally linear relation between samples is captured. Then a low-dimensional solution preserving the captured relation is found by spectral decomposition. A set of eigenvectors with corresponding eigenvalues from spectral decomposition is given as output. (c) Main applications of manifold learning. The upper branch is to reconstruct the hidden manifold structure by choosing the right eigenvectors. The lower branch is using eigenvectors as features based on the meaning of eigenvectors.

methods based on the mathematical definition of manifold, but their results are not in accordance with those from global manifold learning methods [4,16]. We found different researchers have different understanding on “intrinsic dimension”. We broaden the research on this topic by introducing the topological manifold [44]. In mathematics, a manifold in Euclidean space has two dimensions: (1) the “local dimension”—the dimension of manifold, and (2) the “global dimension”—the minimum Euclidean space embedding dimension (MESED). We clarify the misunderstanding on “intrinsic dimension” that locally dimension determining methods [20,21] find the “local dimension”, while global manifold learning methods [4,16] find the “global dimension”. To definitively determine the structure of a manifold, local manifold learning methods should find the “global dimension”. In our research, we propose the manifold topological analysis (MTA) method that can determine the “global dimension” of typical topological manifolds.

The ambiguous meaning of eigenvectors hinders application of LLE in pattern recognition. LLE performs worse than linear methods when many LLE eigenvectors are used, specifically more than 6 eigenvectors according to Saul and Roweis’ [2] research in digit recognition. It was generally considered that LLE eigenvectors that are not contributing to manifold structure capture only clustering and noise, which are derived from an analogous feature in global manifold learning methods [52]. However, by studying locally linear spaces [42,47], topology [44] and general function spaces [43,48], we have found the multi-resolution property of LLE eigenvectors: LLE eigenvectors are multi-resolutional; they capture manifold nonlinearity in the corresponding resolution. We prove this property and develop the manifold multi-resolution analysis (MMA) method. Furthermore, we successfully apply our method in 3D object recognition and 3D object viewpoint space partitioning.

Other research have also found local methods that are not robust against data rescaling and normalization; as a result, global geometric information is unrecoverable [25]. Current local methods are unreliable because they use eigenvalues to choose eigenvectors. Our manifold topological analysis (MTA) method

does not have this problem because the eigenvectors are chosen by topological analysis. We also develop a subroutine included in MTA that scales eigenvectors by eigenvalues, thus, it can recover global geometric information.

Although manifold learning has become popular in the past decade, most research has not fully considered the “manifold hypothesis” [3] (Fig. 1(a)). This “hypothesis” is usually summarized with very few sentences without proof, such as “suppose data samples have an underlying manifold structure”. In some researches, local methods have been disproved [11]. We have found that their data samples do not strictly satisfy the “manifold hypothesis”. In order to make our research strictly follow the “manifold hypothesis”, we introduce a sufficient condition for data samples to satisfy the hypothesis.

In this paper, “topological embedding” and “multi-resolution nonlinearity capture” are introduced. The proposed MTA method is based on “topological embedding” to robustly determine the “global dimension” of a manifold with typical topology, which is important for tracking and perception understanding. The proposed MMA method is based on “multi-resolution nonlinearity capture”. MMA defines LLE eigenvectors as features for pattern recognition. Both MTA and MMA are proved mathematically, and several examples are included. We also examine the method performances in applications in 3D object recognition and 3D object viewpoint space partitioning. This paper has the following organization: In Section 2, we introduce the mathematical definition of topological manifolds and give a sufficient condition to satisfy the “manifold hypothesis”. Then, we summarize the existing LLE algorithms in matrix form. In Section 3, we mathematically analyze the topological manifold and LLE algorithm. Furthermore, we describe “topological embedding dimension” based on the analysis. In Section 4, the MTA method is introduced for typical topology and mathematically proved. This method successfully determines the “global dimension” of typical manifolds. In Section 5, we provide meaning of LLE eigenvectors based on “topological embedding” and general function spaces: LLE eigenvectors are multi-resolutional and capture manifold

nonlinearities. Furthermore, we propose the MMA method with mathematical proofs. In Section 6, we apply our methods in 3D object recognition and 3D object viewpoint space partitioning. We discuss the results and a previous work in Section 7.

2. Sufficient condition for the “manifold hypothesis” and locally linear embedding

In Section 2.1, we first introduce some basic definitions about manifolds and give a sufficient condition for data samples to satisfy the “manifold hypothesis”. In some researches, the data samples do not strictly satisfy the “manifold hypothesis” [11]. Therefore, definite data sample constraints are necessary. In Section 2.2, the original LLE algorithm is quickly introduced in matrix form. Because the LLE algorithm is very familiar, this step is intended to make the algorithm more convenient for further discussions and proofs. The “intrinsic dimension” problem is introduced by an example in the end.

2.1. A sufficient condition for the “manifold hypothesis”

2.1.1. Mathematical definitions

We introduce the following three important definitions that are related to manifolds:

Continuous map: If X and Y are topological spaces, a map $f: X \rightarrow Y$ is said to be continuous if for every open set $U \subset Y$, $f^{-1}(U)$ is open in X .

Homeomorphism: If X and Y are topological spaces, a homeomorphism from X to Y is defined to be a continuous bijective map $\phi: X \rightarrow Y$ with continuous inverse.

Manifold in Euclidean space: A k -dimensional manifold (or a k -dimensional surface) in \mathbb{R}^n is a subset $S^k \subset \mathbb{R}^n$ such that each point has a neighborhood in S^k that is homeomorphic to \mathbb{R}^k .

Topological spaces, the Hausdorff property and second countability are also important in defining manifolds. However, they are not used directly in our proofs. We leave out their definitions [42,44] to make this paper succinct.

We introduce two important dimensions of manifold:

Dimension of manifold: This is already introduced in defining manifold in Euclidean space. According to its definition, it is also called the “locally dimension” in this paper.

Minimum Euclidean space embedding dimension (MESED): “The minimum dimension of those Euclidean spaces in which a given manifold can be smoothly embedded with a one-to-one mapping” [42]. This dimension is also called the “global dimension” or “topological embedding dimension” in this paper.

We give an example in Fig. 2(a). A circle that is embedded in a 2-dimensional Euclidean space is a 1-dimensional manifold. Any given point of the circle (point x is chosen as example in Fig. 2(a)) has a neighborhood in the circle which is homeomorphic to 1-dimensional Euclidean space \mathbb{R} . According to the definition of manifold in Euclidean space, a circle is a 1-dimensional manifold. We could also say its “locally dimension” is 1. On the other hand, the entire circle cannot exist in a 1-dimensional Euclidean space. At least 2-dimensional Euclidean space is needed to embed the circle. This means its MESED or “topological embedding dimension” is 2.

A relation between dimension of manifold and MESED is given by the following theorem:

Strong Whitney embedding theorem [44]: Any connected smooth n -dimensional manifold S^n (required also to be Hausdorff and second-countable) can be smoothly embedded in the Euclidean $2n$ -space \mathbb{R}^{2n} .

2.1.2. Manifold hypothesis and a sufficient condition for the hypothesis

Scientists in many fields face the problem of simplifying high-dimensional data by finding low-dimensional structure in it. If there is a low-dimensional manifold structure hidden in the high-dimensional data, the hidden manifold can be recovered by manifold learning techniques. Fig. 2(b) is an example of a set of image data with a hidden manifold structure. The structure is successfully recovered by LLE as shown in Fig. 2(c).

In favor to reduce the dimension of high-dimensional data, some researchers conduct manifold learning techniques in certain data. However, they fail to find the hidden manifold structure and claim the manifold learning techniques are weak. Actually, they omitted the fact that their testing samples might have a “hidden structure” but do not have a “hidden manifold structure”. According to the definition of manifold in Section 2.1.1, we introduce a sufficient condition to ensure that data samples have a hidden manifold structure. It is divided into 4 constraints:

Constraint 1—continuous mapping: For data samples with a previously known underlying manifold structure, neighborhoods in data space are also neighborhoods in the underlying manifold.

Continuous mapping is an important prerequisite that requires data sampling to preserve neighborhood relations. If the underlying manifold structure is previously unknown, this restriction is automatically satisfied because manifold learning algorithms find only continuous mappings from data to underlying structure.

Constraint 2—locally homeomorphic: For data samples with a previously known underlying manifold S^k , dimension of manifold derived from dimension-determining techniques [20,21] should also be k . For data samples without a previously known

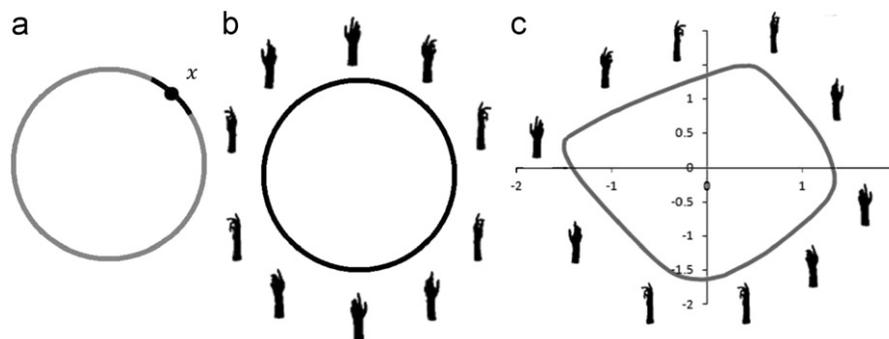


Fig. 2. (a) A circle is used as an example of 1-dimensional manifold with topological embedding dimension as 2. Given any point in the circle (x as an example), its neighborhood (denoted in black) is homeomorphic to 1-dimensional Euclidean space. At least 2-dimensional Euclidean space is needed to embed the circle. (b) Example of image data with an underlying circle structure. 120 silhouettes are sampled by rotating the Hand model m324 from PSB [49]. 10 images are presented here. (c) Manifold structure found by LLE. 120 images are used and 10 are presented.

underlying manifold, the dimension of their underlying manifold should be robustly derived from a dimension determining technique.

The definition of a k -dimensional manifold requires neighborhood of any given data samples to be homeomorphic to local Euclidean subspace \mathbb{R}^k .

For data samples with a known underlying manifold, if the sampling is too sparse or too noisy, the manifold structure might not be recovered. We use dimension-determining techniques [20,21] to evaluate the samples. Such techniques find the locally dimension according to the locally homeomorphic. Dimension returned from these techniques should agree with the previously known dimension.

In most cases, manifold learning techniques are aiming to recover the underlying manifold structure that is previously unknown. If the data sampling is too sparse or noisy, or there isn't any underlying manifold structure at all, it is not expected to recover a manifold structure from manifold learning techniques. Dimension-determining techniques are used as a prerequisite, if the locally dimension cannot be found, there is no possibility to find the global manifold structure. Handwritten digits [2], for example, do not really have an underlying manifold structure, they only satisfy the continuous constraint.

We mention that the use of dimension-determining algorithms does not nullify the use of manifold learning techniques: local dimension determining algorithms [20,21] find only the "local dimension", while manifold learning techniques find the global structure.

Constraint 3—topological space, Hausdorff and second countability: "Data samples are in Euclidean space".

Samples should meet the constraints defined for topology, the Hausdorff property and second countability. In practice, we do not need to show how Euclidean space is used universally. Further, we can only compute finite samples. Thus, the topology, Hausdorff property and second countability constraints are automatically satisfied.

Constraint 4—connectivity: The underlying manifold is simply connected.

We do not discuss manifolds with more than one connected component. If there are more than one connected components, each of them can be analyzed separately. Moreover, there is no need to use manifold learning algorithms to find connected components. There are many robust graph-theory-based algorithms [40,41].

In general, constraints 2 and 4 are the crucial constraints to verify. And we also see that if the data satisfies these constraints, it also satisfies the strong Whitney embedding theorem.

2.2. Locally linear embedding and the "intrinsic dimension" problem

2.2.1. Locally linear embedding

The basic idea of locally linear embedding algorithm is illustrated in Fig. 3. To avoid ambiguity, we introduce the algorithm in matrix form in Section 2.2.1. After that, we show the "intrinsic dimension" problem by an example in Section 2.2.2.

Step 1: Neighborhood search.

Each D -dimensional input data sample is denoted as a horizontal vector:

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD}) \quad (2.1)$$

All N inputs can be represented as a matrix:

$$\mathbb{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix}_{N \times D} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1D} \\ x_{21} & x_{22} & \dots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{ND} \end{bmatrix}_{N \times D} = [\mathbf{X}_1 \ \mathbf{X}_2 \ \dots \ \mathbf{X}_D]_{N \times D} \quad (2.2)$$

For each vector \mathbf{x}_i , we identify its neighborhoods by finding the nearest K vectors using Euclidean distance. And its neighbors are denoted by an N -dimensional horizontal vector:

$$\mathbf{e}_i = (e_{i1}, e_{i2}, \dots, e_{iN}) \quad (2.3)$$

where $e_{ij} = 1$ if \mathbf{x}_j is a neighbor of \mathbf{x}_i , $e_{ij} = 0$ if \mathbf{x}_j is not and $e_{ij} = 0$ if $i=j$.

The overall neighborhood relationship is represented as an $N \times N$ matrix:

$$\mathbb{E} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}_{N \times N} = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1N} \\ e_{21} & e_{22} & \dots & e_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N1} & e_{N2} & \dots & e_{NN} \end{bmatrix}_{N \times N} \quad (2.4)$$

Step 2: Constrained least square fits.

