

Atlas Compatibility Transformation: A Normal Manifold Learning Algorithm

Zhong-Hua Hao Shi-Wei Ma Fan Zhao

School of Mechatronic Engineering and Automation, Shanghai Key Laboratory of Power Station Automation Technology, Shanghai University, Shanghai 200072, China

Abstract: Over the past few years, nonlinear manifold learning has been widely exploited in data analysis and machine learning. This paper presents a novel manifold learning algorithm, named atlas compatibility transformation (ACT). It solves two problems which correspond to two key points in the manifold definition: how to chart a given manifold and how to align the patches to a global coordinate space based on compatibility. For the first problem, we divide the manifold into maximal linear patch (MLP) based on normal vector field of the manifold. For the second problem, we align patches into an optimal global system by solving a generalized eigenvalue problem. Compared with the traditional method, the ACT could deal with noise datasets and fragment datasets. Moreover, the mappings between high dimensional space and low dimensional space are given. Experiments on both synthetic data and real-world data indicate the effect of the proposed algorithm.

Keywords: Nonlinear dimensionality reduction, manifold learning, normal vector field, maximal linear patch, ambient space.

1 Introduction

Since all the objects change gradually and regularly in the world, many high-dimensional data in real-world applications can be modeled as data points lying close to a low-dimensional nonlinear manifold. Discovering the structure of the manifold is a challenging unsupervised learning problem, which is called dimensionality reduction in pattern recognition^[1–10]. The discovered low-dimensional structures can be further used in classification, motion analysis, clustering and data visualization^[11, 12]. A classical example of the low-dimensional manifold embedded in high-dimensional input spaces is that a set of images of an individual face is rotated from one side to another. The key point is that the dimensions of the input spaces can be very high (e.g., the number of pixels for each image in the image dataset), however, the intrinsic dimensionality for dataset is limited by few factors (e.g., the rotation faces lie on a line in input space-one dimensional). The purpose of manifold learning algorithms is to find a low-dimensional embedding that preserves properties (e.g., geodesic distance or local relationships) of the high-dimensional input dataset.

In this paper, we propose a nonlinear dimensionality reduction algorithm, called atlas compatibility transformation (ACT). This method is inspired by the manifold definition in classical differential geometry, which is normal and intuition in mathematics. The ACT charts a manifold into maximal linear patch (MLP) adaptively based on nor-

mal vector field firstly, and then aligns the patches into a global coordinate system based on compatibility condition. Compared with the existing methods, ACT could locate the noise data by scalar curvature field of the manifold. Suppose the intrinsic dimension of manifold is d , for each patch, it can be regarded as a “flat” hyperplane in a $d+1$ Euclidean space, which is called ambient space in mathematics. The curvature should stay in a low range for each patch. However, if a noise point adds to a patch, the curvature value should jump to an extremely high level suddenly. Therefore, the noise points can be located and successive methods can be utilized to clear them away from the training dataset. Moreover, the ACT could deal with fragmentary training data. In this paper, the fragmentary data refers to the training data that are incomplete sampling due to some uncontrollable reason. In ACT, the distribution of missing data can be estimated by the manifold geometry structure depicted by MLPs.

The rest of the paper is organized as follows. We begin with a brief review of the existing dimensionality reduction methods in Section 2. In Section 3, the motivation and basic idea of the proposed ACT are illustrated. Section 4 depicts the charting strategy in detail. Section 5 presents the algorithm that aligns the charts into a global coordinate system. Experimental results are given in Section 6. And Section 7 is devoted to the conclusions and future work discussion.

2 Related work and existing problems

2.1 Related work

Over the past two decades, a great number of algorithms have been proposed to address the problem of dimensionality

Regular paper
Manuscript received September 22, 2013; accepted March 28, 2014
This work was supported by National Natural Science Foundation of China (No. 61171145) and Shanghai Educational Development Foundation (No. 12ZZ083).
Recommended by Guest Editor Xin Sun
© Institute of Automation, Chinese Academy of Science and Springer-Verlag Berlin Heidelberg 2015

reduction^[13]. These algorithms can be divided into two categories: linear dimensionality reduction methods^[14] and nonlinear dimensionality reduction methods.

Principle component analysis (PCA), multi-dimensional scaling (MDS) and linear discriminant analysis (LDA) are three representative linear dimensionality reduction methods. Since these methods cannot discover the curved or nonlinear structures of the input data, they can only deal with flat Euclidean structures.

To discover the nonlinear structures of input data, many nonlinear dimensionality reduction methods have been proposed, such as self-organizing maps^[15], principal curves and its extensions, autoencoder neural networks and generative topographic maps. However, these methods often suffer from difficulties in designing a cost function or tuning too many free parameters. Moreover, most of these methods are computationally expensive, which limits their utility in high-dimensional data sets. Recently, kernel methods^[16, 17] have provided new ways to perform linear algorithms in implicit higher-dimensional feature space. Although these methods can improve the performance of linear ones, most of them are still computationally expensive, and some of them have difficulties in designing cost functions or tuning many parameters.

Over the past few years, manifold learning algorithms^[1–10] have attracted attention of researchers due to their geometric intuition, nonlinear nature and computational feasibility. The basic assumption of manifold learning is that the input data lie on or close to a smooth low-dimensional manifold. Each manifold learning algorithm attempts to uncover the manifold and preserve a different geometrical property of the underlying manifold.

Isomap^[1] and local linear embedding (LLE)^[2] were the first two classical manifold methods published in *Science* in 2000. Isomap constructs a shortest neighborhood graph firstly and then utilizes MDS to seek the low-dimensional embedding. LLE calculates the reconstruction weights for each point which depicts the local geometry structure in a high observation space and then constructs the low-dimensional embedding by preserving this structure. Then, Laplacian eigenmaps (LE)^[3] and Hessian LLE (HLLE)^[5] were proposed in 2003. They estimate the Laplacian and Hessian on the manifold respectively to minimize the squared gradient of a low-dimensional embedding map. In 2004, local tangent space alignment (LTSA)^[7] was proposed. This method constructs the tangent space at each data point firstly and then aligns these tangent spaces in a global coordinate system. Riemannian manifold learning (RML)^[9] and maximal linear embedding (MLE)^[10] were proposed in 2008 and 2011, respectively. The RML computes normal coordinates that map the input high-dimensional data into a low-dimensional space. The MLE divides manifold into maximal linear patch (MLP) firstly and then utilizes a landmark-based alignment method to obtain the low-dimensional embedding.

2.2 Existing problems

While viewing previous work, there are mainly three problems worth to be discussed.

