Riemannian Manifold Learning for Nonlinear Dimensionality Reduction

Tony Lin[1,⋆], Hongbin Zha[1], and Sang Uk Lee[2]

[1] National Laboratory on Machine Perception,
Peking University, Beijing 100871, China
{lintong, zha}@cis.pku.edu.cn
[2] School of Electrical Engineering,
Seoul National University, Seoul 151-742, Korea
sanguk@ipl.snu.ac.kr

Abstract. In recent years, nonlinear dimensionality reduction (NLDR) techniques have attracted much attention in visual perception and many other areas of science. We propose an efficient algorithm called Riemannian manifold learning (RML). A Riemannian manifold can be constructed in the form of a simplicial complex, and thus its intrinsic dimension can be reliably estimated. Then the NLDR problem is solved by constructing Riemannian normal coordinates (RNC). Experimental results demonstrate that our algorithm can learn the data's intrinsic geometric structure, yielding uniformly distributed and well organized low-dimensional embedding data.

1 Introduction

In visual perception, a human face image of size of 64×64 pixels is often represented by a vector in a 4096-dimensional space. Obviously, the 4096-dimensional vector space is too large to allow any efficient image processing. A typical way to avoid this "curse of dimensionality" problem [1] is to use dimensionality reduction techniques. Classical linear methods, such as Principal Component Analysis (PCA) [2] and Multidimensional Scaling (MDS) [3], can only see flat Euclidean structures, and fail to discover the curved and nonlinear structures of the input data. Previous nonlinear extensions of PCA and MDS, including Autoencoder Neural Networks [4], SOM [5], Elastic Nets [6], GTM [7], and Principal Curves [8], suffer from the difficulties in designing cost functions and training too many free parameters, or are limited in low-dimensional data sets. In recent years some nonlinear manifold learning techniques have been developed, such as Isomap [9, 10], LLE [11], Laplacian Eigenmaps [12, 13], Hessian Eigenmaps [14], SDE [15], manifold charting [16], LTSA [17], diffusion maps [18]. Due to their nonlinear nature, geometric intuition and computational practicability, these nonlinear manifold learning techniques have attracted extensive attention

⋆ This work was supported by NSFC (60302005), NSFC (60333010), NKBRP (2004CB318005) and FANEDD (200038), China.

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part I, LNCS 3951, pp. 44–55, 2006.
© Springer-Verlag Berlin Heidelberg 2006

of the researchers from different disciplines. The basic assumption is that the input data lie on or close to a smooth low-dimensional manifold [19].

Each manifold learning algorithm attempts to preserve a different geometrical property of the underlying manifold. Local approaches (e.g. LLE [11], Laplacian Eigenmaps [12], LTSA [17]) aim to preserve the local geometry of the data. They are also called spectral methods, since the low dimensional embedding task is reduced to solving a sparse eigenvalue problem under the unit covariance constraint. However, due to this imposed constraint, the aspect ratio is lost and the global shape of the embedding data can not reflect the underlying manifold. In contrast, global approaches like Isomap [9] attempt to preserve metrics at all scales and therefore give a more faithful embedding. However, Isomap, or isometric mapping, can be only applied to intrinsically flat manifolds, e.g. 2D developable surfaces (cylinders, cones, and tangent surfaces). Conformal mapping [10, 20] appears to be a promising direction.

We propose a general framework called Riemannian manifold learning (RML). The problem is formulated as constructing local coordinate charts for a Riemannian manifold. The most widely used is the Riemannian normal coordinates (RNC) chart. In [21] Brun et al. presented a method for manifold learning directly based on the concept of RNC. In order to calculate the geodesic directions, high sampling density is required and the second order polynomial interpolation is computationally expensive. In this paper, we propose a more efficient method to calculate RNC. The basic idea is to preserve geodesic distances and directions only in a local neighborhood. We also describe a novel method for estimating intrinsic dimension of a Riemannian manifold. Our method is derived by reconstructing the manifold in the form of an simplicial complex, whose dimension is determined as the maximal dimension of its simplices.