In this step, each data sample is linearly represented by its neighbors. For a particular vector \mathbf{x}_i , its K neighbors are temporarily denoted as $\mathbf{x}_{e_1}, \mathbf{x}_{e_2}, \dots, \mathbf{x}_{e_K}$. A locally linear representation is given by finding a set of reconstruction coefficients $w_{e_1}, w_{e_2}, \dots, w_{e_K}$ that minimize the representation error:

$$\varepsilon = \left| \mathbf{x}_i - \sum_{j=1}^K w_{e_j} \mathbf{x}_{e_j} \right| \quad (2.5)$$

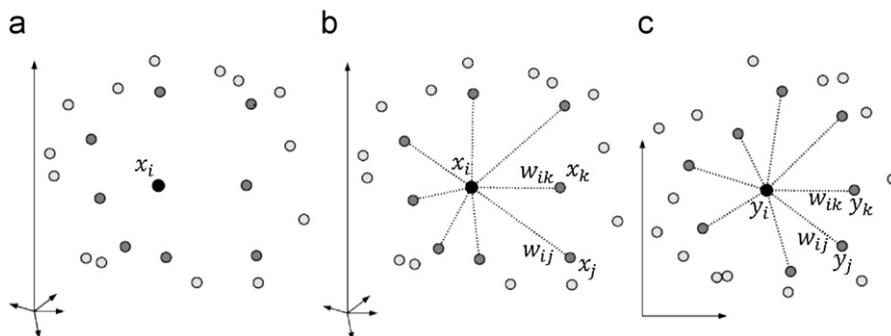


Fig. 3. Steps of locally linear embedding: (a) assign neighbors to each data point \mathbf{x}_i (for example using the K nearest neighbors). (b) Compute the weights w_{ij} that best linearly reconstruct \mathbf{x}_i from its neighbors by solving the constrained least-squares problem in formula (2.5). (c) Compute the low-dimensional embedding vectors \mathbf{y}_i best reconstructed by w_{ij} , minimizing formula (2.13) by finding the smallest eigenmodes of the sparse symmetric matrix in formula (2.15). Although the weights w_{ij} and vectors \mathbf{y}_i are computed by methods in linear algebra, the constraint that points are only reconstructed from neighbors can result in highly nonlinear embeddings. Cited from [1].

with constraints:

$$\sum_{j=1}^K w_{ej} = 1, w_{ej} \geq 0 \quad (2.6)$$

Formula (2.5) can be written as

$$\varepsilon^2 = \left| \mathbf{x}_i - \sum_{j=1}^K w_{ej} \mathbf{x}_{ej} \right|^2 = \left| \sum_{j=1}^K w_{ej} (\mathbf{x}_i - \mathbf{x}_{ej}) \right|^2 = \sum_{jk} w_{ej} w_{ek} \mathbb{G}_{jk} \quad (2.7)$$

where \mathbb{G}_{jk} is the local Gram matrix:

$$\mathbb{G}_{jk} = (\mathbf{x}_i - \mathbf{x}_{ej})^T (\mathbf{x}_i - \mathbf{x}_{ek}) \quad (2.8)$$

Reconstruction weights are solved by the Lagrange multiplier method:

$$w_{ej} = \frac{\sum_k \mathbb{G}_{jk}^{-1}}{\sum_{lm} \mathbb{G}_{lm}^{-1}} \quad (2.9)$$

To make (2.9) uniquely solvable, \mathbb{G}_{jk} is modified as

$$\mathbb{G}_{jk} \leftarrow \mathbb{G}_{jk} + \delta_{jk} \left(\frac{\Delta^2}{K} \right) \text{Tr}(\mathbb{G}) \quad (2.10)$$

where δ_{jk} is 1 if $j=k$ and 0 otherwise. $\text{Tr}(\mathbb{G})$ denotes the trace of \mathbb{G} and $\Delta^2 \ll 1$. Usually, $\Delta=0.1$.

The overall representation weights are arranged in a representation weight matrix:

$$\mathbb{W} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_N \end{bmatrix}_{N \times N} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \dots & w_{NN} \end{bmatrix}_{N \times N} \quad (2.11)$$

where w_{ij} is the weight of \mathbf{x}_i from \mathbf{x}_j . If \mathbf{x}_i is not a neighbor of \mathbf{x}_j , $w_{ij}=0$. Notice that $w_{ij}=0$ does not guarantee \mathbf{x}_j is not a neighbor of \mathbf{x}_i ; the neighborhood relationship is denoted separately by \mathbb{E} .

Step 3: Eigenvalue problem.

The final step finds another set of d -dimensional vectors $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ that best fit the representation weights \mathbb{W} . $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ is said to be the d -dimensional embedding of the original data. $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ is arranged in the matrix form as follows:

$$\mathbb{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}_{N \times d} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1d} \\ y_{21} & y_{22} & \dots & y_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \dots & y_{Nd} \end{bmatrix}_{N \times d} = [\mathbf{Y}_1 \ \mathbf{Y}_2 \ \dots \ \mathbf{Y}_d]_{N \times d} \quad (2.12)$$

\mathbb{Y} minimizes the square average cost function Φ :

$$\Phi(\mathbb{Y}) = \|\mathbb{Y} - \mathbb{W}\mathbb{Y}\|^2 \quad (2.13)$$

or

$$\begin{aligned} \Phi(\mathbb{Y}) &= \|\mathbb{Y} - \mathbb{W}\mathbb{Y}\|^2 = \sum_{k=1}^d |\mathbf{Y}_k - \mathbb{W}\mathbf{Y}_k|^2 = \sum_{k=1}^d (\mathbf{Y}_k - \mathbb{W}\mathbf{Y}_k)^T (\mathbf{Y}_k - \mathbb{W}\mathbf{Y}_k) \\ &= \sum_{k=1}^d \mathbf{Y}_k^T (I - \mathbb{W})^T (I - \mathbb{W}) \mathbf{Y}_k = \sum_{k=1}^d \mathbf{Y}_k^T \mathbb{M} \mathbf{Y}_k \end{aligned} \quad (2.14)$$

where

$$\mathbb{M}_{N \times N} = (I - \mathbb{W}_{N \times N})^T (I - \mathbb{W}_{N \times N}) \quad (2.15)$$

and $\{\mathbf{Y}_i\}_{i=1,2,\dots,d}$ is orthogonal and regulated:

$$\mathbf{Y}_i^T \mathbf{Y}_j = 0, i \neq j, i, j = 1, 2, \dots, d \quad (2.16)$$

$$E(|\mathbf{Y}_i|^2) = \frac{1}{N} \mathbf{Y}_i^T \mathbf{Y}_i = 1, \quad i = 1, 2, \dots, d \quad (2.17)$$

The optimal solution to (2.14) with constraints (2.16) and (2.17) is calculated using the Lagrange multiplier method:

$$\Phi(\mathbb{Y}) = \sum_{k=1}^d \mathbf{Y}_k^T \mathbb{M} \mathbf{Y}_k - \sum_{k=1}^d \lambda_k \left(\frac{1}{N} \mathbf{Y}_k^T \mathbf{Y}_k - 1 \right) \quad (2.18)$$

Differentiate (2.18) by $\mathbf{Y}_k, k=1, 2, \dots, d$:

$$\Rightarrow \mathbf{Y}_k^T \mathbb{M} - \lambda_k \mathbf{Y}_k^T = 0 \Rightarrow \mathbb{M} \mathbf{Y}_k = \lambda_k \mathbf{Y}_k \quad (2.19)$$

As a result, $\{\lambda_k\}$ are eigenvalues of \mathbb{M} , while $\{\mathbf{Y}_k\}$ are the corresponding eigenvectors.

From (2.14):

$$\Phi(\mathbb{Y}) = \sum_{k=1}^d \mathbf{Y}_k^T \mathbb{M} \mathbf{Y}_k = \sum_{k=1}^d \mathbf{Y}_k^T \lambda_k \mathbf{Y}_k = \sum_{k=1}^d \lambda_k \mathbf{Y}_k^T \mathbf{Y}_k = \frac{1}{N} \sum_{k=1}^d \lambda_k \quad (2.20)$$

Therefore, for each given d , $\Phi(\mathbb{Y})$ can be minimized by finding the bottom eigenvalues of \mathbb{M} and their corresponding eigenvectors.

Obviously, eigenvector $\mathbf{Y}_0^T = \pm(1, 1, \dots, 1)$ have the eigenvalue $\lambda_0=0$; thus, it should be discarded because they reveal nothing about the original data.

To solve for each given d , we find the lowest $d+1$ eigenvalues of \mathbb{M} and their corresponding $d+1$ eigenvectors, and discard $\mathbf{Y}_0^T = \pm(1, 1, \dots, 1)$; the remaining d eigenvectors are the best d -dimensional embedding of the original data.

2.2.2. An example of the “intrinsic dimension” problem

The existing LLE algorithm ends without giving the method to determine the “intrinsic dimension” d , which is crucial for the manifold structure. We use the data of hand rotation (Fig. 2(b)) to show this problem. We visualize the manifold structure in Fig. 4 for $d=1, d=2, d=3$ and $d=4$. They look distinctively. Therefore, if the “intrinsic dimension” is unknown, we cannot determine

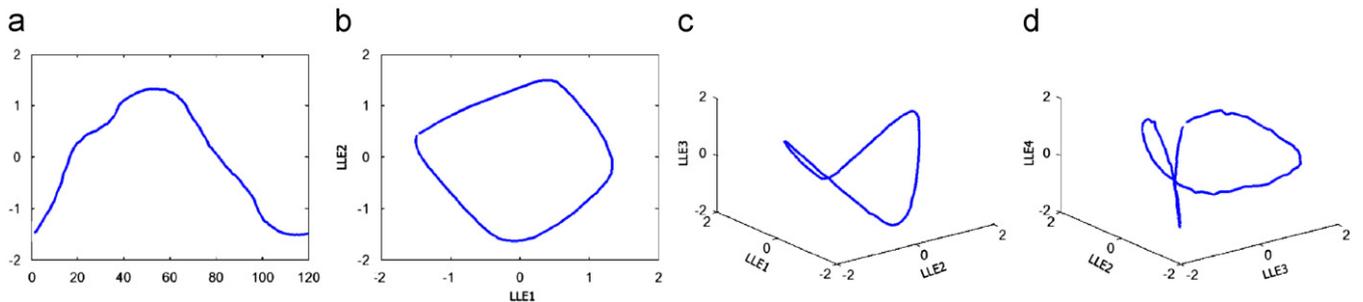


Fig. 4. Reconstruct the underlying manifold structure of hand rotating images (Fig. 2(b)). Assuming different “intrinsic dimension” d , different structures are reconstructed. (a) $d=1$, the indices of images are used as the x -axis. (b) $d=2$. (c) $d=3$. (d) $d=4$, only the 2nd, 3rd and 4th LLE eigenvectors are visualized.

which structure is right. Other examples of this problem can also be found in [2].

3. Mathematical analyses of topological manifolds and the LLE algorithm

In this section we mathematically analyze topological manifolds and LLE algorithm to develop the key ideas for the manifold topological analysis method (MTA) and manifold multi-resolution analysis method (MMA). Specifically, the “centralization tradeoff” introduced in Section 3.2 is a key idea for MMA, and “spectral decomposition” introduced in Section 3.4 and “topological embedding dimension” introduced in Section 3.5 are key ideas for both MTA and MMA, respectively. We develop these ideas step by step with interim analysis in Sections 3.1 and 3.3.

3.1. Locally linear spaces

From the definition of a d -dimensional manifold, each data sample has locally linear spaces. As illustrated in Fig. 5, we introduce the following locally linear spaces of a particular sample \mathbf{x}_i with its neighbors $\mathbf{x}_{e1}, \mathbf{x}_{e2}, \dots, \mathbf{x}_{eK}$:

- The original data space: \mathbb{R}^D .
- The local tangential space of the manifold: \mathbb{R}^d .
- The span space by neighborhoods $\mathbf{x}_{e1}, \mathbf{x}_{e2}, \dots, \mathbf{x}_{eK}$: \mathbb{R}^S .

\mathbb{R}^d is derived directly from the manifold hypothesis, and \mathbf{x}_i is sampled from the d -dimensional manifold; it has a local neighborhood homeomorphic to \mathbb{R}^d and is tangent to the manifold. Because \mathbf{x}_i is sampled with noise, it might not be located exactly on the manifold. We give a tolerance such that the space is spanned by its nearest point on the manifold.

\mathbb{R}^S has the expression:

$$\mathbb{R}^S = \left\{ \mathbf{x} \mid \sum_{j=1}^K w_{ej} \mathbf{x}_{ej}, \sum_{j=1}^K w_{ej} = 1 \right\} \quad (3.1)$$

Constraint $\sum_{j=1}^K w_{ej} = 1$ is important because it makes the spanned space local.

We discuss the dimension S of this space. From expression (3.1), S is no greater than $K-1$:

$$S \leq K-1 \quad (3.2)$$

Because locally linear representation is expected to capture the manifold tangential space \mathbb{R}^d , it is required that S be no smaller than d :

$$S \geq d \quad (3.3)$$

In ideal situations, $\{\mathbf{x}_{ej}\}$ should all lie in the tangential space \mathbb{R}^d , as required by definition of manifold. Thus, S should be no

greater than d . However, in practices (1) data are sampled discretely, which means data might be located outside \mathbb{R}^d because of manifold nonlinearity; (2) data are sampled with noise, which also means that the data might be located outside \mathbb{R}^d . Formula (3.3) is satisfied if the number of neighbors K is chosen to be big enough.

Generally, the three spaces are related according to

$$\mathbb{R}^d \subseteq \mathbb{R}^S \subseteq \mathbb{R}^D \quad (3.4)$$

3.2. “Degeneration” and “centralization tradeoff”

“Degeneration” is originally discussed by Saul and Roweis [2] for the LLE algorithm to uniquely solve the representation weights defined in (2.10). From (3.1) and (3.2), such degeneration occurs when $S < K-1$. From (2.7), a weaker condition to claim “degeneration” is

$$D < K-1 \quad (3.5)$$

Formula (3.5) is more practical because D and K are known.

$\delta_{jk}(\Delta^2/K)\text{Tr}(\mathbb{G})$ in (2.10) punishes big representation weights. Equivalently, formula (2.10) chooses those weights which $\{w_{ej}\}$ minimize the square sum:

$$\sum_{j=1}^K w_{ej}^2 \quad (3.6)$$

Intuitively, this is to place $\mathbf{x}_i = \sum_{j=1}^K w_{ej} \mathbf{x}_{ej}$ at the center of its neighborhoods $\{\mathbf{x}_{ej}\}$. This is called the “centralization optimal”.

In this section, we stress the tradeoff between two optimal: “least square fit” (2.7) and “centralization” (3.6). It is a key preposition to support the MMA method. In practice, data are discretely sampled from nonlinear manifold. Given a particular point, it is not expected to lie in the center of its neighborhoods. To minimize the square fit error (2.7), square sum of weights (3.6) is not minimized. Formula (2.10) gives a tradeoff between these two optimals, where Δ denotes how important “centralization” is in compared with “least square fit”. $\Delta=0.1$ by default in the LLE algorithm [2]. Intuitively, “least square fit” means “pull $\mathbf{x}_i = \sum_{j=1}^K w_{ej} \mathbf{x}_{ej}$ away from center of $\{\mathbf{x}_{ej}\}$ ”.

This tradeoff gives LLE the ability to capture the nonlinearity of manifold. We use it to prove the MMA method in Section 5.

3.3. Continuous mapping

In addition to embedding, we demonstrate that the eigenvectors $\{\mathbf{Y}_k\}$ (2.12) and (2.19) can be seen as mappings from data samples to real numbers \mathbb{R} :

$$\mathbf{Y}_k : \{\mathbf{x}_i\} \rightarrow \mathbb{R} \quad (3.7)$$

Given the definition of continuous mapping in Section 2.1, $\{\mathbf{Y}_k\}$ can be considered to be continuous maps because LLE maintains neighborhoods.

According to the manifold hypothesis, $\{\mathbf{x}_i\}$ are sampled from the underlying manifold \mathbb{S} . This means $\{\mathbf{Y}_k\}$ are continuous maps from the underlying manifold \mathbb{S} to real numbers \mathbb{R} .

$$\mathbf{Y}_k : \mathbb{S} \rightarrow \mathbb{R} \quad (3.8)$$

Technically, mapping newly sampled data could be accomplished by well-developed function interpolations techniques. Some research have attempted to modify LLE from embedding to mapping [38,52]. In our research, we do not discuss the techniques in detail but stress the idea that $\{\mathbf{Y}_k\}$ are continuous mappings from the underlying manifold \mathbb{S} to the real numbers \mathbb{R} .