The manifold learning with noise data. Manifold learning algorithms are sensitive to noisy data that lie outside the manifold (outliers)^[4, 18]. Several manifold learning algorithms can be adopted to extract the intrinsic features of manifold in high dimensional space by preserving the local geometric characteristics. However, due to the locality geometry preservation, these manifold learning methods are sensitive to noise. Moreover, experiments show that, for different types of noises, there is no single method that could be immune to all of them and the dimensional reduction results are uncontrollable and unpredictable. Compared with the existing methods, the ACT could distinguish noise data from manifold dataset by analyzing curvature of the manifold.

The manifold learning with fragmentary data. High-density sampling is an important assumption for all of the existing manifold learning algorithms. However, during the sampling process, some data may be lost or not sampled. Compared with the existing methods, the ACT could estimate the geometry distribution of the missing data by analyzing the basic vectors of neighboring MLPs.

The incremental-learning problem. Most manifold learning methods operate in “batch”, in which the mapping between input high dimensional space and low-dimensional embedding space is not given. Therefore, new data points could not get their coordinates in the embedding space directly. However, incremental learning is essential for practical applications. Recently, many methods have been proposed to solve this problem^[19, 20]. Our algorithm gives the mapping between original input high-dimensional space and low-dimensional embedding space through a mapping set which contains mappings between MLPs and their coordinate areas in the embedding space. The embedded coordinates for new data points could be computed as long as the MLPs that contain new data points are determined and the data points in the original space can be reconstructed from the embedded coordinates.

The proposed ACT method in this paper provides a solution to the above problems, with the five characteristics briefly summarized in Section 1. The details of the algorithm are described in the following sections.

3 Motivation and basic idea

In this paper, the proposed ACT method is inspired by the manifold definition in classical differential geometry. Therefore, we list the definition below firstly and then move on to the motivation and basic idea.

3.1 Manifold definition

In differential geometry, a topological space M that has an open cover O_α , i.e., \bigcup_α is an n -dimensional manifold if: 1) for each open set O , \exists homeomorphism mapping

$\psi_\alpha : O_\alpha \rightarrow V_\alpha$ (V_α is an open set in \mathbf{R}^n), 2) if $O_\alpha \cap O_\beta \neq \emptyset$, the complex mapping $\psi_\beta \circ \psi_\alpha^{-1}$ has C^∞ differentiable structure (smooth).

In differential geometry, (O_α, ψ_α) is called a “chart” which maps an open set O_α to a Euclidean space through a homeomorphic mapping ψ_α . Any manifold can be described by a collection of charts, also known as an atlas. It is worth mentioning that in manifold M (topological space), the elements have no coordinates initially. However, they get coordinates in the Euclidean space after selecting a set of coordinate basis. Therefore, (O_α, ψ_α) is also called a local coordinate system.

The condition 2) is called “compatibility”, which gives the differential structure to charts. It indicates that where the domains of charts overlap (the shadow area in Fig. 1), the coordinates defined by each chart are required to be differentiable with respect to the coordinates defined by every chart in the atlas. The mappings that relate the coordinates defined by the various charts to one another are called transition maps. Suppose an element $p \in O_\alpha$, and that $\psi_\alpha(p)$ maps p to space V_{p1} . $\{X_1, X_2, \dots, X_n\}$ is a coordinate basis of V_{p1} and $\{x^1, x^2, \dots, x^n\}$ is the coordinate of p in V_{p1} . If $O_\alpha \cap O_\beta \neq \emptyset$, the elements in the intersection area of $O_\alpha \cap O_\beta$ could not only obtain coordinates $\{x^\mu\} (\mu = 1, 2, \dots, n)$ in V_{p1} by ψ_α , but also obtain coordinates $\{x^\nu\} (\nu = 1, 2, \dots, n)$ in V_{p2} by ψ_β . The transition mapping between (O_α, ψ_α) and (O_β, ψ_β) , $\psi_\beta \circ \psi_\alpha^{-1}$, can be depicted by the equation for coordinate system transformation.

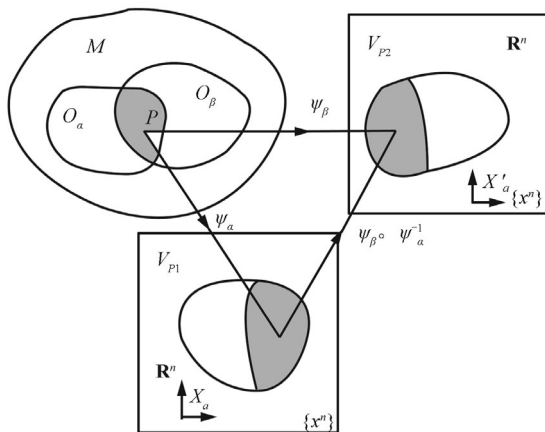


Fig. 1 Manifold definition

3.2 Motivation and basic idea

Manifold definition gives us a nature way for nonlinear dimensionality reduction: 1) charting the manifold reasonably, 2) aligning the charts into a single global space based on compatibility. It is the basic idea of atlas compatibility transformation. Correspondingly, the two key problems we will confront are how to chart the manifold reasonably and how to find the translation maps that align patches into a single global coordinate system.

The methods that deal with these two problems are the main novelty of the ACT. For the first problem, the manifold is divided into maximal linear patches (MLPs) based on the curvature of the manifold. For the second problem, we seek for the proper mapping that maps each patch to a global coordinate system. What we want to preserve are the relationships among MLPs by minimizing the sum of the coordinate errors of MLPs intersection areas in the global coordinate system. Our alignment method follows [6, 7]. The following two sections present the way to solve these key problems in detail.

4 Manifold charting

In manifold definition, one open set on the manifold and its homeomorphism mapping constitute one chart. In this paper, an open set is called a patch of the manifold. Since homeomorphism mapping maps the chart to a Euclidean space, the patch should be flat. In other words, the patch should be approximate to a linear space.

Several manifold methods, such as ISOMAP, LLE, LE, usually define their patches on each data point by k -NN or ε -ball, generally of fixed and small size. However, it may divide a large linear patch into multiple smaller ones. Evidently, these small patches could not depict the structure of the manifold. Moreover, small changes to the size of the trusted set can make the resulting embedding unstable in some cases. Some efforts have been made to alleviate the effect on the fixed neighborhood size^[8, 10].

In this paper, based on the curvature of the manifold, we divide the manifold into maximal linear patches adaptively by measuring the degree of manifold curvature at each local area. The curvature of manifold is estimated by the normal vector field of the scatter data points that distribute on the manifold. Firstly, a neighborhood graph is constructed for each point of the data. Then, PCA is utilized to obtain the normal vector field of the manifold. The manifold curvature can be also estimated at this step. Finally, manifold is divided into MLPs by measuring the intersection angles between neighborhood normal vectors by Cosine theorem.