2 Mathematical Preliminaries

In this section we briefly review some basic concepts of Riemannian geometry [22]. A bijective map is called a homeomorphism if it is continuous in both directions. A (topological) manifold M of dimension m is a Hausdorff space for which every point has a neighborhood U homeomorphic to an open set V of R^m with $\phi : U \to V \subset R^m$. (U, ϕ) is called a local coordinate chart. An atlas for M means a collection of charts $\{(U_\alpha, \phi_\alpha)|\alpha \in J\}$ such that $\{U_\alpha|\alpha \in J\}$ is an open cover of M. A manifold M is called a differential manifold if there is an atlas of M, $\{(U_\alpha, \phi_\alpha)|\alpha \in J\}$, such that for any $\alpha, \beta \in J$, the composite $\phi_\alpha \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \to R^m$ is differentiable of class C^∞. A differential manifold M endowed with a smooth inner product (called Riemannian metric) $g(u, v)$ or $\langle u, v\rangle$ on each tangent space T_pM is called a Riemannian manifold (M, g).

An exponential map $exp_p(v)$ is a transform from a tangent vector $v \in T_pM$ into a point $q \in \gamma \subset M$ such that $dist(p, q) = ||v|| = \langle v, v\rangle^{1/2}$, where γ is the unique geodesic traveling through p such that its tangent vector at p is v. A geodesic is a smooth curve which locally join their points along the shortest path.

All the geodesics passing through p are called radial geodesics. The local coordinates defined by the chart (U, exp_p^{-1}) are called Riemannian Normal Coordinates (RNC) with center p. Note that the RNC mapping preserves the distances on radial geodesics. A simple example is paring an orange, which maps a sphere onto a plane, while the distances on the great circles of the sphere are preserved.

3 Manifold Assumption

Most manifold learning algorithms [9, 11, 19] assume that a set of image data may generate a low-dimensional manifold in a high-dimensional image space. Here we present a simple geometric imaging model (shown in Fig. 1) for human face images to clarify this assumption. Varying poses and lighting conditions are considered in this model, as they are two important factors in face detection and recognition. The model may be adapted to image data of other objects (e.g. cars), if similar imaging conditions are encountered.

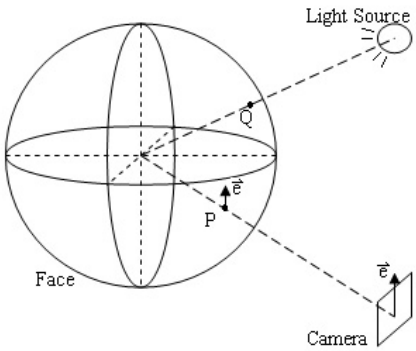

Fig. 1. A geometric imaging model for human face images

We model the head of a human as a unit sphere S^2, where the frontal hemisphere is the human face. Different poses are obtained by moving the camera, as the human face is kept in stationary. The focal length is assumed to be unchanged in the imaging process. We also assume that the distance from the camera to the face is fixed, so the face images have similar scales. Commonly, the center axis of the camera is set to passing through the center of the sphere. The camera is allowed to have some degree of planar rotations. The lighting is simply modeled with a point light source far away from the sphere. Under these variations, this set of face images generates a 5-dimensional manifold, which is homeomorphic to

$$M = \{PQ\boldsymbol{e} | P \in S^2, Q \in S^2, \boldsymbol{e} \in S^1\},$$

where P and Q are two intersection points on S^2 by the center axis of the camera and the lighting ray, and $\boldsymbol{e}$ is a unit vector to show the planar rotation angle

of the camera. This representation is just a simple extension of the parametric representation of a surface, $\boldsymbol{r} = \boldsymbol{r}(u, v)$, where (u, v) are two varying parameters. If the lighting variation is ignored, a 3-dimensional manifold may be generated:

$$M' = \{P\boldsymbol{e} | P \in S^2, \boldsymbol{e} \in S^1\}.$$

This is called a circle bundle on a sphere, which is one of the simplest tangent bundles. This manifold can be visualized as the earth running in its circular orbit in the 4-dimensional space-time.

4 Manifold Reconstruction

The m-manifold M generated from a set of data points in R^n is modeled with an approximating simplicial complex, whose dimension serves as a reliable estimation of the intrinsic dimension of M. Our manifold reconstruction is a simplified version of Freedman's method [23], which involves a computationally expensive optimization for convex hulls.