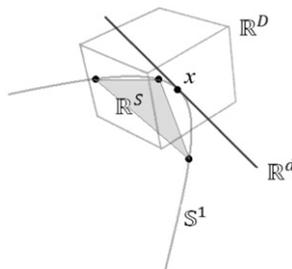


Fig. 5. Illustration of locally linear spaces. 1-dimensional manifold \mathbb{S}^1 embedded in \mathbb{R}^3 is used as an example. For a particular point \mathbf{x} , \mathbb{R}^d is its 1-dimensional manifold tangential space; \mathbb{R}^S is the 2-dimensional space spanned by neighbors; and \mathbb{R}^D is the 3-dimensional data sample space.

3.4. Spectral decomposition

Other than embedding, it is rewarding to consider the third step of LLE as spectral analysis of the original data and their underlying manifold. All the eigenvalues of $\mathbb{M}_{N \times N}$ (2.15) can be sorted such that

$$\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d \leq \dots \leq \lambda_N \quad (3.9)$$

and their corresponding eigenvectors are related according to

$$\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_d, \dots, \mathbf{Y}_N \quad (3.10)$$

where λ_i measures prominence of \mathbf{Y}_i and \mathbf{Y}_i as a feature of the original data [40].

This idea is similar to PCA, in which the original data are analyzed by spectral decomposition; the highest eigenvalues and eigenvectors of PCA define the expansion while the lowest eigenvalues and eigenvectors define the clustering. Some research has used LLE as a spectral analysis tool [12,15,53] based on this idea.

However, we found that the analogy to PCA does not hold completely because LLE-eigenvalues and LLE-eigenvectors have their own meanings: *Each eigenvector of LLE is a continuous mapping from the manifold to real number \mathbb{R} and captures nonlinearities from the manifold; its corresponding eigenvalue measures the prominence of captured nonlinearity.*

Our MMA method is developed and proved in Section 5. Nonlinearity capture is introduced and proved in detail in Section 5.3.

3.5. Topological embedding

It is mentioned in Section 3.3 that each LLE eigenvectors is a continuous map from manifold to \mathbb{R} . Because continuous map preserves topology, we can find the low-dimensional topological manifold structure from high-dimensional data using the right LLE eigenvectors. This is the basic idea of LLE.

For manifold with MESED=2, it can be visualized using two eigenvectors (denoted as \mathbf{Y}_i and \mathbf{Y}_j). The coordinates \mathbf{y}_n of sample \mathbf{x}_n is

$$\mathbf{y}_n = (y_{ni}, y_{nj}) = (\mathbf{Y}_i(\mathbf{x}_n), \mathbf{Y}_j(\mathbf{x}_n)) \quad (3.11)$$

And the manifold \mathbb{S} can be expressed as

$$\mathbb{S} = (\mathbf{Y}_i(\mathbb{S}), \mathbf{Y}_j(\mathbb{S})) \quad (3.11')$$

Similarly, manifold with MESED=3 can be visualized as

$$\mathbb{S} = (\mathbf{Y}_i(\mathbb{S}), \mathbf{Y}_j(\mathbb{S}), \mathbf{Y}_k(\mathbb{S})) \quad (3.12)$$

Formula (3.11)' can also be considered as an implicit expression of $\mathbf{Y}_j(\mathbb{S})$ in \mathbb{R}^2 with the coordinates given by $\mathbf{Y}_i(\mathbb{S})$. And (3.12) can be considered as implicit expression of $\mathbf{Y}_j(\mathbb{S})$ in \mathbb{R}^3 with the coordinates given by $\mathbf{Y}_i(\mathbb{S})$ and $\mathbf{Y}_k(\mathbb{S})$. This implicit expression gives us a method to observe the behavior of all eigenvectors without using a known index.

As previously shown by the example in Section 2.2.2, we now discuss the well-documented criticism of LLE: it cannot determine the “intrinsic dimension” of the underlying manifold. Newer local manifold learning algorithms [6,7,9,10] are aiming to improve geometric performance, but they cannot determine the “intrinsic dimension” either.

Saul and Roweis [2] have attempted to determine the “intrinsic dimension” using eigenvalues, which is an analogy to global manifold learning algorithms. However, their result is negative. We point out it is a “false analogy”: as discussed in Sections 3.2 and 3.4, the second step of LLE captures only the local relationships in the manifold tangential space \mathbb{R}^d or the neighborhood

spanned space \mathbb{R}^S . This is different than PCA and Isomap, where the global relationship for each sample in \mathbb{R}^D has been captured.

Other than the “false analogy”, we find the meaning of “intrinsic dimension” is ambiguous. Different researchers have different understandings on “intrinsic dimension”. To avoid ambiguity, we have introduced the dimension of manifold and MESED in Section 2.1.1 instead. We clarify that Kégl [20] and Levina and Bickel [21] who propose local dimension determining methods [20,21] consider it as the dimension of manifold, while Tenenbaum et al. [4] and Fu and Thomas [16] who propose global manifold learning algorithms [4,16,46] consider it as the MESED.

We find a method to determine the MESED based on the particular behavior of LLE eigenvectors.

LLE first captures local information of manifold \mathbb{S}^d . Because the “locally dimension” (dimension of manifold) is d , the information captured is also d -dimensional. Then LLE tries to find global structures to fit all the local information. From the strong Whitney embedding theorem, the global structure can always be found in \mathbb{R}^T , where the MESED: $T \leq 2d$. Solving the eigenproblem (2.19), we can find N eigenvectors that are orthogonal (2.16), where N is the number of data samples. Analyzed in Section 3.3, those eigenvectors that preserve the continuity will preserve the topology. Since $d \ll N$, we have $T \leq 2d \ll N$, and most eigenvectors preserve the continuity in order to minimize the reconstruction error (2.14). Therefore, the number of those eigenvectors preserving the continuity is much more than T . Since there is only T -dimensional freedom in topology, the eigenvectors will repeat the same topology while finding other ways to satisfy the orthogonal constraint (2.16).

In Section 5, we discuss and prove that the eigenvectors of LLE outside \mathbb{R}^T find freedom in general function spaces to satisfy the orthogonal regulation. Moreover, the general function spaces are specified as multi-resolutional spaces. In Section 4, we introduce a method to find eigenvectors inside \mathbb{R}^T by observing the eigenvectors' multi-resolution behavior, which is developed as the MTA method.

4. Manifold topological analysis (MTA) method

In this section, we introduce the manifold topological analysis (MTA) method for typical topological manifolds, which are most frequently used in computer vision, especially in tracking and perception understanding. We analyze the behavior of the eigenvectors and eigenvalues of these manifolds. In addition, we propose a method to determine their topological embedding dimension and a method to improve the geometric performance of LLE using eigenvalues.

4.1. Typical topological manifolds in computer vision

We introduce 4 typical topological manifolds:

(1) Line: \mathbb{R} . (2) Circle: \mathbb{S}^1 . (3) Plane: \mathbb{P} . (4) Cylinder: *cyl*.

There are only 2 kinds of 1-dimensional topology: a line \mathbb{R} and circle \mathbb{S}^1 . They can generate 2-dimensional topology using a Cartesian product. Thus, they are basic and important.

There are 3 basic 2-dimensional topologies: plane \mathbb{P} , sphere \mathbb{S}^2 and torus \mathbb{T} . However, we choose \mathbb{P} and *cyl* as typical topologies. The following are the reasons.

Plane \mathbb{P} is the most popular manifold in manifold learning algorithms. The rectangle, swiss-roll and S-shape plane are all planes in topology. Almost all manifold learning algorithms have experimented with plane \mathbb{P} [1–9,16,17]. In computer vision, there are many samples with an underlying topological plane. Plane \mathbb{P} is chosen as one typical topological manifold.

The sphere \mathbb{S}^2 is a very important 2-dimensional topology. For example, the viewpoint space of a 3D-object is a sphere, as shown in Fig. 12(a). However, it is seldom discussed in manifold learning methods because it is impossible to visualize a dimension reduction of a sphere. To show a dimension reduction of sphere, the original space must be at least 4-dimensional. The other reason we do not choose a sphere is: a sphere is difficult to describe with an explicit expression in Euclidean space. For example, a unit sphere is described explicitly in polar coordinates by $R=1$ or implicitly in Euclidean space by $x^2+y^2+z^2=1$. The explicit expression in Euclidean space can be derived from the polar expression (R,θ,φ) . However, the expression varies dramatically depending on the choice of the original point (θ,φ) . The manifold topological analysis method would be very complex to apply to a sphere. However, a sphere can be analyzed by the manifold multi-resolution analysis method (MTA) introduced in Section 5. An example of a viewpoint sphere analysis is given in Section 6.2.

The cylinder *cyl* is not a basic 2-dimensional topology, which can be seen as a combination of two planes [44]. However, it is very important because it is much easier to analyze than a sphere. In 3D computer vision, it is often used instead of a sphere [56]. Thus, we choose a cylinder as a typical 2-dimensional topology.

The torus \mathbb{T} is not chosen because it seldom occurs in computer vision.

Classification of 3 or higher dimensional topology has not been completely developed in mathematics [44]. And all 3 or higher dimensional topology cannot be visualized, except the cube \mathbb{R}^3 . Therefore, they are not discussed in MTA. We introduce and prove our MMA method in Section 5, which is a feature extraction method that is applicable to all topologies.

We introduce and prove the MTA method in the rest of this section.

4.2. Manifold topological analysis method

We introduce the MTA method for typical topological manifolds:

Step 1: Determine whether or not the data samples satisfy the “manifold hypothesis” introduced in Section 2.1.2. If yes, calculate the “local dimension” d of their underlying manifold \mathbb{S} with a local dimension determining method [20,21] if d is previously unknown.

Step 2: Conduct LLE on data samples and return more than $4d$ of the lowest eigenvectors $\{\mathbf{Y}_i\}$ along with their corresponding eigenvalues $\{\lambda_i\}$ (including the zero eigenvalue and the eigenvector $\lambda_0, \mathbf{Y}_0^T = \pm(1, 1, \dots, 1)$).

Step 3: Observe the behavior of eigenvectors according to coordinates known a priori $(\mathbf{Y}_i : \mathbb{S} \rightarrow \mathbb{R})$ or implicitly $((\mathbf{Y}_i(\mathbb{S}), \mathbf{Y}_j(\mathbb{S}))$ (3.11) for a 1-dimensional manifold ($d=1$) and $(\mathbf{Y}_i(\mathbb{S}), \mathbf{Y}_j(\mathbb{S}), \mathbf{Y}_k(\mathbb{S}))$ (3.12) for a 2-dimensional manifold ($d=2$)). Calculate their degree of periodicity.

Step 4: Classify eigenvectors into groups by their degree of periodicity and find the basic (least periodic) eigenvectors. The number of basic eigenvectors is equal to the dimension of the topological embedding \mathcal{T} , and their implicit expression reconstructs the topological structure of the manifold.

Step 5: Rescale basic eigenvectors $\mathbf{Y}_i(\mathbb{S})$ to $c_i \mathbf{Y}_i(\mathbb{S})$, where c_i is determined by its corresponding eigenvalues $c_i/c_j = \sqrt{\lambda_j/\lambda_i}$. The implicit expression of rescaled basic eigenvectors reconstructs the topological structure of the manifold, while allowing the global geometry to be recovered.

Step 1 was discussed in Section 2.1.2. Step 2 was previously discussed in Section 3.4 and why $4d$ eigenvectors are chosen will be discussed in Section 4.5. Step 3 will be introduced and proved by enumeration in Section 4.3. The proof of Step 4 was discussed in Section 3.5 and will be expanded in Section 4.4. Step 5 will be introduced and proved in Section 4.5.

4.3. MTA on typical topological manifolds

In this section, we observe the behavior of eigenvectors of all typical topological manifolds, and categorize them mathematically. In MTA, only $4d$ eigenvectors are needed, and this will be explained in Section 4.5. In order to bring more information to the readers, we will give more than $4d$ eigenvectors in examples.

4.3.1. Line \mathbb{R}

We start from an ideal situation in which the data are uniformly sampled from a straight line segment. This is illustrated in Fig. 6(a). Fig. 6(b) and (c) shows the first 6 eigenvectors when 80 points are sampled uniformly; the neighborhood number is chosen as $K=2$. Fig. 6(b) is the explicit expression $\mathbf{Y}_i : \{j\} \rightarrow \mathbb{R}$, where j is the sample index. Fig. 6(c) is the implicit expression: $(\mathbf{Y}_1(\{\mathbf{x}_i\}), \mathbf{Y}_j(\{\mathbf{x}_i\}))$ (3.11). Significantly, Fig. 6(b) shows a set of sinusoidal functions in ascending order. It can be proved that they are standard sinusoidal functions:

$$\left\{ \pm \sin\left(n \frac{\pi x}{2l}\right) \right\} \quad (4.1)$$

where l is the length of the line segment $[a,b]$, $x \in [a,b]$ and $n=0,1,2,\dots$. For $K=2$, the reconstruction matrix is given as

$$\mathbb{W} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & \dots & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0.5 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{N \times N} \quad (4.2)$$

We will prove (4.1) in Section 4.3.2. It can also be proved that the implicit expressions shown in Fig. 6(c) are Chebyshev polynomials:

$$T_n(t) = \cos(\arccos(nt)), t \in [-1, 1], n = 0, 1, 2, \dots \quad (4.3)$$

We highlight the standard sinusoidal function bases and Chebyshev polynomial bases because there are two important spectral analysis methods with these two bases: Fourier series decomposition and Chebyshev decomposition. As a result, we were inspired to develop the manifold multi-resolution analysis method in Section 5.