In a practical project, the disordered high-sample image set is the object that the manifold learning algorithm deals with. Each image of the dataset corresponds to a point in the high dimensional image space. Since there are regularities when an object changes gradually, the points that correspond to images lie on a manifold in the high dimensional image space. Therefore, the treated object is the scattered point cloud data. Lots of algorithms have been proposed for estimating the curvature of the scattered point cloud data^[21]. Our method is listed below:

1) Constructing the neighborhood graph for each data point by k -NN or ε -ball. Since each graph is “small” enough, it could be regarded as a local tangent space of the point.

2) Utilizing PCA to find the eigenvalues and eigenvectors of each neighborhood graph. Supposing P is a point on the manifold and its k -neighborhood is $X_p = [x_1, x_2, \dots, x_k]$,

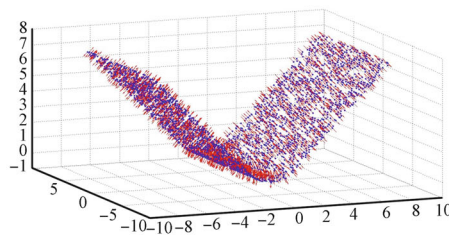
we have

$$\begin{aligned} (X_P - \bar{x}e^T) \times (X_P - \bar{x}e^T)^T &= Q\Lambda Q^T \\ \bar{x} &= X_P \times \frac{e}{N}, \quad Q = [v_1, v_2, \dots, v_n] \\ \Lambda &= \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}, \quad e = \underbrace{[1, 1, \dots, 1]^T}_k \end{aligned} \quad (1)$$

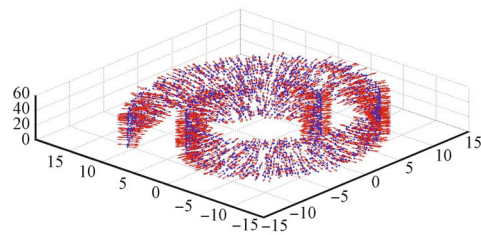
where $(X - \bar{x}e^T)$ is the centralized matrix, $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues with decent order, and v_1, v_2, \dots, v_n are the corresponding eigenvectors. Especially, n is the dimensionality of the original input space. Since the covariance matrix $(X_P - \bar{x}e^T) \times (X_P - \bar{x}e^T)^T$ is symmetric, all the eigenvalues are positive real numbers and eigenvectors are pairwise orthogonal. In this case, the eigenvalue λ_i depicts the scatter range of points that are in the neighborhood graph along v_i direction. Suppose d is the intrinsic dimensionality of the manifold. Therefore, v_1, v_2, \dots, v_d can be regard as a batch of coordinate basis vectors of tangent space of P and v_{d+1} is the normal vector of point P . Hence, the normal vector field can be obtained.

3) Estimate curvature scalar field of the manifold. It is visible in Fig. 2 that the curvature at point P can be presented by

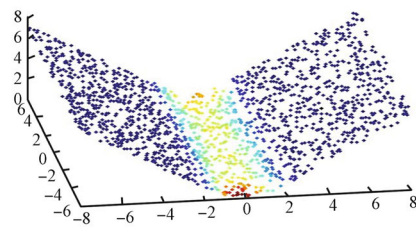
$$c_k(p) = \frac{\lambda_{d+1}}{\lambda_1 + \lambda_2 + \dots + \lambda_{d+1}}. \quad (2)$$



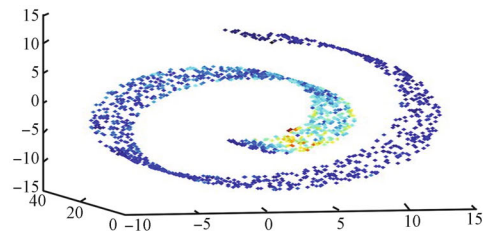
(a) Normal vector field of V-like shape data



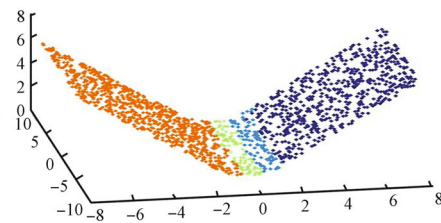
(b) Normal vector field of swiss-roll data



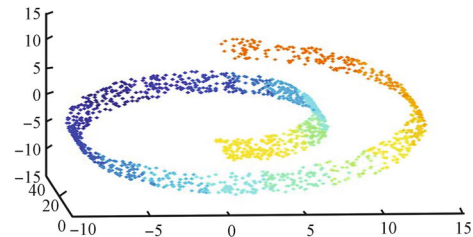
(c) Scalar curvature field of V-like shape data



(d) Scalar curvature field of Swiss-roll data



(e) The result of charting a V-like shape data



(f) The result of charting a Swiss-roll data

Fig. 2 Manifold charting

4) Chart the manifold. MLPs have two constrain conditions: 1) Each patch should be approximated to a linear space. 2) The patch should be the largest open set under a certain curvature range. In order to meet the two conditions above, firstly, a random point is selected as the center of the first patch. Then, we calculate the intersect angels between the vector of the center point and other normal vectors by Cosine theorem:

$$\cos(\theta) = \frac{nv_c \times nv_r^T}{nv_c \times nv_c^T + nv_r \times nv_r^T}, \quad r = 1, 2, \dots, m \quad (3)$$

$$Patch_i = \{x | \theta_{c,x} < \theta_t | \theta_{c,x} > \pi - \theta_t\}. \quad (4)$$

For each patch, the intersection angel θ is constrained to a constant curvature range that we set in advance, where θ_t is a threshold angel which can be adjusted by requirements in different cases. The smaller θ_t is, the more charts are divided. And each chart is more similar to a linear space. In other words, the manifold is charted when θ_t is small. On the contrary, the manifold is charted roughly when θ_t has a big value.

5) Finally, a new center point is selected and a new patch is divided by iterating the second step. The loop keeps running until all the data points are distributed into different patches. It is worth noticing that we must guarantee there

is an intersection area among adjacent patches. Two classical synthetic data sets, V-like shape data and Swiss-roll data, are listed in Fig. 2 to show the effectiveness of our method.

In Figs. 2(a) and 2(b), the blue points stand for the dataset and red rows show the normal vector field of the manifold that consists of cloud point data. In Figs. 2(c) and 2(d), the color depicts how the manifold is twisted. The points with hot color stands for the high curvature. On the contrary, the cool color points stand for the flat area. In Figs. 2(e) and 2(f), the final results of manifold charting are shown. Different patches are labeled with different colors. It is obvious that each patch holds the maximal linear structure. In Figs. 2(e), the V-like shape data are mainly consisted of two linear patches. In Fig. 2(f), since the curvature of inner circle is larger than the outer one's, the patches at the inner circle is smaller than the ones at the outer circle.