The key to the reconstruction problem from unstructured sample points is to recover the edge connections within a local neighborhood. The neighborhood of one point $p \in M$, denoted $NBD(p)$, is defined as the K nearest points to p. K is often set as $c \times m'$, where c is a constant number between 2 and 5, and m' is an initial estimation of the intrinsic dimension m. Then we select k $(1 \leq k \leq K)$ edge points from the K neighbors, such that the edge connections are built between p and each edge point. Note that the number of edge points, k, is varying with p. A point q is said to be an edge point of p if no other point r separates p and q by the normal plane passing through r and perpendicular to the line (p, r). Formally, the edge point set of point p is defined as

$$EP(p) = \{q \in NBD(p) \mid \langle p - r, q - r \rangle \geq 0, \text{ any } r \in NBD(p)\}.$$

It is easy to show that by this definition, the angle between any two adjacent edges is acute or right, while obtuse angles are prohibited. This property guarantees to yield well-shaped simplices, which are basic building blocks to construct the target simplicial complex. The underlying reason for this property is explained by a simple example shown in Fig. 2. It is often believed that the 1D reconstruction in (b) is much better than the 2D reconstruction in (c). These points are more likely to be sampled from a 1D curve, rather than a 2D surface. The width of the 2D complex in (c) is too small and thus can be ignored. In fact, any thin rope in the physical world can be modeled as a 1D curve by ignoring its radius. This definition of an edge point permits edge connections like (b) while (c) is prohibited.

Simplices in each dimension are constructed by grouping adjacent edges. For example, if (p, q) is an edge and r is any other point, a triangle (p, q, r) is generated when there are two edges (p, r) and (q, r). This procedure repeats from low-dimensional to high-dimensional, until there are no new simplices generated. The target simplicial complex is composed of all the simplices. The dimension of the complex is a good estimate of the intrinsic dimension of M.

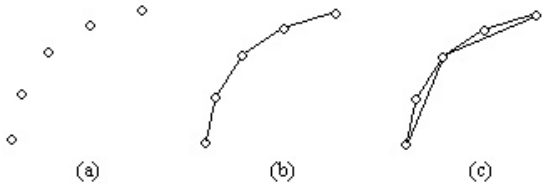

Fig. 2. Reconstruction of five points sampled from a curve. (a) Unorganized points. (b) 1D reconstruction. (c) 2D reconstruction.

5 Manifold Charting

Manifold charting consists of two steps: (1) Compute the tangent space and set up a Cartesian coordinate system; (2) Use Dijkstra's algorithm to find single-source shortest paths, and calculate the Riemannian Normal Coordinates (RNC) for each end point of the shortest paths.

In principle, the base point p for a RNC chart may be freely selected. Here we choose the base point close to the center of the input data. For each candidate point, the maximal geodesic distance (called geodesic radius) is computed using Dijkstra's algorithm. One point with the minimal geodesic radius is the optimal base point.

A local coordinate chart is set up by computing the tangent space T_pM:

$$x_0 + \text{span}\{x_1 - x_0, \ldots, x_m - x_0\},$$

where $\{x_0, x_1, \ldots, x_m\}$ are $(m+1)$ geometrically independent edge points (or nearest neighbors) of p. Any point on the tangent space can be represented as

$$x_0 + \sum_{i=1}^{m} \lambda_i (x_i - x_0).$$

An orthonormal frame, denoted $(p\ ; \boldsymbol{e_1}, \ldots, \boldsymbol{e_m})$, is computed from the vectors $\{x_1 - x_0, \ldots, x_m - x_0\}$ by using the Gram-Schmidt orthogonalization.

Then the Dijkstra's algorithm [24] is exploited to find single-source shortest paths in the graph determined by the simplicial complex. Each time a new shortest path is found, we compute the RNC of the end point on this path. If the end point q is an edge point of p, we directly compute the projection of q, denoted $q' \in R^m$, onto the tangent space frame $(p\ ; \boldsymbol{e_1}, \ldots, \boldsymbol{e_m})$ by solving the following least squares problem

$$\min_{X \in R^m} \|q - (p + \sum_{i=1}^{m} x_i \boldsymbol{e_i})\|^2,$$

where $X = (x_1, x_2, \ldots, x_m)$ are the projection coordinates of q' in the tangent space. The RNC of q is given by

$$\frac{\|q - p\|_{R^n}}{\|X\|_{R^m}} X\ ,$$

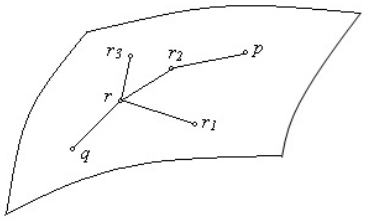

Fig. 3. An example illustrating how to compute the RNC of q. In this case, q is not an edge point of the base point p.

since the RNC preserves the distances on each radial geodesic, which is approximated by the corresponding shortest path.