Fig. 6(d)–(i) shows many visual examples whose underlying manifold is a line \mathbb{R} . We found their eigenvectors are still a set of non-standard sinusoidal functions with an ascending degree of sinusousness. Their degree of sinusousness can be measured as the number of local extremas or the number of points that cross the line $\mathbf{Y}_i(\mathbb{R})=0$. As in formula (4.1), we denote these typical eigenvectors of line \mathbb{R} as

$$\mathbf{Y}_0^{\mathbb{R}}, \mathbf{Y}_1^{\mathbb{R}}, \mathbf{Y}_2^{\mathbb{R}}, \mathbf{Y}_3^{\mathbb{R}}, \dots \quad (4.4)$$

4.3.2. Circle \mathbb{S}^1

We start from an ideal situation in which the data are uniformly sampled from a unit circle, Fig. 7(a). We use a visual example, which is shown in Fig. 7(b). For $K=2$ its locally linear representation weights matrix is

$$\mathbb{W} = \begin{bmatrix} 0 & 0.5 & 0 & \dots & 0 & 0.5 \\ 0.5 & 0 & 0.5 & \dots & 0 & 0 \\ 0 & 0.5 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0.5 \\ 0.5 & 0 & 0 & \dots & 0.5 & 0 \end{bmatrix}_{N \times N} \quad (4.5)$$

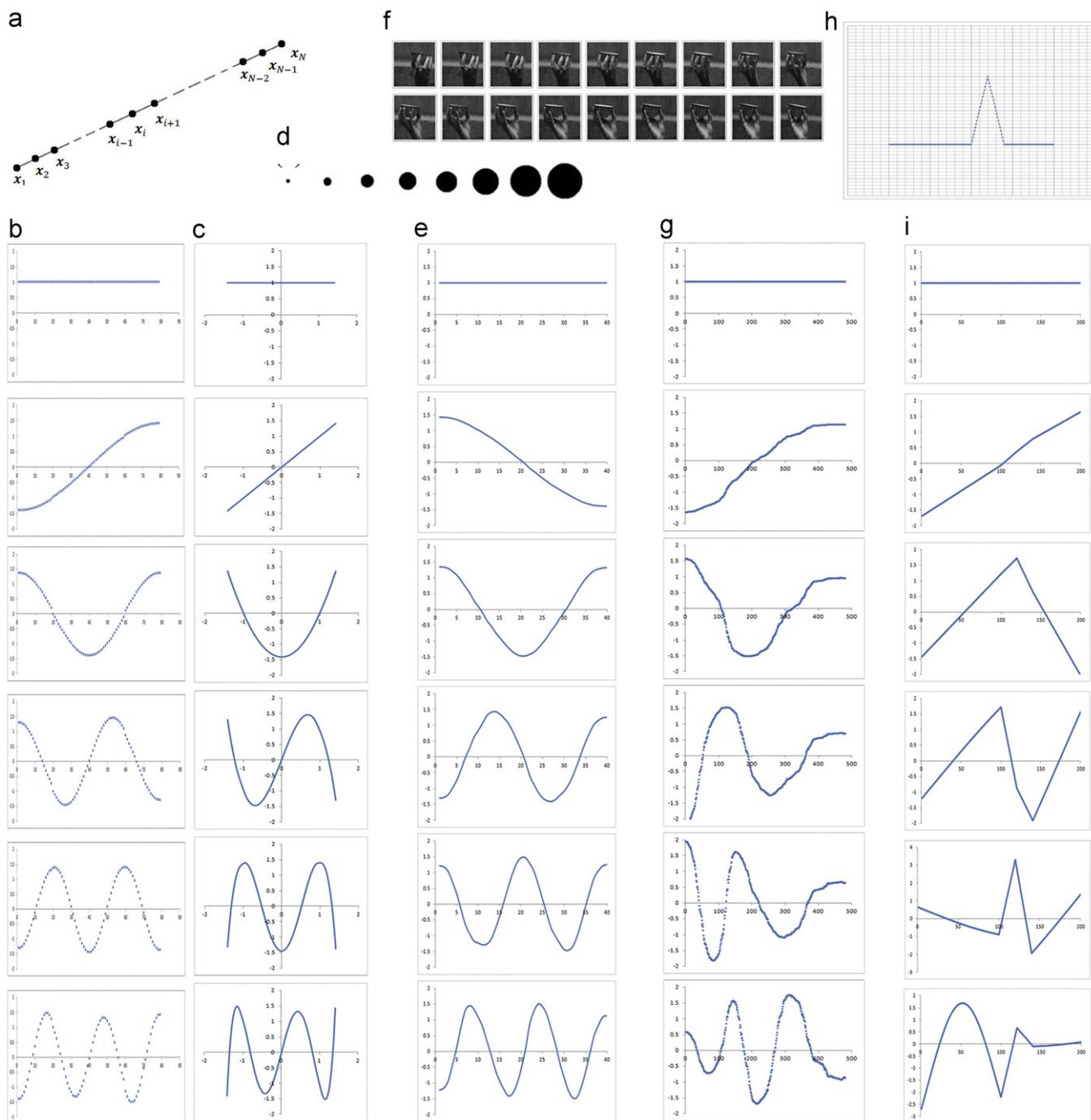


Fig. 6. LLE eigenvectors of line \mathbb{R} : (a) uniform samples from a straight line segment; (b) their lowest 6 eigenvectors in the explicit expression; (c) their lowest 6 eigenvectors in the implicit expression; (d) and (e) an image series of an expanding black circle and its lowest 6 eigenvectors in the explicit expression; (f) a rotating hand (481 images) from the CMU image database; (g) their lowest 6 eigenvectors in the explicit expression; (h) a nonlinear line segment in \mathbb{R}^2 ; and (i) its lowest 6 eigenvectors in the explicit expression.

Fig. 7(c) is the explicit expression $\mathbf{Y}_i : \{j\} \rightarrow \mathbb{R}$, where j is the sample index. Fig. 7(d) is the implicit expression $\mathbf{Y}_1(\{\mathbf{x}_i\}) \rightarrow \mathbf{Y}_j(\{\mathbf{x}_i\})$. Comparing with line \mathbb{R} situation, Fig. 7(c) also shows a set of sinusoidal functions in ascending order. The difference is that the eigenvectors of circle \mathbb{S}^1 , $\mathbf{Y}_i : \{\mathbb{S}^1\} \rightarrow \mathbb{R}$, are periodic functions, because the route on circle \mathbb{S}^1 is periodic. Eigenvectors are expected to be standard sine and cosine functions because of the ideal symmetry of \mathbb{W} :

$$\mathbf{Y}_0^{\mathbb{S}^1} \equiv 1 \quad \text{or} \quad -1$$

$$\mathbf{Y}_{n+}^{\mathbb{S}^1} = \pm \sin(n\theta + \varphi_{n+})$$

$$\mathbf{Y}_{n-}^{\mathbb{S}^1} = \pm \sin(n\theta + \varphi_{n-})$$

$$\theta \in [0, 2\pi), \quad |\varphi_{n+} - \varphi_{n-}| = \frac{\pi}{2} \tag{4.6}$$

Formula (4.6) is also proved as a method to generate complete function bases in general function analysis [48]. Each sinuous level (or frequency) has 2 orthogonal random phase sine functions. We denote the one with the smaller eigenvalue as $\mathbf{Y}_{n+}^{\mathbb{S}^1}$ and the other as $\mathbf{Y}_{n-}^{\mathbb{S}^1}$. These sinuous functions are the base functions of Fourier series decomposition for periodic functions.

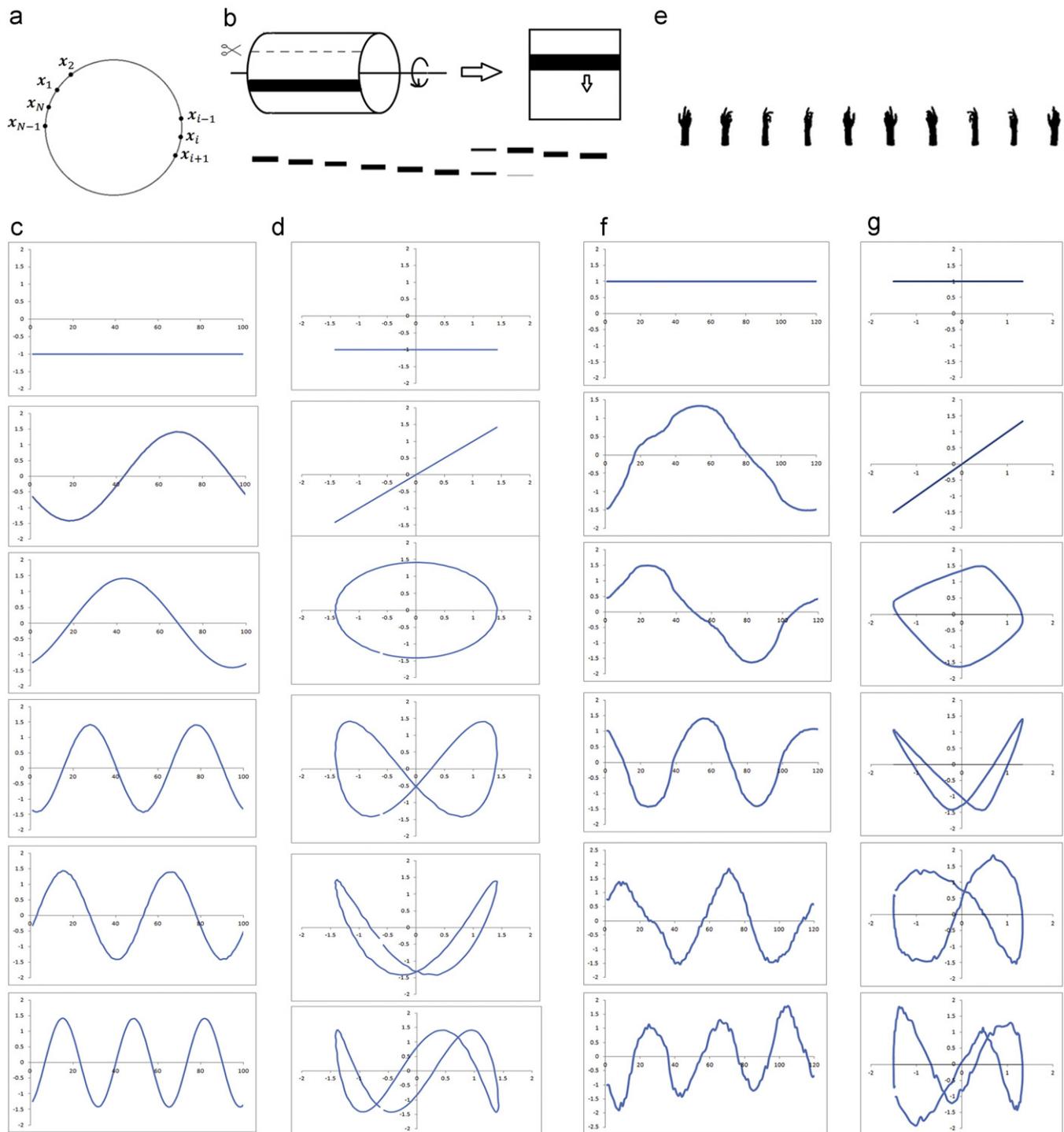


Fig. 7. LLE eigenvectors of circle S^1 : (a) uniform samples from a unit circle; (b) the method to generate a 100×100 pixel images series (100 images) with the underlying manifold structure shown in (a); (c) their lowest 6 eigenvectors in the explicit expression; (d) their lowest 6 eigenvectors in the implicit expression. (e) 120-images series sampled from the rotating hand model m324 from PSB [49]; (f) their lowest 6 eigenvectors in the explicit expression; and (g) their lowest 6 eigenvectors in the implicit expression.

We prove $\{Y_i^R\}$ in (4.1) are standard sine functions: Observe the weight matrixes (4.2) and (4.5). The only difference is the weights on the end points, x_1^R and x_N^R on line \mathbb{R} . Circle S^1 has no end points. If the circle is observed from the side, it collapses to a line. There is a perspective point at which x_{N-i}^S is sheltered from x_{2+i}^S , and x_1^S and $x_{N/2+1}^S$ become end points (N is an even number and $i=0,1,\dots,N/2-1$). Because x_{N-i}^S are sheltered, add the

weights of x_{N-i}^S to x_{2+i}^S , and remove x_{N-i}^S ; the new weight matrix is the same as (4.2). In the same way, deleting the sheltered part of $Y_{n\pm}^S$, we get the yield Y_n^R . Thus the theorem is proved. \square

In practice, the frequency feature of eigenvectors is very robust. Fig. 7(e)–(g) shows an example. Generally, eigenvectors of circle S^1 are denoted as

$$Y_0^S, Y_{n+}^S, Y_{n-}^S \quad n = 1, 2, \dots \tag{4.7}$$

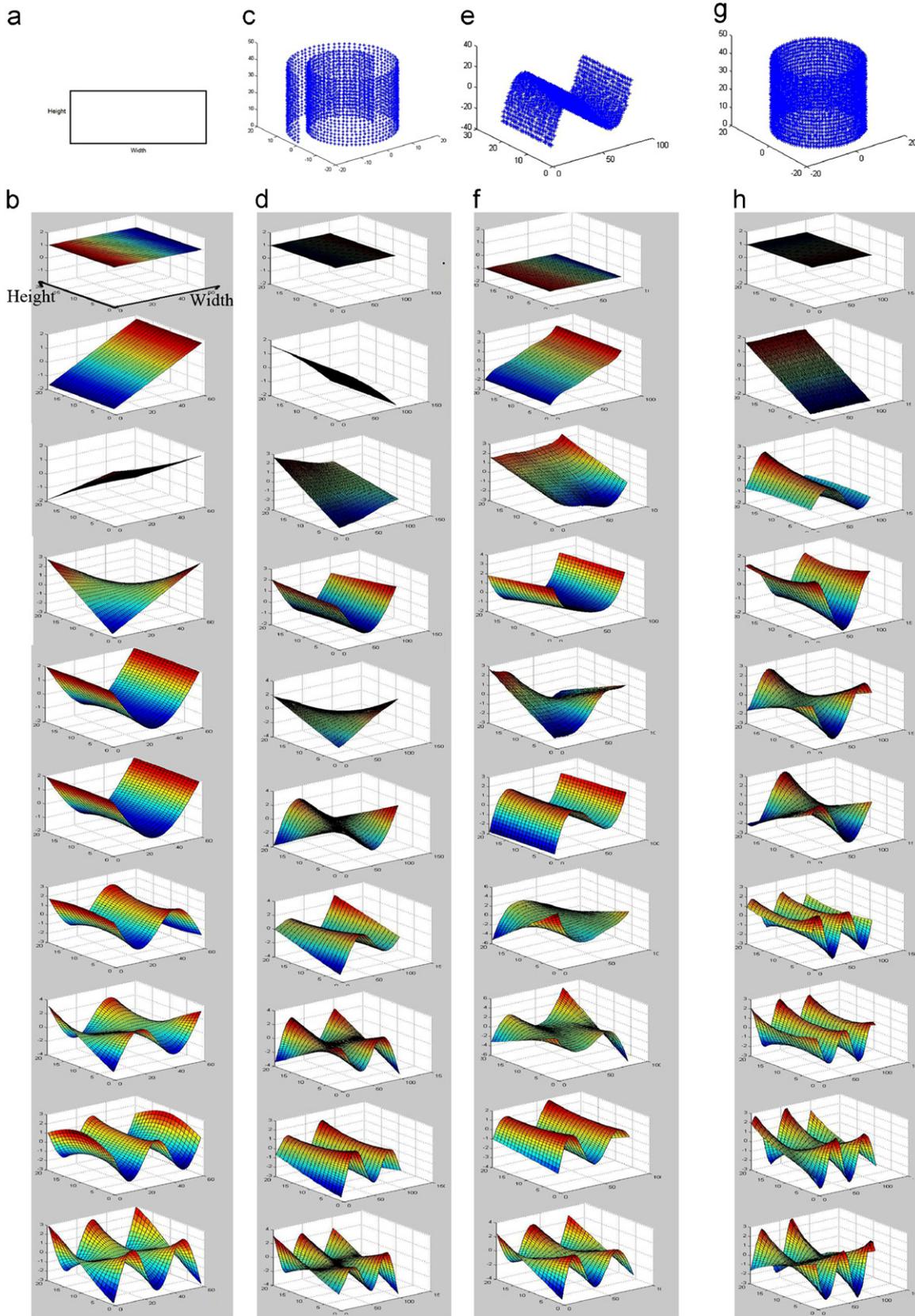


Fig. 8. (a)–(f) LLE eigenvectors of plane \mathbb{P} . (a) and (b) The rectangle and its lowest 10 eigenvectors in the explicit expression, respectively; (c) and (d) noise-sampled “Swiss-roll” and its eigenvectors, respectively; (e) and (f) noise-sampled “S-shape” and its eigenvectors, respectively; (g) and (h) LLE eigenvectors of cylinder *cyl*: a noise-sampled cylinder and its lowest 10 eigenvectors in the explicit expression, respectively.