Remark 1. The intrinsic dimensionality of the manifold is a critical problem in manifold learning. Intuitively, we always observe a d -dimensional manifold in a $d+1$ -dimensional surrounding Euclidean space which is called ambient space in mathematics. In real world data, the dimensionality of the manifold corresponds to the variation degree of freedom of the object. Similar to the classical methods^[1,11], the true dimensionality of the manifold can be estimated by the residual variance of PCA as applied to the neighborhood graph constructed in Step 1.

Generally speaking, the dimensionality of the input space is extremely larger than that of the ambient space. In our experiments, PCA is utilized to reduce the dimensionality of the input dataset to ambient space before manifold charting. The delightful fact is that the ambient space can be estimated after linear dimensionality reduction. It has two advantages. The first one is that the dimensionality of the ambient space is theoretically one more than the manifold intrinsic dimensionality. It could ensure the correctness of the intrinsic dimensionality of the manifold that has been estimated. The second one is that the normal vectors of data points are unique in the ambient space.

Remark 2. The number of intersection data points of adjacent charts should be more than d (the intrinsic dimensionality of the manifold). d equations are needed to determine the parameters of $d \times (d+1)$ translation matrix.

5 Alignment

We assume that the manifold F is divided into n -th patches $\{P_1, P_2, \dots, P_n\}$ after Section 4. The following problem is how to align them in a global coordinate system. In manifold definition, there are intersection areas between adjacent patches. It is evident that the intersection data points from different patches should have the same coordinates in the global coordinate system. Therefore, in this paper, a global coordinate system is constructed by minimizing coordinate divergence of the intersection data points in the global coordinate system. The process is listed below.

Step 1. Calculating the local coordinates of each patch (homeomorphism mapping). Let $P_i = \{x_1^i, x_2^i, \dots, x_k^i\}$ be a matrix of data points in the i -th patch. The local coordinates satisfy

$$\min_{x, \theta, Q} \sum_{j=1}^k \|x_j^i - (\bar{x}^j + Q_i^T \theta_j)\|^2 = \min_{x, \theta, Q} \|P_i - (xe^T + Q_i^T \theta)\|^2 \quad (5)$$

where Q is the d left singular vectors of $P_i(I - \frac{1}{k}ee^T)$ corresponding to its d largest singular values, d is the intrinsic dimensionality of the manifold. The local coordinates of θ_i are

$$\Theta_i = Q_i P_i \left(I - \frac{1}{k} ee^T \right) = [\theta_1^i, \theta_2^i, \dots, \theta_k^i], \theta_j^i = Q_i (x_j^i - \bar{x}^i). \quad (6)$$

Step 2. Constructing object function and constraint condition. Suppose P^{ij} is the intersection point set of adjacent patches P_i and P_j . The point's local coordinates of these two patches are $\theta_{\text{int}}^i = \{\theta_1^i, \theta_2^i, \dots, \theta_m^i\}$ and $\theta_{\text{int}}^j = \{\theta_1^j, \theta_2^j, \dots, \theta_m^j\}$, m is the number of points in P^{ij} . However, in global coordinate system, intersection points should be coincided with

$$\begin{aligned} r_i \theta_{\text{int}}^i + l_i + E &= r_j \theta_{\text{int}}^j + l_j + E \Leftrightarrow \\ \Theta_{\text{int}}^i R_i - \Theta_{\text{int}}^j R_j &= E_c \Leftrightarrow \\ [\Theta_{\text{int}}^i, -\Theta_{\text{int}}^j] \begin{bmatrix} R_i \\ R_j \end{bmatrix} &= E_c \Leftrightarrow \\ E_{\text{eu}} &= E_c^T \times E_c \\ \Theta_{\text{int}}^i &= \begin{bmatrix} (\theta_1^i)^T & 1 \\ (\theta_2^i)^T & 1 \\ \dots & \dots \\ (\theta_m^i)^T & 1 \end{bmatrix} \\ R_i &= \begin{bmatrix} r_i^T \\ l_i^T \end{bmatrix} \end{aligned} \quad (7)$$

where r_i is a rotation matrix and l_i is a offset matrix for P_i , E_c is coordinate components error of the intersection points of p^{ij} , and E_{eu} is the summation of Euclidean distance errors of intersection points. Therefore, to obtain an optimum global coordinate system, we seek to find proper R_i to minimize the summation of Euclidean distance errors of all the intersection points that are from adjacent patches.

$$E = \begin{bmatrix} E_c^1 \\ E_c^2 \\ \dots \\ E_c^n \end{bmatrix} = \begin{bmatrix} X_{\text{int}}^{1,2} \\ \dots \\ X_{\text{int}}^{1,n} \\ X_{\text{int}}^{2,2} \\ \dots \\ X_{\text{int}}^{2,n} \\ \dots \\ X_{\text{int}}^{n,n} \end{bmatrix} \times \begin{bmatrix} R_1 \\ R_2 \\ \dots \\ R_n \end{bmatrix} = XR$$

$$e = \sum_{i=1}^n E_{eu}^i = E^T \times E = R^T X^T X R$$

$$X_{int}^{i,j} = \begin{bmatrix} \underbrace{0, 0, \dots, 0}_{(i-1) \times (d+1)}, \underbrace{\Theta_{int}^i}_{(j-i-1) \times (d+1)}, \underbrace{0, 0, \dots, 0}_{(n-j-1) \times (d+1)}, -\underbrace{\Theta_{int}^j}_{(n-j-1) \times (d+1)}, \underbrace{0, 0, \dots, 0}_{(n-j-1) \times (d+1)} \end{bmatrix} \quad (8)$$

where Θ_{int}^i and Θ_{int}^j are changed simultaneously at either a different i or j .

Suppose the global coordinates of data points we will construct are $T = \{t_1, t_2, \dots, t_n\}$. To uniquely determine T , we impose a constrain condition $TT^T = I_d$. And for each patch, the global coordinate coordinates is $T_i = \{t_1^i, t_2^i, \dots, t_m^i\}$. The relationship between global coordinates and local coordinates are $T_i = R_i^T \Theta_i$, where Θ_i is the local coordinates of data points in patch i that are obtained from (6).

$$T = \begin{bmatrix} R_1 \\ R_2 \\ \dots \\ R_n \end{bmatrix} \times \begin{bmatrix} \Theta_1 & \Theta_2 & \dots & \Theta_n \\ 1 & 1 & \dots & 1 \end{bmatrix} \Rightarrow TT^T = R^T \Theta \Theta^T R. \quad (9)$$

Therefore, in our paper, the cost function ξ and constraint condition C are

$$\xi : \text{mine} = R^T X^T X R \quad (10)$$

$$C : R^T \Theta \Theta^T R = I. \quad (11)$$

It is worth pointing out that the cost function is convex. Therefore, the solution to this problem is given by the smallest generalized eigenvectors v with $(X^T X)v = \lambda(\Theta \Theta^T)v$. The optimal R is given by the d eigenvectors corresponding to the 2nd to $(d+1)$ -th smallest generalized eigenvectors.

Remark 3. The relationship among homeomorphic mapping, affine mapping and translation mapping is illustrated in Fig. 3. The homeomorphic mapping ψ is obtained from (6), which maps the patch into the n -dimensional Euclidean space. $P_i \psi_i$ is a chart in the manifold definition. R_i is the optimal affine mapping that maps the i -th coordinate system into the global coordinate system. The translation mappings between the i -th and j -th patches in local coordinate systems and global coordinate system are $\psi_i \circ \psi_j^{-1}$ and $(\psi_i R_i) \circ (\psi_j R_j)^{-1}$, respectively. It is worth noticing that all maps are invertible.

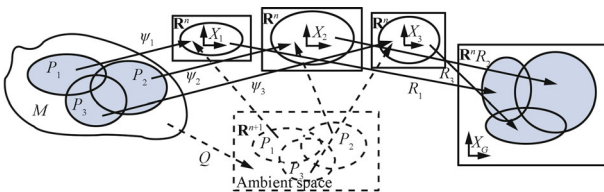


Fig. 3 The relationship between homeomorphic mapping, affine mapping and translation mapping

Remark 4. The intersection points that belong to different patches have different global coordinates. In this paper, we take their average value as their global coordinates.

Remark 5. It could map the input space to the ambient space firstly by a linear mapping Q , then calculate the homeomorphic mapping and affine mapping in this space.

6 Experimental results

In this section, the experiments are divided into three parts to show the effectiveness of the ACT. Firstly, the noise synthetic data and fragment synthetic data are utilized to show the effect of the ACT in preprocessing of manifold learning. Secondly, we test the ACT on six synthetic data sets. At last, two face data sets are utilized to test the effect of the ACT on dimension reduction and data reconstruction. Some of the compared results are generated by the Mani Matlab demo^[22].

6.1 Experiment on noise data and fragment data

The characteristics of ACT are illustrated for dealing with noise data and fragment data, respectively. As mentioned above, the noise data and fragment data may cause unwilling results for dimensional reduction. “V” data and “Swiss-roll” data with noise and fragment are utilized to show this in Figs. 4 and 5.

For generating noise data, ten random points in synthetic data set are additive white Gaussian noise in range of $[-3, 3]$ at z -axis direction. It can be seen from Fig. 4 (a) that the three methods (MDS, Isomap, and Diffusion map) yield good results and the other methods give incorrect results. In Fig. 4 (b), it is obvious that these methods produce an satisfactory results. Moreover, the Hessian LLE could not get results in these two datasets when the parameter $KNN = 8$. Moreover, in other experiments, different types of noise are added to synthetic data sets and the dimensionality reduction results are poor and unpredictable in most cases. Therefore, identifying and removing noise data is essential in preprocessing.

Since the curvature of local area that contains noise point jumps much higher than that of the neighborhood area, the noise data can be identified by curvature with sudden change. In this paper, the curvature of data sets can be estimated in Section 4.2 which could locate the noise points roughly. In Figs. 4 (c) and 4 (d), the 200 and 400 largest curvature value points are marked respectively. It can be seen that the noise points are pointed out entirely in Fig. 4 (c) and the main noise points are marked in Fig. 4 (d). Since the curvature is estimated in a local area by the ratio of eigenvalue, the neighborhood data points have the same curvature value with the noise point. Therefore, these data points are also marked in Fig. 4. It is worth noticing that these neighborhood points can be removed by subsequent processes.