Table 1
Eigenvectors of typical topological manifolds.

Topology	Manifold dimension	Topological embedding dimension	Eigenvectors	Basic eigenvectors
Line \mathbb{R}	1	1	$\mathbf{Y}_i^{\mathbb{R}} \quad i = 0, 1, 2, \dots$	$\mathbf{Y}_1^{\mathbb{R}}$
Circle \mathbb{S}^1	1	2	$\mathbf{Y}_i^{\mathbb{S}^1} \quad i = 0, 1 \pm, 2 \pm, \dots$	$\mathbf{Y}_{1+}^{\mathbb{S}^1}, \mathbf{Y}_{1-}^{\mathbb{S}^1}$
Plane \mathbb{P}	2	2	$\mathbf{Y}_{ij}^{\mathbb{P}} \quad i, j, k = 0, 1, 2, \dots$	$\mathbf{Y}_{0,1}^{\mathbb{P}}, \mathbf{Y}_{1,0}^{\mathbb{P}}$
Cylinder <i>cyl</i>	2	3	$\mathbf{Y}_{ij}^{cyl} \quad i = 0, 1, 2, \dots \quad j = 0, 1 \pm, 2 \pm, \dots$	$\mathbf{Y}_{1,0}^{cyl}, \mathbf{Y}_{0,1+}^{cyl}, \mathbf{Y}_{0,1-}^{cyl}$

4.3.3. Plane \mathbb{P}

The plane is easy to analyze because it can be generated as a Cartesian product of two lines:

$$\mathbb{P} = \mathbb{R}^2 = (\mathbb{R}, \mathbb{R}) \tag{4.8}$$

Similarly, eigenvectors of plane $\mathbf{Y}^{\mathbb{P}}$ are Cartesian products of line eigenvectors $\mathbf{Y}^{\mathbb{R}}$. Fig. 8(a) and (b) shows the eigen-decomposition of a rectangle. A rectangle has a “height orientation” and “width orientation”. We do not use a square because its orientation is arbitrary. The first 10 eigenvectors can be denoted as

$$\mathbf{Y}_{ij}^{\mathbb{P}} = (\mathbf{Y}_i^{\mathbb{R}}, \mathbf{Y}_j^{\mathbb{R}}) \tag{4.9}$$

$$(i, j) = (0, 0), (0, 1), (1, 0), (1, 1), (0, 2), (1, 2), (0, 3), (1, 3), (0, 4), (2, 5) \tag{4.10}$$

where i refers to the “height orientation” and j refers to the “width orientation”.

Such eigenvector behavior is robust against nonlinearity and noise. The “Swiss roll” and “S-shape plane” sampled with noise are popular; Fig. 8(c)–(f) shows their eigen-decomposition.

4.3.4. Cylinder *cyl*

Similar as plane, cylinder can be generated by Cartesian product of a line and a circle:

$$cyl = (\mathbb{R}, \mathbb{S}^1) \tag{4.11}$$

Moreover, its eigenvectors \mathbf{Y}_{ij}^{cyl} can be expressed as a Cartesian product of line-eigenvectors $\mathbf{Y}_i^{\mathbb{R}}$ and circle-eigenvectors $\mathbf{Y}_j^{\mathbb{S}^1}$:

$$\mathbf{Y}_{ij}^{cyl} = (\mathbf{Y}_i^{\mathbb{R}}, \mathbf{Y}_j^{\mathbb{S}^1}) \tag{4.12}$$

Fig. 8(g) and (h) shows the eigen-decomposition of a cylinder sampled with noise. The lowest 10 eigenvectors can be expressed as (4.12), where

$$(i, j) = (0, 0), (1, 0), (0, 1+), (0, 1-), (1, 1+), (1, 1-), (0, 2+), (0, 2-), (1, 2+), (1, 2-) \tag{4.13}$$

4.4. Determining the “topological embedding dimension” of typical topological manifolds

In Section 3.5, we have discussed that manifold with $MESED = \mathcal{T}$ can be embedded in $\mathbb{R}^{\mathcal{T}}$. $\mathbb{R}^{\mathcal{T}}$ can have at most \mathcal{T} orthogonal eigenvectors. The rest of the eigenvectors will repeat the topology and find orthogonal freedom by varying its frequency. For a typical topological manifold, we see its eigenvectors can be grouped by their topology. Eigenvectors in the same topology group varies in frequency to satisfy the orthogonal constraint (2.16). In each group, there is an eigenvector with the least frequency. It is labeled with the smallest non-zero sub-indices, and is named the “basic eigenvectors”. We sort them as in Table 1.

When the topological embedding dimension is unknown, we analyze the behavior of the eigenvectors. The number of basic eigenvectors is equal to the dimension of the topological

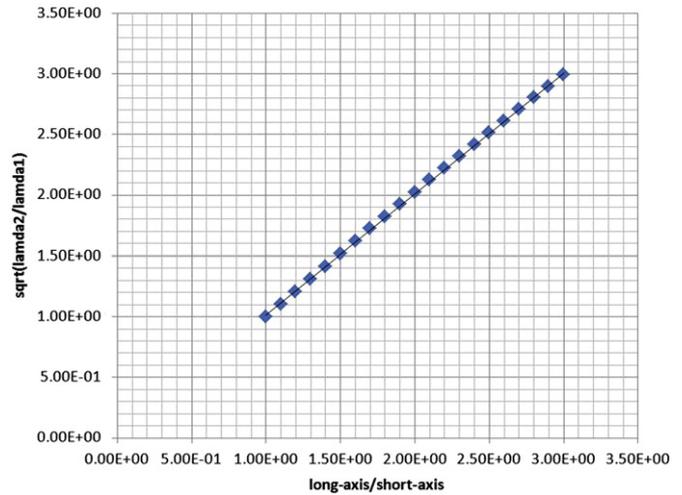


Fig. 9. Relationship between the global scale of the original data samples and the LLE eigenvalues. The x-axis is the ratio of the long-axis and short-axis of an ellipse; the y-axis is the square root of the ratio of the first 2 eigenvalues.

embedding (MESED). In addition, basic eigenvectors reconstruct the global structure of an underlying manifold, which is the 4th step of MTA.

4.5. Improved geometric reconstruction using eigenvalues

We improve the geometric reconstruction ability of LLE using eigenvalues. From (2.14), the relationship between an eigenvalue and the scaling of its corresponding eigenvector is

$$\Phi(c\mathbf{Y}_k) = c^2 \Phi(\mathbf{Y}_k) = c^2 \lambda_k, \quad c \in \mathbb{R} \tag{4.14}$$

Eigenvectors are regulated according to (2.17) and are not recovered after topological reconstruction. From (4.14), their scale information c_i can be recovered by

$$\frac{c_i}{c_j} = \sqrt{\frac{\lambda_j}{\lambda_i}} \tag{4.15}$$

This is the 5th step of MTA.

We use an ellipse as an example. Its long axis is denoted as a , and the short axis is denoted as b . Topologically, an ellipse is a circle \mathbb{S}^1 , and implicit expression $(\mathbf{Y}_{1+}^{\mathbb{S}^1}, \mathbf{Y}_{1-}^{\mathbb{S}^1})$ reconstructs an ellipse as a round circle. The global geometric information is lost. So does other locally manifold learning methods [6,7,9,10]. From (4.15), we set $c_{1+} = 1$ and $c_{1-} = \sqrt{\lambda_{1+}/\lambda_{1-}}$, and $(c_{1+} \mathbf{Y}_{1+}^{\mathbb{S}^1}, c_{1-} \mathbf{Y}_{1-}^{\mathbb{S}^1})$ successfully reconstructs an ellipse. When data are well sampled, this reconstruction is precise. Fig. 9 shows the relationship $\sqrt{\lambda_{1-}/\lambda_{1+}}$ when a/b varies from 1.0 to 3.0; 2400 samples are sampled uniformly from the ellipse.

Another weakness of the original LLE is that the smallest eigenvalues do not always correspond to basic eigenvectors

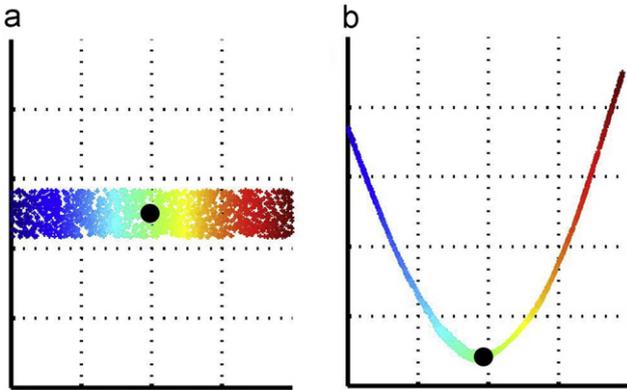


Fig. 10. Local methods are unreliable because of rescaling. (a) Plane \mathbb{P} (rectangle) with a high width/height ratio. (b) A false underlying structure found by a local method. Cited from [25].

because of scaling. Thus, reconstruction with the lowest eigenvectors is not robust. Goldberg et al. [25] discuss over-rescaling the original data. They also propose a criterion to describe how much scaling causes reconstruction failure. They show an example using rectangle \mathbb{P} , which we cite in Fig. 10. We cite their example to explain this weakness and to demonstrate that our method is robust against scaling.

As shown in Fig. 8(b), the first 5 eigenvalues of rectangle \mathbb{P} when its width-height ratio is not too great are

$$0 = \lambda_{0,0} < \lambda_{0,1} < \lambda_{1,0} < \lambda_{1,1} < \lambda_{0,2} \quad (4.16)$$

where λ_{ij} is the eigenvalue of eigenvector $\mathbf{Y}_{ij}^{\mathbb{P}} = (\mathbf{Y}_i^{\text{height}}, \mathbf{Y}_j^{\text{width}})$. If the scale in the “width orientation” maintained as the scale in the “height orientation” decreases, $\lambda_{1,0}$ and $\lambda_{1,1}$ increase. $\lambda_{0,0}$, $\lambda_{0,1}$ and $\lambda_{0,2}$ remain the same because they do not have a “height orientation” component. The scale in the “height orientation” is decreased until $\lambda_{0,2} < \lambda_{1,0}$ and $\lambda_{0,2} < \lambda_{1,1}$. This can always be fulfilled according to (4.14). Then, the smallest 2 non-zero eigenvalues are $\lambda_{0,1}$ and $\lambda_{0,2}$. Using the original LLE, the reconstruction fails because a plane \mathbb{P} is reconstructed as a line \mathbb{R} :

$$\mathbf{Y}_{0,1}^{\mathbb{P}} \rightarrow \mathbf{Y}_{0,2}^{\mathbb{P}} \Leftrightarrow (\mathbf{Y}_0^{\text{height}}, \mathbf{Y}_1^{\text{width}}) \rightarrow (\mathbf{Y}_0^{\text{height}}, \mathbf{Y}_2^{\text{width}}) \Rightarrow \mathbf{Y}_1^{\mathbb{R}} \rightarrow \mathbf{Y}_2^{\mathbb{R}} \quad (4.17)$$

It is evident that our method is robust against scaling and can recover the original scale information.

We now discuss why $4d$ eigenvectors are chosen in Step 2 of MTA. According to the strong Whitney embedding theorem, for a d -dimensional manifold, we need to observe at least $2d$ eigenvectors. And in this section, order of eigenvectors will change because of over scaling. We need to observe more in case not to omit the base eigenvectors. Generally, eigenvector with frequency f has eigenvalue more than 10 times greater than that with frequency $f-1$ [2]. For a rectangle \mathbb{P} , if $\lambda_{0,3} < \lambda_{1,0}$, roughly the width-height rate is greater than 10. For such rectangle sampled discretely with noise, it is ambiguous to define whether it is a narrow rectangle \mathbb{P} or a bold line \mathbb{R} . Similarly, in practice, it is ambiguous to define an ellipse with too big long-axis to short-axis rate. So generally, we observe $4d$ eigenvectors in Step 2 of MTA as default.

5. Manifold multi-resolution analysis (MMA) method

In Section 4, we describe the MTA method, which is a reliable way to determine the typical global structure for tracking and perception understanding. In the field of pattern recognition, LLE

is used for dimension reduction and as a feature extraction tool. In this field, the underlying structure need not be given. However, non-basic eigenvectors are still required to have exact meanings.

In this section, we introduce and prove the manifold multi-resolution analysis (MMA) method, which extracts nonlinearity from a manifold with multi-resolution. This method does not reconstruct the global structure of manifold; thus, it is universally applicable to all those manifolds that satisfy the “manifold hypothesis”. In Section 5.1, we discuss the origins of MMA. Based on this preparation, we introduce the method in Section 5.2. Further, in Section 5.3, we prove the multi-resolutional nonlinearity capture property of LLE eigenvectors. Two examples are given in Section 5.4.

5.1. Ideas of manifold multi-resolution analysis

Many studies have used LLE eigenvectors as features in pattern recognition [1,2,10–16,18,26,35,52], which is analogous to how PCA, Isomap [4] or other global eigen-methods have been used. As discussed in Section 3.5, LLE is a local method, and using it in the same way as global methods is not acceptable. For example, one inaccurate statement is that non-basic eigenvectors of LLE capture only noise because non-basic eigenvector of global methods capture only noise. In reality, basic LLE eigenvectors define the topology without periodic repeating, but higher eigenvectors do not. This is why LLE performs better than PCA if only a few eigenvectors are used and performs worse if more eigenvectors are used.

In Sections 4.3.1 and 4.3.2, we have mentioned function multi-resolution analysis methods such as Fourier series decomposition and Chebyshev decomposition. In addition to eigen-decomposition of typical manifolds in Section 4.3, we also discuss the multi-resolution property of LLE-eigenvectors. If we can find the meaning of each eigenvector, we can use LLE as a manifold multi-resolution analysis tool. This method is given in Section 5.2, and the proof is given in Section 5.3.

5.2. Manifold multi-resolution analysis (MMA) method

We claim that LLE eigenvectors are multi-resolutional; LLE eigenvectors capture manifold nonlinearity by their curves in different resolutions.

The MMA method includes the following steps:

Step 1: Determine whether or not the data samples satisfy the “manifold hypothesis” introduced in Section 2.1.2. If yes, calculate the “local dimension” d of their underlying manifold \mathbb{S} with a local dimension determining method [20,21] if d is not known a priori.

Step 2: Conduct LLE on data samples and return the 2nd least eigenvectors $\{\mathbf{Y}_i\}$, where n is the desired analysis resolution.