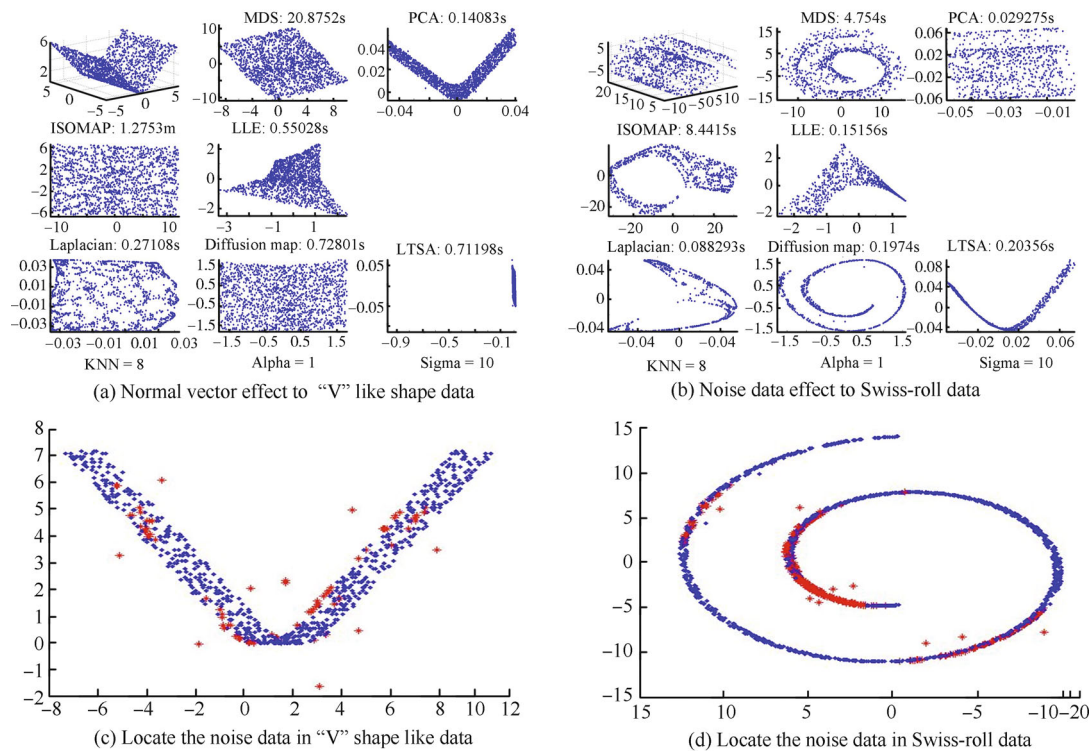


Fig. 4 Dimensionality reduction for noise data

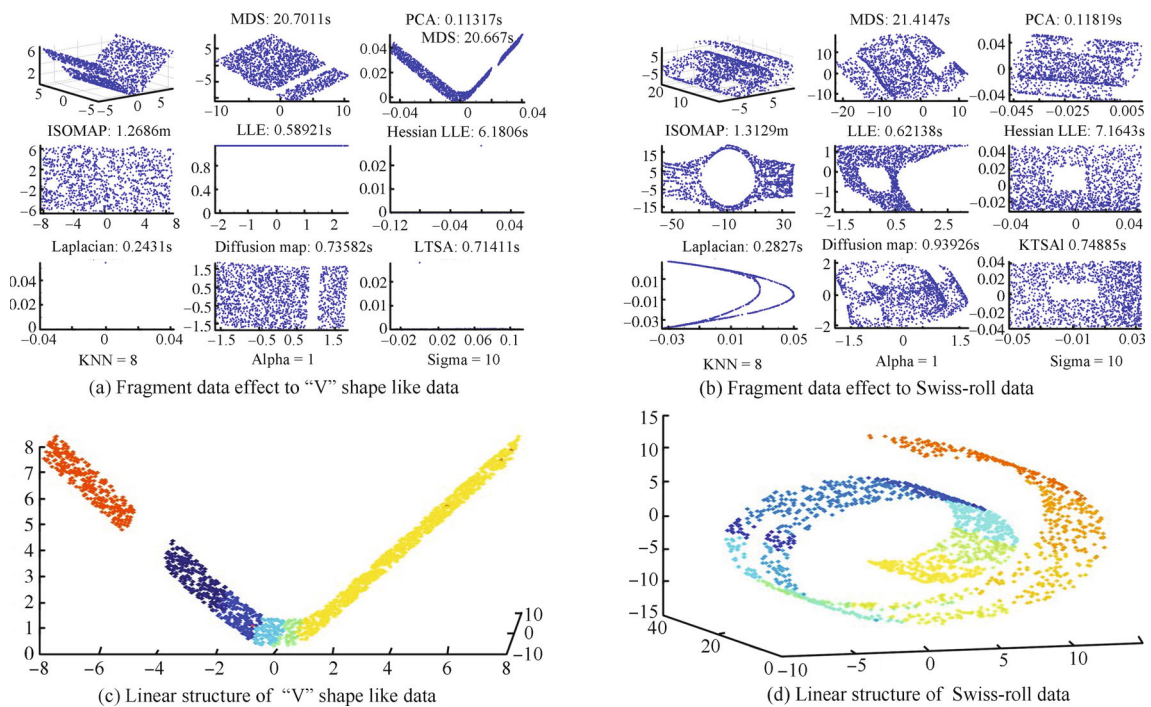


Fig. 5 Dimensionality reduction for fragment data

For fragment data, we remove some data points from the synthetic data sets. It can be seen from Fig. 5 that the two methods (MDS and Diffusion map) generate satisfactory results in Fig. 5 (a) and another two methods (Hessian

LLE and LTSA) produce good results in Fig. 5 (b). Same as noise data, there is no method that could get satisfactory results for all fragment data sets. Therefore, how to fill up the missing data is also an important problem in preprocessing.

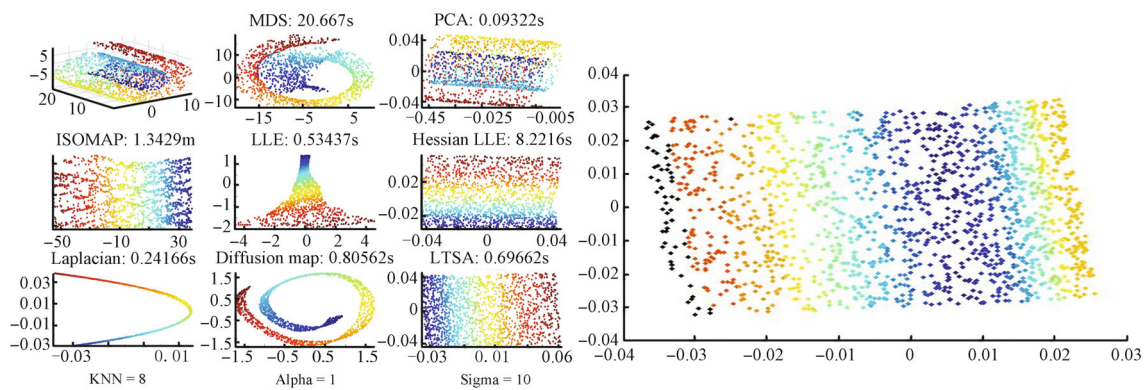
In this paper, after dividing the manifold into maximal linear patches, the distribution of missing data points can be estimated by the neighborhood patches. We select the parameter equal to 40 degrees as utilized in Section 4.3 and it can be seen that the manifolds are properly charted. Since the coordinate basis of each patch is calculated by (5), the basis vectors of the area that miss data point distribution can be simply estimated by linear combination.

6.2 Experiment on synthetic data

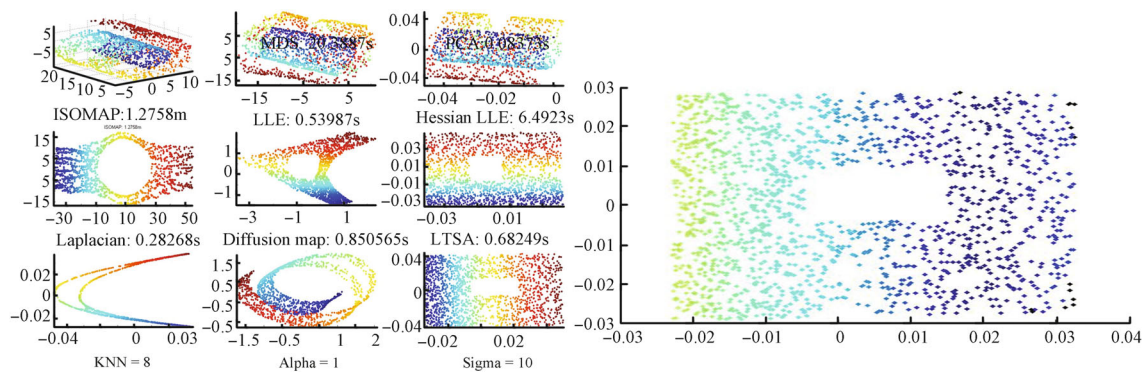
To evaluate the performance of our ACT, we test it on five classical synthetic data sets and compare the results with eight classical dimensionality reduction meth-

ods. In our experiments, each synthetic datasets consists of 1600 data points. The number of nearest neighborhoods $KNN = 8$ and target dimension $d = 2$. In ACT, the parameter in (4) is set to 30 degrees. The efficient method could preserve the intrinsic geometrical structure of the manifold after dimensional reduction.