Step 3: Find the local extrema of each eigenvector \mathbf{Y}_i according to the following method:

For a particular point \mathbf{x}_i , find its neighborhood $\{\mathbf{x}_{ej}\}$ in the neighborhood relation matrix \mathbb{E} (2.4). If $\mathbf{Y}_k(\mathbf{x}_i) > \mathbf{Y}_k(\mathbf{x}_{ej})$ (or $\mathbf{Y}_k(\mathbf{x}_i) < \mathbf{Y}_k(\mathbf{x}_{ej})$) for all $\mathbf{x}_{ej} \in \{\mathbf{x}_{ej}\}$, then \mathbf{x}_i is a local maximum (or minimum).

The number of local extremas in eigenvector \mathbf{Y}_i represents its resolution.

Step 4: Find local curvature patches according to the following method:

Start from a particular local extrema \mathbf{x}_i . Add its neighbor \mathbf{x}_{ej} to its set of curve parts $\{\mathbf{x}_{i,C}\}$ if $\mathbf{Y}_k(\mathbf{x}_i) \times \mathbf{Y}_k(\mathbf{x}_{ej}) > 0$. For each point in the set $\{\mathbf{x}_{i,C}\}$, repeat the search until no new point is added.

Each set $\{\mathbf{x}_{i,C}\}$ is a local curve patch that captures nonlinearity from the manifold in the corresponding resolution. The most nonlinear location is identified by its local maximum.

5.3. Proof of the multi-resolutional nonlinearity capture property of LLE eigenvectors

Step 1: The multi-resolution property of LLE-eigenvectors.

In this step, we show that, for any given n -dimensional manifold \mathbb{S}^n , its eigenvectors are multi-resolutional, and each of its eigenvectors can be divided into small curves by equation $\mathbf{Y}_i=0$.

As discussed in Section 3.5, a d -dimensional manifold has a topological embedding space \mathbb{R}^T that only supports T orthogonal eigenvectors. These T eigenvectors are called basic eigenvectors in Section 4.4. Basic eigenvectors can be orthogonal to each other because of freedom in topology. Moreover, as discussed in Section 3.3, eigenvectors should preserve the continuity and topology of the original manifold. Non-basic eigenvectors are not topologically orthogonal to basic eigenvectors because there is no topological freedom.

Freedom is found in continuous functional space by varying the resolution, which is common in general function analysis [48]. For example $\mathbf{Y}_T^{\mathbb{S}^n} : \mathbb{S}^n \rightarrow \mathbb{R}$ can generate an eigenvector by increasing the resolution.

\mathbb{S}^n can be divided into two sub-patches \mathbb{S}_1^n and \mathbb{S}_2^n with mappings $h_1(\mathbb{S}_1^n) = \mathbb{S}^n$ and $h_2(\mathbb{S}_2^n) = \mathbb{S}^n$. Then, we find coefficients c_1 and c_2 to generate the eigenvector $\mathbf{Y}_{2T}^{\mathbb{S}^n}$:

$$\mathbf{Y}_{2T}^{\mathbb{S}^n}(t) = \begin{cases} c_1 \mathbf{Y}_T^{\mathbb{S}^n}(h_1(t)) & t \in \mathbb{S}_1^n \\ c_2 \mathbf{Y}_T^{\mathbb{S}^n}(h_2(t)) & t \in \mathbb{S}_2^n \end{cases} \quad (5.1)$$

which satisfies

$$\begin{aligned} & \int_{t \in \mathbb{S}^n} \mathbf{Y}_{2T}^{\mathbb{S}^n}(t) \mathbf{Y}_T^{\mathbb{S}^n}(t) dt \\ &= \int_{t \in \mathbb{S}_1^n} \mathbf{Y}_{2T}^{\mathbb{S}^n}(t) \mathbf{Y}_T^{\mathbb{S}^n}(t) dt + \int_{t \in \mathbb{S}_2^n} \mathbf{Y}_{2T}^{\mathbb{S}^n}(t) \mathbf{Y}_T^{\mathbb{S}^n}(t) dt \\ &= \frac{1}{b_1} \int_{t \in \mathbb{S}_1^n} c_1 \mathbf{Y}_T^{\mathbb{S}^n}(t) \mathbf{Y}_T^{\mathbb{S}^n}(t) dt + \frac{1}{b_2} \int_{t \in \mathbb{S}_2^n} c_2 \mathbf{Y}_T^{\mathbb{S}^n}(t) \mathbf{Y}_T^{\mathbb{S}^n}(t) dt \\ &= \frac{c_1}{b_1} + \frac{c_2}{b_2} = 0 \end{aligned} \quad (5.2)$$

where b_1 and b_2 are the coefficients that are modified as a result of changing the integration domain. The resolution of $\mathbf{Y}_{2T}^{\mathbb{S}^n}$ is doubled according to (5.1).

We note that $\mathbf{Y}_0^{\mathbb{S}^n} \equiv \pm 1$, which is orthogonal to $\mathbf{Y}_0^{\mathbb{S}^n}$ indicates that $\mathbf{Y}_i^{\mathbb{S}^n}(t)$ are centralized:

$$\int_{t \in \mathbb{S}^n} \mathbf{Y}_0^{\mathbb{S}^n}(t) \mathbf{Y}_i^{\mathbb{S}^n}(t) dt = \pm \int_{t \in \mathbb{S}^n} \mathbf{Y}_i^{\mathbb{S}^n}(t) dt = E(\mathbf{Y}_i^{\mathbb{S}^n}(t)) = 0 \quad (5.3)$$

Thus, \mathbb{S}^n can be divided into two sections by $\mathbf{Y}_i^{\mathbb{S}^n} = 0$: $\{t | t \in \mathbb{S}^n, \mathbf{Y}_i^{\mathbb{S}^n}(t) > 0\}$ and $\{t | t \in \mathbb{S}^n, \mathbf{Y}_i^{\mathbb{S}^n}(t) < 0\}$. We note that $\mathbf{Y}_i^{\mathbb{S}^n}$ is a continuous mapping, which is discussed in Section 3.3; thus, \mathbb{S}^n is divided into several continuous patches. According to (5.1), a higher resolution function has more and smaller patches.

According to the strong Whitney embedding theory, each resolution level can stand only $2d$ eigenvectors. Therefore, at least $2nd$ eigenvectors are needed for the desired resolution n .

Step 2: Local curve sections can capture local nonlinearities in the original manifold.

$\mathbf{Y}_k^{\mathbb{S}^n}$ is a continuous mapping on bounded domain, thus, each local patch has a local extrema. For a particular vector \mathbf{x}_i , its K neighbors are temporarily denoted as $\mathbf{x}_{e1}, \mathbf{x}_{e2}, \dots, \mathbf{x}_{eK}$. We may assume $\mathbf{x}_{e1}, \mathbf{x}_{e2}, \dots, \mathbf{x}_{eK}$ is sorted by

$$\mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{e1}) \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{e2}) \leq \dots \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{eK}) \quad (5.4)$$

If \mathbf{x}_i is not a local extrema:

$$\mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{e1}) \leq \dots \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{ej}) \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_i) \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{ej+1}) \leq \dots \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{eK}), \quad (5.5)$$

$$j = 1, 2, 3, \dots, K-1$$

We say \mathbf{x}_i lies inside its neighbors.

If \mathbf{x}_i is a local extremum, for example, if \mathbf{x}_i is a local maximum:

$$\mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{e1}) \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{e2}) \leq \dots \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{eK}) \leq \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_i) \quad (5.6)$$

We say \mathbf{x}_i lies outside its neighbors.

In Section 3.2, we represent the tradeoff between “least square fit” (2.7) and “centralization” (3.5). Regarding to eigenvector $\mathbf{Y}_k^{\mathbb{S}^n}$, if \mathbf{x}_i is located in a generally nonlinear location in the original manifold, then \mathbf{x}_i is considered at the center of its neighbors. Furthermore, \mathbf{y}_i should be at the center of $\{\mathbf{y}_{ej} = \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{ej})\}$ to reduce the reconstruction error $|\mathbf{y}_i - \sum_{j=1}^K w_{ej} \mathbf{Y}_k^{\mathbb{S}^n}(\mathbf{x}_{ej})|^2$ (2.13). In the same way, if \mathbf{x}_i is from a location that is the most nonlinear, \mathbf{x}_i is considered to be far away from the center of its neighborhoods. In addition, \mathbf{y}_i should be at a local extrema, that is far away from its neighbors $\{\mathbf{y}_{ej}\}$. Similarly, curve patches can capture nonlinearity from the local patch of the original manifold.

Step 3: Curve sections of eigenvectors with different scales capture nonlinearity at different scales.

The last step of the proof discusses the global property of the capture. Previously, we focused on the reconstruction errors raised by nonlinearity of the manifold. This kind of error is from the locally orthogonal space of the manifold $\mathbb{R}^D \setminus \mathbb{R}^d$. The notation in local spaces was described in Section 3.1.

Typically, reconstruction error results from the locally tangential space \mathbb{R}^d is small. Unlike the orthogonal space, points in the tangential space can always be reconstructed without error, which is discussed in Sections 3.1 and 3.2. Thus, global rescaling does not increase reconstruction error according to (4.14). However, if we rescale each local part of the manifold with a different scale (nonlinearly), the reconstruction point $\mathbf{x}'_i = \sum_{j=1}^K w_{ej} \mathbf{x}_{ej}$ deviates from the point \mathbf{x}_i and increases reconstruction error.

In Step 2, we proved that curved parts of the eigenvectors are likely to capture nonlinearities of the original manifold to reduce the reconstruction error (2.5). If there are nonlinearities in the manifold, curved patches move to those places to capture them. Such movement causes a non-uniform redistribution of samples, which increases the reconstruction error.

Compared with error from nonlinearities, error from redistribution is very small because it is determined by distribution differentiation. However, the ultimate goal is to minimize the overall reconstruction error, thus, nonlinearity capture and redistribution are tradeoffs. Curve patches do not move very extremely to capture nonlinearity. Otherwise, the reconstruction error reduced by capture does not compensate for the error raised by tangential redistribution.

In other words, a certain curve section is most likely to capture nonlinearity in the same scale to optimize the total reconstruction error. Thus, LLE is a useful tool to capture nonlinearities at different resolutions.

We have proved the MMA method. \square

5.4. Examples

We give two examples.

- (1) The first one shows “move and capture”: a line with 3 nonlinear places, as shown in Fig. 6(h). As proved in Section 4.4,

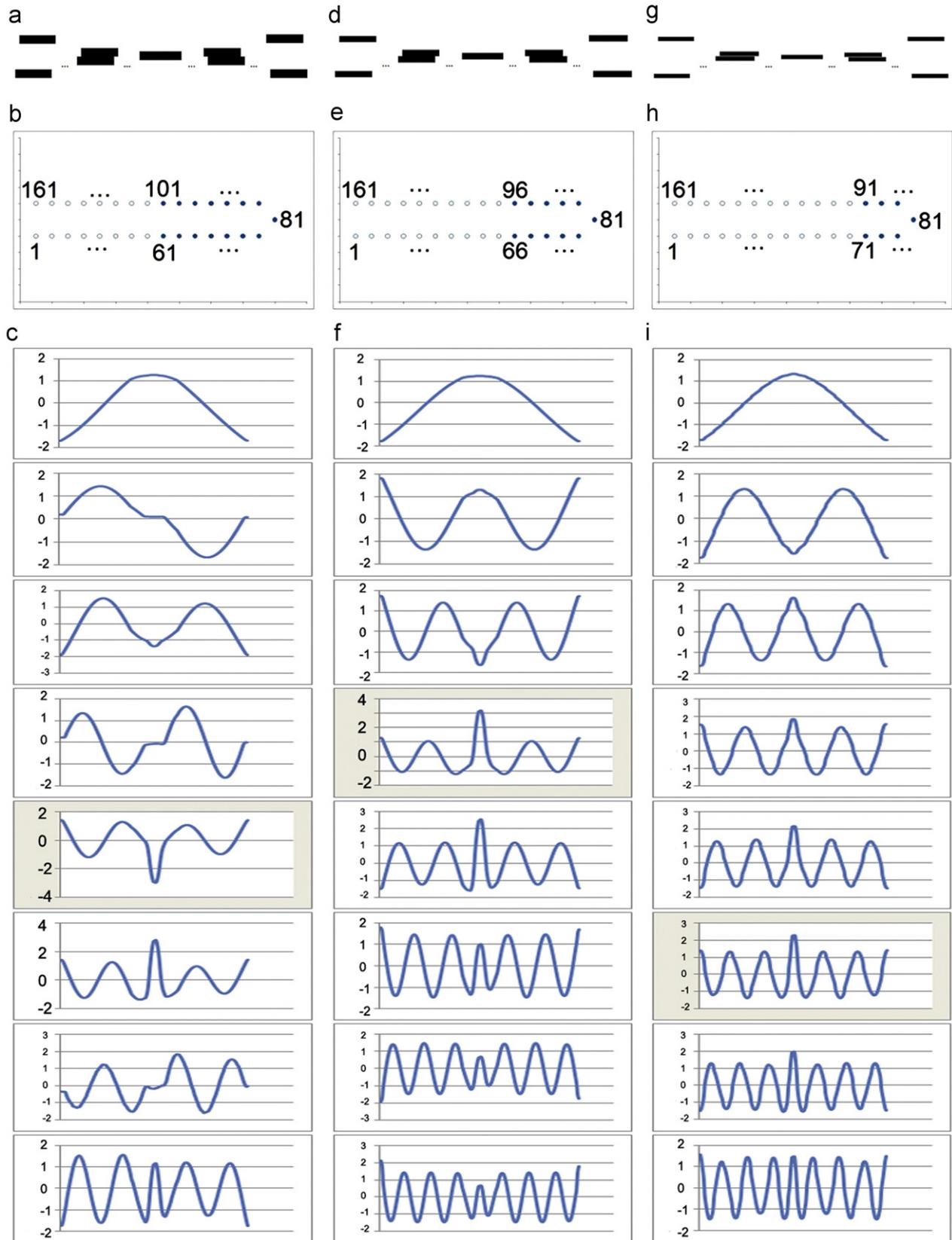


Fig. 11. An example to demonstrate the monotonic relationship between the resolution of eigenvectors and the scale of manifold nonlinearity: (a) a series of images that consist of 161 frames describing “the movement of two rectangles”. The height of the rectangles is 40 pixels. (b) The topological structure of data samples. Those images with an occlusion of the two rectangles are denoted with solid points; (c) the lowest 8 LLE eigenvectors. The most obvious eigenvectors with “occlusion” are highlighted. (d)–(f) The same experiment as above, but the height of the rectangles is 30 pixels. (g)–(i) The same experiment as above, but the height of the rectangles is 20 pixels.

the eigenvector Y_1^R captures only global topology. Moreover, as proved in Section 5.3, Y_2^R captures one nonlinearity, while Y_3^R captures two and Y_4^R captures all the three. The standard eigenvectors are compared with these eigenvectors in Fig. 6(b). Local extremas have moved from their original locations to locations with nonlinearity, which shows the “move and capture”.

- (2) The second example shows “capture in the corresponding resolution”. Fig. 11(a) is a series of images that shows two rectangles that “move together”, “occlude each other” and “separate”. Specifically, the size of the canvas is 200×200 pixels, and the size of the rectangles is 180×40 pixels. The motion is composed of 161 images. In the 61st through 101st images, the two rectangles occlude each other. The underlying topology is shown in Fig. 11(b), where images that show “occlusion” are denoted as solid points. The “occlusion” is a nonlinearity on the manifold. The first 8 LLE eigenvectors are shown in Fig. 11(c). We see some of them capture the nonlinearity with a local extrema. We have highlighted the one with most significant local extrema.

We performed the same experiment but rescaled the rectangles to 180×30 pixels and 180×20 pixels, as shown in Fig. 11(d)–(i), where the nonlinearity raised by “occlusion” has a smaller scale. Such differences in scale are shown in their underlying topology. We have also highlighted the eigenvector with the most significant local extrema for “occlusion”.

Table 2
The product invariance between the scale of nonlinearity and the resolution of eigenvector.

Scale of nonlinearity (number of images with “occlusion”)	40	30	20
Resolution of eigenvector (number of periodic)	3	4	6
Product	120	120	120

We observe the relationship between the scale of nonlinearity and the resolution of the highlighted eigenvectors in Table 2, which shows eigenvectors with different resolutions that capture nonlinearity in the corresponding resolution.

6. Applications in 3D object recognition

We give two applications in 3D object recognition using the methods introduced in Sections 4 and 5.

6.1. 3D object recognition

This application demonstrates the use of non-basic LLE eigenvectors.

One efficient method of 3D object recognition is to present the original 3D object as a set of its 2D projections [50], as shown in Fig. 12. The difficulty of this method is choosing the 2D projections that lie on the viewpoint sphere S^2 , which is a 2-dimensional manifold. Usually, such projections are chosen to have equal intervals. Here, we choose representative projections by finding its local extremas using MMA. Referring to Fig. 6(h) and (i), local extremas can represent the original manifold perfectly by linear interpolation.

We introduce our experiment in detail: Fig. 13(a) shows the 3D airplane models we used in the recognition experiment, which are chosen from PSB [49]. Fig. 13(b) illustrates the method of obtaining 2D projections, where the blue part is used as training set and the red part is used as testing set. We note that the training set and testing set are different because of focus projection. The ratio of the model scale to the focus length is about 2:5, which causes significant shape deformation.

For each particular model, its training set lies on a 1-dimensional manifold. We choose its representation projections by finding the local extrema of LLE-eigenvectors. This is similar to choosing local extremas in Fig. 6(g). The Fourier descriptor [54] was extracted from projections. The classification library is

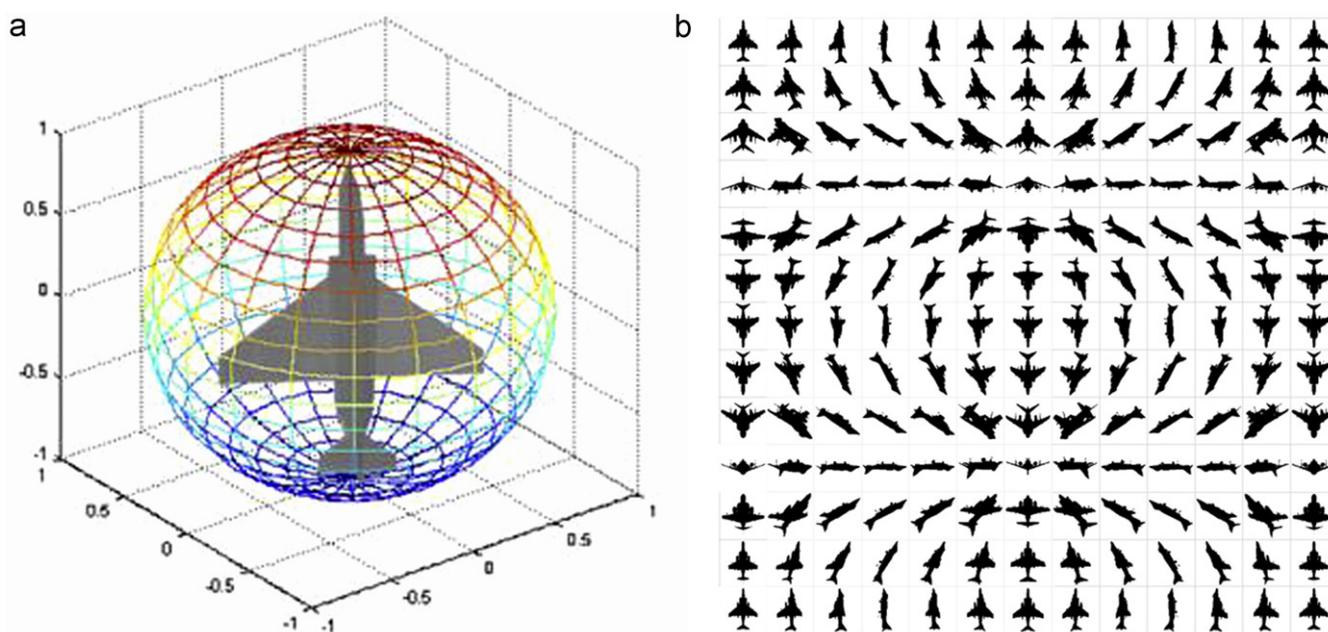


Fig. 12. (a) Viewpoint space of 3D objects, which is a 2-dimensional manifold S^2 . (b) Projection images used to describe the original 3D objects.

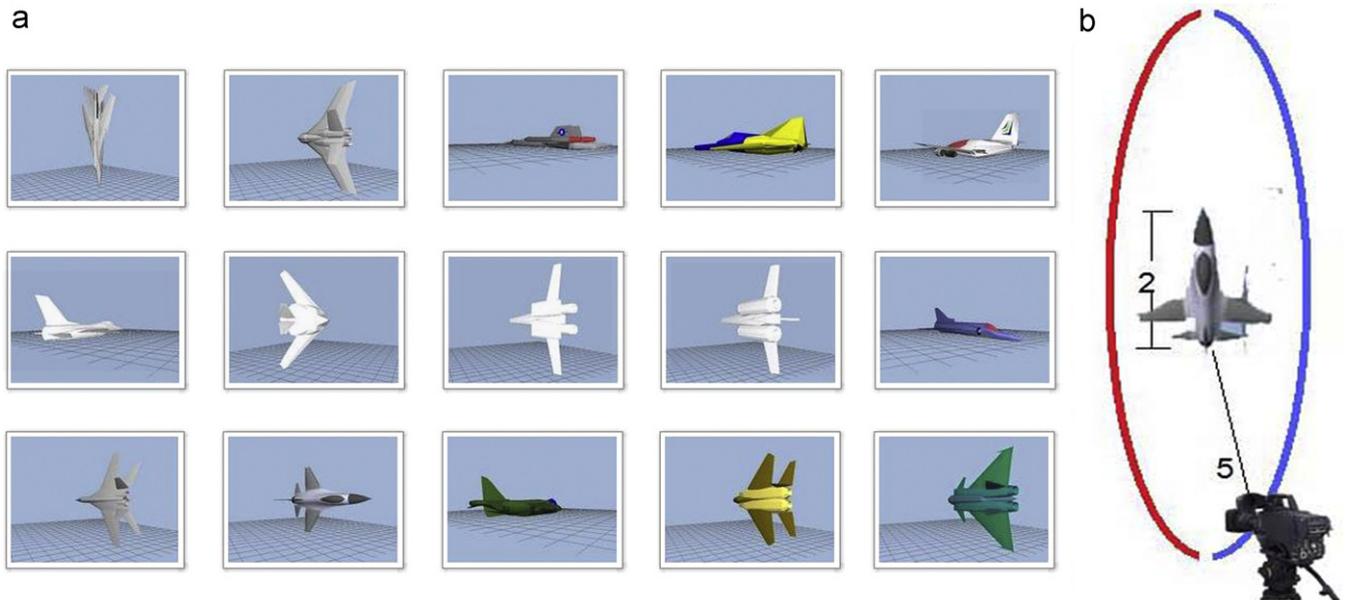


Fig. 13. (a) 15 airplane models chosen from PSB [49] for the recognition test. (b) Method to obtain projection images. The blue part is used as the training set and the red part is used as the testing set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Recognition of 5 models using different eigenvectors.

Eigenvector used to choose representative images	Number of images chosen for the library of each model	Images tested	Recognition rate: representative images chosen by	
			MMA	Equal interval
Y_2^R	3	1800	0.591	0.461
Y_3^R	4	1800	0.614	0.585
Y_4^R	5	1800	0.674	0.669

Table 4
Recognition rate of different number of models using Y_2^R .

Number of models	Images tested	Recognition rate: representative images chosen by	
		MMA	Equal interval
5	1800	0.591	0.461
10	3600	0.458	0.374
15	5400	0.384	0.341

established by storing representative images of each model. Another library with representative images chosen with equal intervals was used for comparison.

In the recognition test, the test image is compared with all representative images in the classification library and classified to the class of its nearest neighbor. Tables 3 and 4 show 2 tests: (1) recognition of 5 models using different non-basic eigenvectors; and (2) recognition of different numbers of models.

Manifold multi-resolution analysis (MMA) performs better in all tests because MMA captures nonlinearities of underlying manifold, which are important representatives.

6.2. 3D object viewpoint space partitioning

The aspect graph was a popular way to find a 2D representation of a 3D object [50]. However, several problems hinder its application: (1) disastrous computational complexity: $O(n^6)$ – $O(n^9)$, where n is the number of edges of the 3D mesh model, which is about 10,000 for the airplane model from PSB [49]; (2) the partitioning is over-detailed, with about 10,000–100,000 partitions for the airplane model from PSB; and (3) it cannot give a hierarchical representation.

3D object viewpoint space partitioning by MMA overcomes all these shortages: (1) the computational complexity of LLE is only $O(n^2)$ – $O(n^3)$, where n is the number of data samples, which is about 2000 for 3D models; (2) partition resolution can be controlled by the number of eigenvectors used; and (3) partitions are hierarchical.

Fig. 14 shows a comparison by partitioning the viewpoint space airplane model.

7. Discussion

7.1. Learn underlying manifold of human walking

Learning the underlying structure of “human walking” is a successful application of LLE in tracking and perception understanding, as shown in Fig. 15 [28,33]. Their results illustrate that LLE successfully separates the two half cycles of human walking cycle, which cannot be separated by traditional linear methods. However, their analysis is not strictly correct.

An “8” shape is not strictly a “manifold” because it does not satisfy the definition of a 1-dimensional manifold in Section 2.1.1. Specifically, the “cross” on “8” is not homeomorphic to \mathbb{R} .

According to the typical eigenvectors of circle \mathbb{S}^1 in Section 4.3.2, an implicit expression of eigenvectors, $(Y_{2+}^{\mathbb{S}^1}, Y_{2-}^{\mathbb{S}^1}, Y_{1+}^{\mathbb{S}^1})$, is always an “8” shape, even though the original manifold is a circle (Fig. 4(d)).

The correct explanation is that $Y_{1+}^{\mathbb{S}^1}$ and $Y_{1-}^{\mathbb{S}^1}$ find the underlying structure of human walking as a circle and discriminate the two half

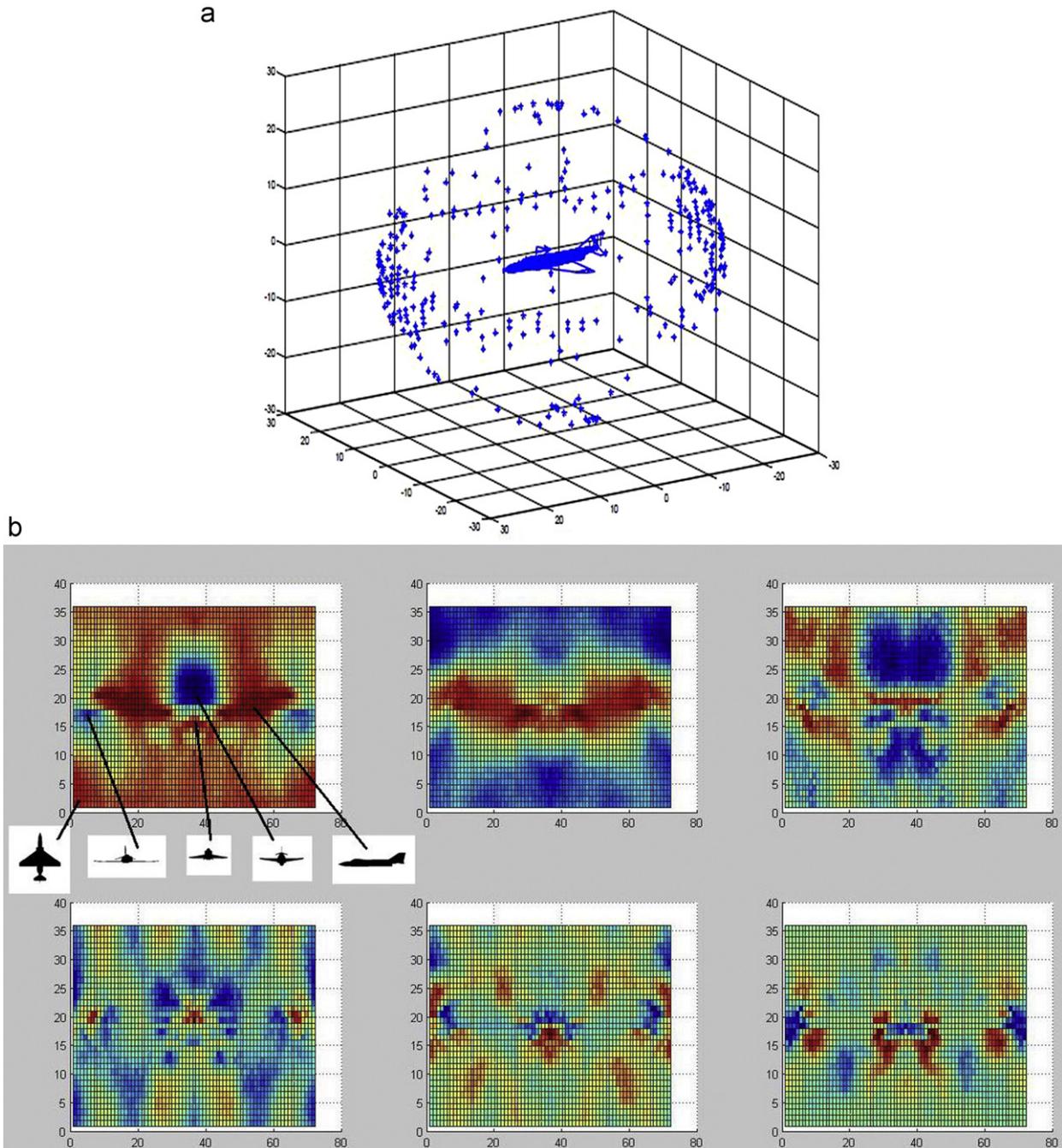


Fig. 14. 3D object viewpoint space partition. (a) Cited from [55]: the aspect graph method by catastrophe theory. Too many partitions are generated, and they are non-hierarchical. (b) The MMA method is shown with the bottom 6 eigenvectors from the explicit expression. For the first eigenvector, some projections on local extremas are shown. The series of eigenvectors is shown, and the partition is hierarchical, and the partition resolution can be controlled.

cycles. $\mathbf{Y}_{2+}^{\mathcal{S}^1}$ and $\mathbf{Y}_{2-}^{\mathcal{S}^1}$ find the underlying structure of human walking as a circle, but do not discriminate half cycles. As analyzed in Section 4.5, eigenvalues λ_{1+} and λ_{1-} are greater than λ_{2+} and λ_{2-} because the difference between 2 half cycles is very slight. Because the original LLE uses the lowest eigenvectors, the 3 lowest eigenvectors in this situation are $\mathbf{Y}_{2+}^{\mathcal{S}^1}$, $\mathbf{Y}_{2-}^{\mathcal{S}^1}$ and $\mathbf{Y}_{1+}^{\mathcal{S}^1}$, which implies an “8”.