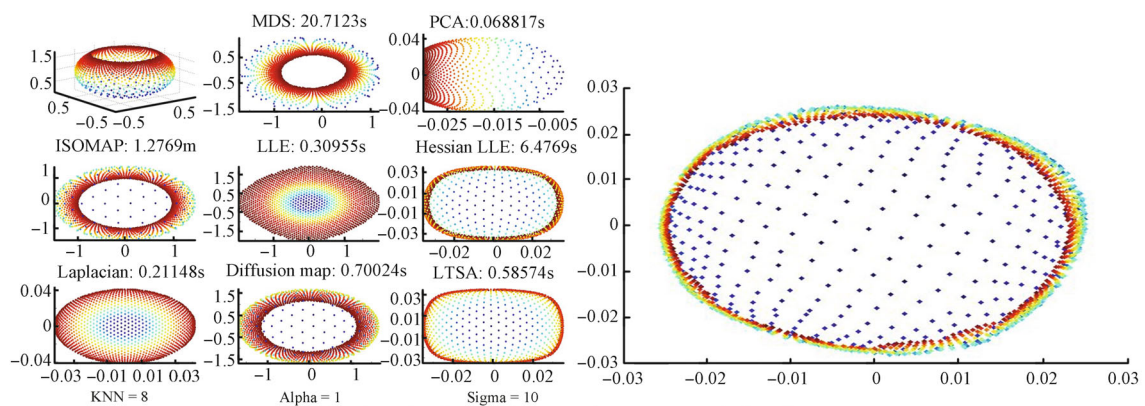
It can be seen from Fig. 6 that the ACT could produce correct results except punctured sphere data, which performs as well as the LSTA. Moreover, it could depict the linear structure of the manifold when the manifold has visible linear structure, which is shown with different colors in Figs. 6 (a)–(c). PCA and MDS just suit to deal with linear datasets and cannot unfold manifold with nonlinear structure. Other methods are unstable on different datasets.



(a) Dimensionality reduction comparison on swiss-roll dataset



(b) Dimensionality reduction comparison on Swiss hole dataset



(c) Dimensionality reduction comparison on punctured sphere dataset

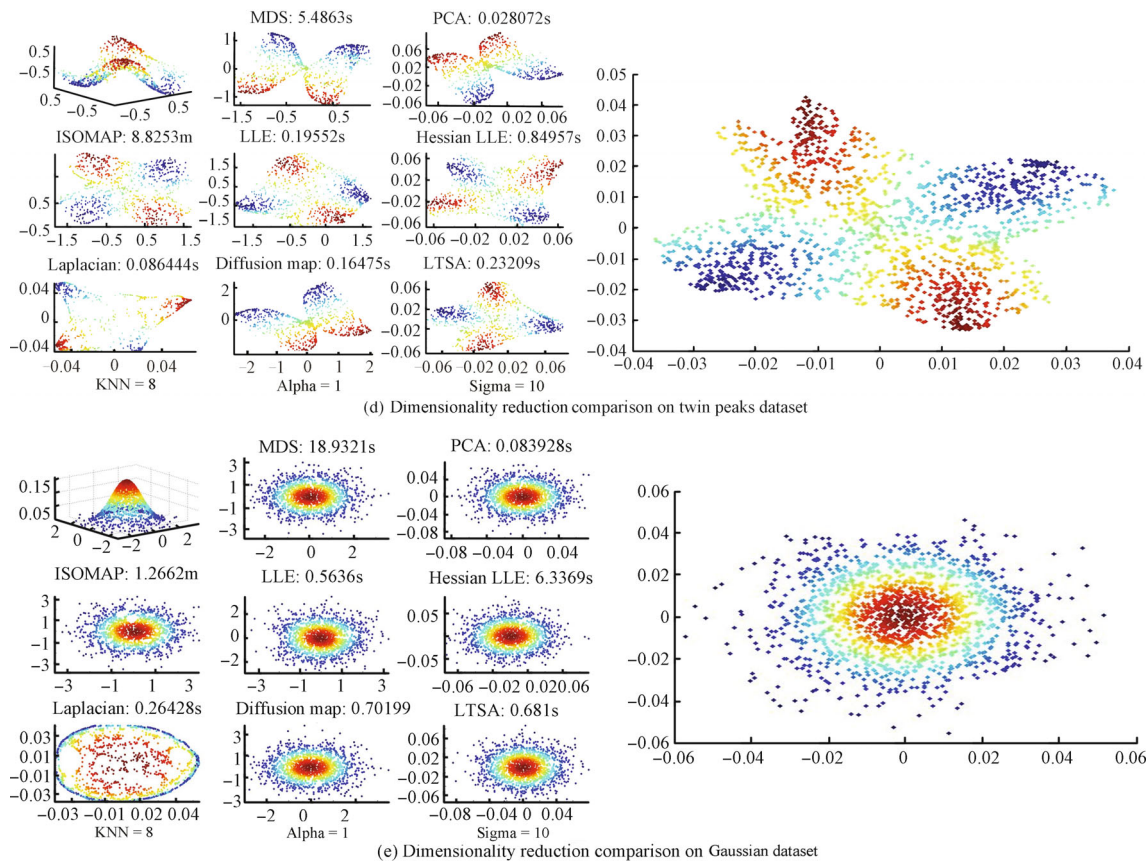


Fig. 6 Results comparison on synthetic data sets

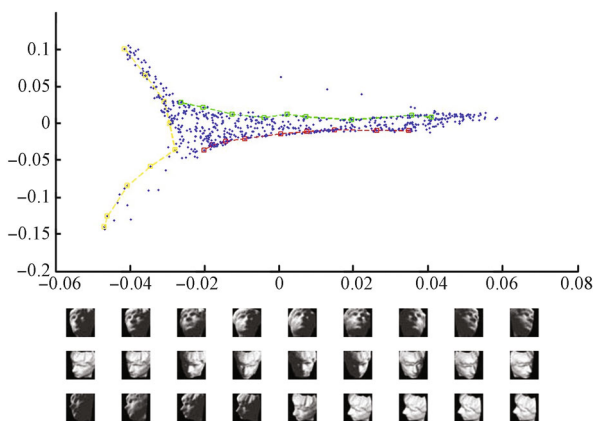


Fig. 7 Two-dimensional embedding of ACT of Isomap face data using ACT

To evaluate our ACT algorithm on image datasets, two data sets (ISOMAP face data and LLE face data) are used to perform dimensionality reduction. The following figures show results in detail.

Isomap face image set consists of 698 pixel images, in which a 3D face model changes with pose and lighting condition. Each image is converted to an $m = 4096$ dimensional image vector. We apply the ACT with $degree = 40$ and $d = 2$. The result of dimension reduction is shown in Fig. 7. Three paths along the boundaries are extracted. In

Fig. 6, the three image rows from top to bottom correspond to lines 1, 2 and 3, respectively. It can be seen that the pose and lighting variations can be depicted by the low-dimensional coordinate system.