MTA would be a better analysis method to determine inter-half-cycle behavior according to $\mathbf{Y}_{1+}^{\mathcal{S}^1}$, $\mathbf{Y}_{1-}^{\mathcal{S}^1}$, λ_{1+} and λ_{1-} , and intra-half-cycle behavior with $\mathbf{Y}_{2+}^{\mathcal{S}^1}$, $\mathbf{Y}_{2-}^{\mathcal{S}^1}$, λ_{2+} and λ_{2-} . For example, as

shown in Fig. 15(e), $\mathbf{Y}_{1+}^{\mathcal{S}^1}$ (LLE (3)) captures the nonlinearity at the “cross” by its local extrema.

7.2. Universality of MTA and MMA on all local methods

MTA can be extended to all the following local methods: LEM [6], HLE [7], DFM [9] and LSTA [10]. They differ from LLE in the way that they capture local information. In the proof of MTA in Sections 3.4, 3.5, 4.3, 4.4 and 4.5, we do not use the specified

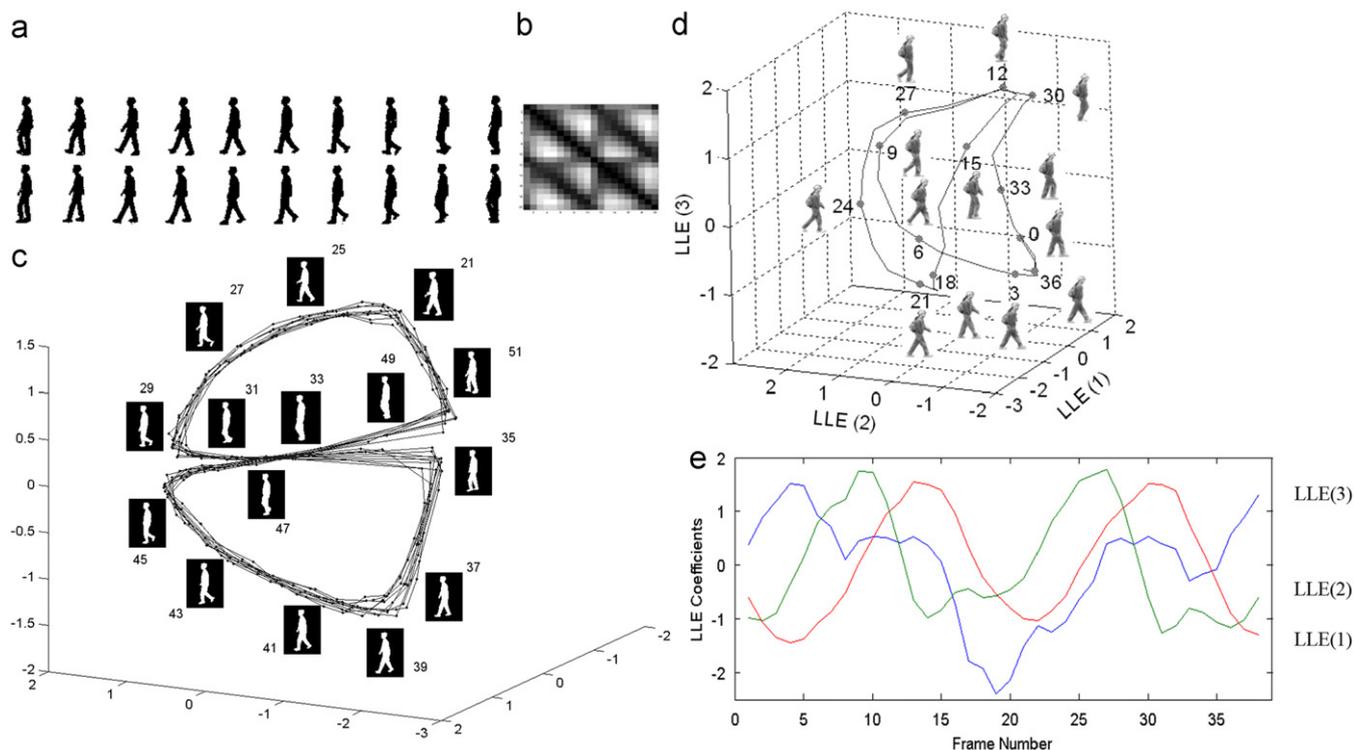


Fig. 15. Experiments on learning the underlying manifold of human walking. (a) A cycle of human walking that consists of two half cycles. (b) The correlation matrix, which shows the high resemblance of the two half cycles. (c) A 3-dimensional reconstruction of the underlying manifold by LLE; the two half cycles are successfully separated. (d) same as (c). (e) The lowest 3 LLE eigenvectors, where LLE(1) is $Y_{2+}^{S_1}$ (red line) (4.6); LLE(2) is $Y_{2-}^{S_1}$ (green line); and LLE(3) is $Y_{1+}^{S_1}$ (blue line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.) (a)–(c) cited from [28]; (d)–(e) cited from [33].

formulas (2.5) or (2.13); thus, any of them can be chosen to replace LLE in the second step of MTA.

In the proof of MMA, we have used the tradeoff between two optimal conditions (2.5) and (3.6) to prove the existence of “nonlinearity capture”. One can use other local methods to replace LLE in Step 2 of MMA if the “nonlinearity capture” ability is proved. The proof is one of our future works.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 60502013) and the National High Technology Research and Development Program of China (863 Program, No. 2006AA01Z115).

References

- [1] Sam T. Roweis, Lawrence K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (2000).
- [2] Lawrence K. Saul, Sam T. Roweis, Think globally, fit locally: unsupervised learning of low dimensional manifolds, *Journal of Machine Learning Research* 4 (2003) 119–155.
- [3] H. Sebastian Seung, Daniel D. Lee, The manifold ways of perception, *Science* 290 (5500) (2000) 2268–2269.
- [4] Joshua B. Tenenbaum, Vin de Silva, John C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* 290 (2000).
- [5] Joshua B. Tenenbaum, Mapping a manifold of perceptual observations. *Advances in neural information processing systems*, 1998.
- [6] M. Belkin, P. Niyogi, Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural computation*, MIT Press, 2003.
- [7] David L. Donoho, Carrie Grimes. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. In: *Proceedings of the National Academy of Sciences*, National Academy Sciences, 2003.
- [8] Kilian Q. Weinberger, Lawrence K. Saul, Unsupervised learning of image manifolds by semidefinite programming, *IJCV* 70 (1) (2006) 77–90.

- [9] R.R. Coifman, S. Lafon, Diffusion maps, *Applied and Computational Harmonic Analysis* 21 (1) (2006) 5–30.
- [10] X. Huo, A.K. Smith. Performance analysis of a manifold learning algorithm in dimension reduction. Technical Paper, Statistics in Georgia Tech, Georgia Institute of Technology, March 2006.
- [11] Laurens van der Maaten, Geoffrey Hinton, Visualizing data using t-SNE, *Journal of Machine Learning Research* 9 (2008) 2579–2605.
- [12] Lin Xiao, Jun Sun, Stephen Boyd. A duality view of spectral methods for dimensionality reduction. In: *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA, 2006.
- [13] Fei Sha, Lawrence K. Saul. Analysis and extension of spectral methods for nonlinear dimensionality reduction. In: *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany, 2005.
- [14] L.J.P. van der Maaten, E.O. Postma, H.J. van den Herik. Dimensionality reduction: a comparative review. Online Preprint, 2008.
- [15] Matthew Brand, Kun Huang. A unifying theorem for spectral embedding and clustering. In: *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Key West, FL, January 2003.
- [16] Yun Fu, Thomas S. Huang. Unsupervised locally embedded clustering for automatic high-dimensional data labeling. In: *Proceedings of the IEEE Conference on ICASSP'07*, 2007, pp. 1057–1060.
- [17] Matthew Brand. Charting a manifold. *Advances in Neural Information Processing Systems*, 2003.
- [18] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang. Graph embedding: a general framework for dimensionality reduction. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- [19] Martin H.C. Lawr, Anil K. Jain, *IEEE TPAMI* 28 (3) (2006).
- [20] B. Kégl. Intrinsic dimension estimation using packing numbers. *Advances in Neural Information Processing Systems*, 2003.
- [21] Elizaveta Levina, Peter J. Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in Neural Information Processing Systems*, vol. 17, 2005.
- [22] Ameet Talwalkar, Sanjiv Kumar, Henry Rowley, Large-Scale Manifold Learning, *CVPR*, 2008.
- [23] Piotr Dollár, Vincent Rabaud, Serge Belongie. Non-isometric manifold learning: analysis and an algorithm. In: *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, OR, 2007.
- [24] Mikhail Belkin, Partha Niyogi, Vikas Sindhwani. On manifold regularization. In: *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTAT 2005)*.
- [25] Y. Goldberg, A. Zakai, D. Kushnir., Manifold learning: the price of normalization, *Journal of Machine Learning Research* 9 (2008) 1909–1939.

- [26] Charles J. Alpert, So-Zen Yao. Spectral partitioning: the more eigenvectors the better. In: Proceedings of the 32nd ACM/IEEE Design Automation Conference; 1995.
- [27] Chan-Su Lee, Ahmed Elgammal. Simultaneous inference of view and body pose using torus manifolds. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006).
- [28] Ahmed Elgammal, Chan-su Lee, Inferring 3D Body Pose from Silhouettes using Activity Manifold Learning, CVPR, 2004.
- [29] Ahmed Elgammal, Chan-su Lee, Separating Style and Content on a Nonlinear Manifold, CVPR, 2004.
- [30] Chan-su Lee, Ahmed Elgammal, Modeling View and Posture Manifolds for Tracking, ICCV, 2007.
- [31] Dick de Ridder, Robert P.W. Duin, Locally Linear Embedding for Classification, CVPR, 2004.
- [32] Yun Fu, Thomas S. Huang. Graph embedded analysis for head pose estimation. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06).
- [33] H. Lim, V. Morariu, O.I. Camps, M. Sznaiier, Dynamic Appearance Modeling for Human Tracking, CVPR, 2006.
- [34] Therdsak Tangkuampien, Tat-Jun Chin. Locally linear embedding for markerless human motion capture using multiple cameras. In: Proceedings of the Digital Imaging Computing: Techniques and Applications (DICTA 2005).
- [35] Junping Zhang, Stan Z. Li, Jue Wang. Manifold learning and applications in recognition. Intelligent Multimedia Processing with Soft Computing, vol. 168, 2005, pp. 281–300.
- [36] Vlad I. Morariu, Modeling Correspondences for Multi-Camera Tracking Using Nonlinear Manifold Learning and Target Dynamics, CVPR, 2006.
- [37] C.Mario Christoudias, Trevor Darrell, On Modelling Nonlinear Shape-and-Texture Appearance Manifolds, CVPR, 2005.
- [38] Changshui Zhang, Jun Wang, Nanyuan Zhao, David Zhang, Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction, Pattern Recognition (2004).
- [39] Ali Rahimi, Ben Recht, Trevor Darrell, Learning Appearance Manifolds from Video, CVPR, 2005.
- [40] Fan R.K. Chung. Spectral graph theory. Conference Board of the Mathematical Sciences, Regional Conference Series in Mathematics Number, vol. 92, American Mathematics Society, 1994.
- [41] Reinhard Diestel, Graph Theory, 3rd ed, Springer, New York, 2005.
- [42] V.A. Zorich, Mathematical Analysis, Springer, 2004.
- [43] Lloyd Nicholas Trefethen. Spectral method in Matlab. society for industrial and applied mathematics, 2000.
- [44] John M. Lee, Introduction to Topological Manifolds, Springer, 2000.
- [45] Jean-Philippe Vert, Koji Tsuda, Bernhard Schoelkopf, A Primer on Kernel Methods, MIT press, Cambridge, MA, 2004.
- [46] J.B. Kruskal, M. Wish, Multidimensional Scaling, Psychometrika, Springer, 1964.
- [47] X. Zhang, Advanced Algebra, Tsinghua University Press, Beijing, 1998.
- [48] Erwin Kreyszig, Introductory Functional Analysis with Applications, Wiley, 1989.
- [49] P. Shilane, P. Min, M. Kazhdan, T. Funkhouser. The Princeton shape benchmark[C]. In: Proceeding of the IEEE Shape Modeling International 2004(SMI'04), IEEE Computer Society, Washington, DC, 2004, pp. 167–178.
- [50] Robert D. Schiffenbauer. A survey of aspect graphs. Technical Report TR-CIS-2001-01, Department of Computer and Information Science, Polytechnic University, Brooklyn, Long Island, New York, February 2001.
- [51] D. Baymer, T. Poggio, Image representations for visual learning, Science, New Series 272 (5270) (1996) 1905–1909.
- [52] Yun Fu, Thomas S. Huang. Locally linear embedded eigenspace analysis. In: Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, 2006.
- [53] Andrew Y. Ng, Michael I. Jordan, Yair Weiss. On spectral clustering: analysis and an algorithm. Advances in Neural Information Processing Systems, 2002.
- [54] D.S. Zhang, G.J. Lu. Shape retrieval using Fourier descriptors. In: Proceedings of the International Conference on Multimedia and Distance Education. Fargo, ND, USA, June 2001, pp. 1–9.
- [55] X. Yu, H. Ma, S. You, Z. Yuan. A solution to efficient viewpoint space partition in 3D object recognition. In: Proceedings of the International Conference on Image Graphics, 2009.
- [56] A. Thomas, V. Ferrari, B. Leibe, T. Tuytelaars, B. Schiele, L.V. Gool, Towards Multi-view Object Class Detection, CVPR, 2006.

Shaodi You received his bachelor's degree in Electronic Engineering from Tsinghua University, PR. China in 2009. He also finished a special set of courses in advanced mathematics and physics. He joined the 3D Image Simulation Lab of the Image & Graphics School of the department since 2008 for the Student Research Training project. His research is mainly focusing on 3D object analysis and recognition. He is now a master student in the Computer Vision Lab, The University of Tokyo.

Huimin Ma received her PhD in Mechanical Electronic Engineering from the Beijing Institute of Technology in PR. China in 2001. She is currently an associate professor of the Image & Graphics School of Department of Electronic Engineering, Tsinghua University, PR. China, and her specialty is signal and information processing. Research interests include 3-D object pattern recognition and tracking, modeling and simulation, image processing, etc. Dr. Ma is also the executive director and the vice-secretary-general of China Society of Image and Graphics.