The LLE face image set consists of 1965 pixel images that are 28×20 , in which a single person strikes a variety of pose and expressions. Each image is converted to an $m = 560$ dimensional image vector. We apply the ACT with $degree = 40$ and $d = 2$. The result of dimension reduction is shown in Fig. 8. Two patches are extracted which correspond to two image rows below Fig. 8. It can be seen that the horizon axis depicts the variation of face rotation and the vertical axis reflects the expression changes.

7 Conclusions and future work

A manifold learning method (atlas compatibility transformation) is proposed. Firstly, the manifold is divided into maximal linear charts based on normal vector field. Then, charts are aligned into a global coordinate system by minimizing the error of intersection areas among patches. Our method can preserve both local geometry and global structure of the manifold. Compared with previous manifold learning algorithms, the ACT could deal with noise datasets and fragment datasets. Moreover, it more corresponds to the manifold definition in differential geometry. And the incremental-learning problem is solved by given invertible

maps between high input space and low embedding space. Experimental results indicate that the ACT could get excellent performance on both synthetic and real data sets.

It is worth pointing out that intersection area between charts has great impact on the dimensional reduction results. Hence, how to select intersection area is interesting in future work. Moreover, in 1956, Stein showed that Euclidean distance is not the best measurement manner in a high-dimensional space. Therefore, adopting different distance is another research direction.

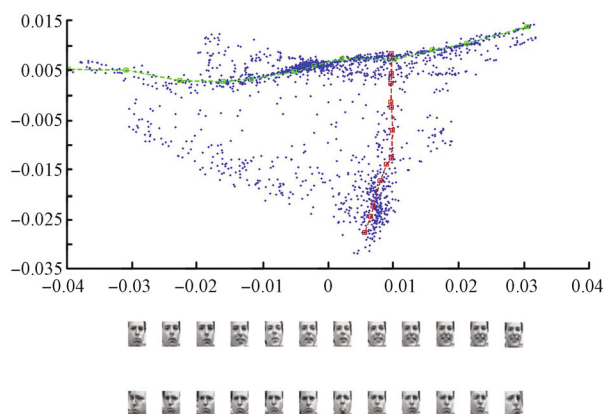


Fig. 8 Two-dimensional embedding of LLE face data using ACT

References

- [1] J. Tenenbaum, V. De Silva, J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [2] S. Roweis, L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [3] M. Belkin, P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems 14*, pp. 585–591, 2002.
- [4] M. Brand. Charting a manifold. *Advances in Neural Information Processing Systems*, vol. 15, pp. 961–968, 2003.
- [5] D. L. Donoho, C. Grimes. Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data. In *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 10, pp. 5591–5596, 2003.
- [6] Y. Teh, S. Roweis. Automatic alignment of hidden representations. *Vancouver, Canada, Neural Information Processing Systems 15*, pp. 841–848, 2003.
- [7] Z. Zhang, H. Zha. Principle manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal on Scientific Computing*, vol. 26, no. 1, pp. 313–338, 2004.
- [8] Z. Zhang, J. Wang, H. Zha. Adaptive manifold learning. *Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 253–365, 2012.
- [9] T. Lin, H. Zha. Riemannian manifold learning. *Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 796–809, 2008.
- [10] R. Wang, S. Shan, X. Chen, J. Chen, W. Gao. Maximal linear embedding for dimensionality reduction. *Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1776–1792, 2011.
- [11] O. Tuzel, F. Porikli, P. Pedestrian detection via classification on Riemannian manifolds. *Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008.
- [12] A. Elgammal, C. Lee. Tracking people on a torus. *Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 520–538, 2009.
- [13] H. Yin. Nonlinear dimensionality reduction and data visualization: A review. *International Journal of Automation and Computing*, vol. 4, no. 3, pp. 294–303, 2007.
- [14] G. Strang. *Introduction to Linear Algebra*, Cambridge, UK: Wellesley-Cambridge Press, 2009.
- [15] T. Kohonen. *Self-organizing Maps*, Berlin, Germany: Springer-Verlag, 2001.
- [16] J. S. Wang, X. D. Ren. GLCM based extraction of flame image texture features and KPCA-GLVQ recognition method for rotary kiln combustion working conditions. *International Journal of Automation and Computing*, vol. 11, no. 1, pp. 72–77, 2014.
- [17] J. Ham, D. Lee, S. Mika, B. Schölkopf. A kernel view of the dimensionality reduction of manifolds. In *Proceedings of the 21st International Conference on Machine Learning*, Banff, Alberta, Canada, pp. 47–54, 2004.
- [18] M. Balasubramanian, E. L. Schwartz. The isomap algorithm and topological stability. *Science*, vol. 295, no. 5552, pp. 7, 2002.
- [19] M. H. Law, A. K. Jain. Incremental nonlinear dimensionality reduction by manifold learning. *Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 377–391, 2006.
- [20] R. R. Coifman, S. Lafon. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 31–52, 2006.
- [21] L. H. Wang, B. Z. Yuan. Curvature and density based feature point detection for point cloud data. In *Proceedings of the 3rd IET International Conference on Wireless, Mobile and Multimedia Networks*, IEEE, Beijing, China, pp. 377–380, 2010.
- [22] T. Wittman. *Mani Matlab Demo*, [Online], Available: <http://www.math.umn.edu/wittman/mani/>, 2005.

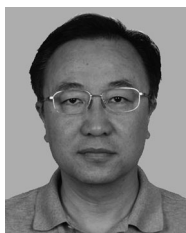


Zhong-Hua Hao received the B.Sc. degree in automation from Three Gorges University, China in 2007. He received the M.Sc. degree in control theory and control engineering from Kunming University of Science and Technology, China in 2000. Now, he is a Ph.D. candidate in Shanghai University, China.

His research interests include digital image processing, data mining, machine learning and pattern recognition.

E-mail: haozhonghua85@gmail.com

ORCID iD: 0000-0002-8309-0333



Shi-Wei Ma received B.Sc. and M.Sc. degrees in electronics from Lanzhou University, China in 1986 and 1991, respectively, and received the Ph.D. degree in control theory and engineering from Shanghai University, China in 2000. From 2001 to 2003, he was a Japan Science and Technology research fellow at the National Institute of Industrial Safety of Japan. From 2003 to 2008, he was an associate professor. Since 2008, he has been a professor, both in the Department of Automation in Shanghai University, China.

His research interests include signal processing, pattern recog-

nition and intelligent system.

E-mail: masw@shu.edu.cn (Corresponding author)

ORCID iD: 0000-0001-6039-5030



Fan Zhao received the B.Sc. degree in information engineering from Northwestern Polytechnical University, China in 2012. She is a master student in control science and engineering at Shanghai University, China.

Her research interests include digital image processing and machine learning.

E-mail: zaofun@shu.edu.cn