If the end point $q \in M \subset R^n$ is not an edge point of p, the RNC of q (denoted q') is computed by solving a quadratically constrained linear least squares problem. Let point r be the previous point on the shortest path from p to q. Let $\{r_1, \ldots, r_k\}$ be the edge points (or nearest neighbors if needed) of r, whose RNCs have been computed. The number of these points, k, is required to be larger than or equal to m in order to guarantee the constrained least squares problem to be correctly solved. (One exception may occur at the beginning of the Dijkstra's algorithm, when k is less than m. In this case, point q is treated as an edge point of p to compute its RNC.) Fig. 3 shows such an example with $k = 3$. The basic idea is that we want to preserve the angles in the neighborhood of r, while keeping the geodesic distance from q to r unchanged. This leads to the following linear least squares problem

$$\cos\theta = \frac{\langle q - r, r_i - r\rangle}{\|q - r\| \cdot \|r_i - r\|} \approx \cos\theta' = \frac{\langle q' - r', r_i' - r'\rangle}{\|q' - r'\| \cdot \|r_i' - r'\|}, \ i = 1, 2, \ldots, k$$

with a quadratic constraint

$$\|q - r\| = \|q' - r'\|,$$

where q', r', and r_i' are the RNCs of q, r, and r_i. Our goal is to compute $q' \in R^m$.

We get the following linear least squares problem with quadratic constraints [25]:

$$\min_{x \in R^m} \|A_{k\times m} x_{m\times 1} - b_{k\times 1}\|^2 \text{ subject to } \|x_{m\times 1}\|^2 = \alpha^2 \ (k \geq m).$$

This problem can be solved by the following Lagrange multipliers optimization

$$\phi(x, \lambda) = \|b - Ax\|^2 + \lambda(\|x\|^2 - \alpha^2) = (b^T - x^T A^T)(b - Ax) + \lambda(x^T x - \alpha^2).$$

Setting the gradient of this function with respect to x (and not λ) equal to zero yields the equation

$$\frac{\partial \phi}{\partial x} = -2A^T b + 2A^T Ax + 2\lambda x = 0,$$

which has the solution

$$x = (A^T A + \lambda I)^{-1} A^T b$$

provided the inverse of $(A^T A + \lambda I)$ exists. Substituting this result into the constraint $\|x\|^2 = \alpha^2$, we have

$$\psi(\lambda) = b^T A (A^T A + \lambda I)^{-2} A^T b - \alpha^2 = 0.$$

Let $A = U \Sigma V^T$ be the singular value decomposition of A. Then our constraint equation becomes

$$\begin{aligned}\psi(\lambda) = 0 &= b^T U \Sigma V^T (V \Sigma^T U^T U \Sigma V^T + \lambda I)^{-2} V \Sigma^T U^T b - \alpha^2 \\ &= b^T U \Sigma V^T (V (\Sigma^T \Sigma + \lambda I) V^T)^{-2} V \Sigma^T U^T b - \alpha^2 \\ &= b^T U \Sigma V^T (V (\Sigma^T \Sigma + \lambda I) V^T V (\Sigma^T \Sigma + \lambda I) V^T)^{-1} V \Sigma^T U^T b - \alpha^2 \\ &= b^T U \Sigma (\Sigma^T \Sigma + \lambda I)^{-2} \Sigma^T U^T b - \alpha^2.\end{aligned}$$

Letting $\beta = U^T b$, we get

$$\psi(\lambda) = \sum_{i=1}^{m} \frac{\beta_i^2 \sigma_i^2}{(\sigma_i^2 + \lambda)^2} - \alpha^2 = 0.$$

It is easy to verify that $\psi(\lambda)$ decrease from ∞ to $-\alpha^2$ as λ goes from $-\sigma_m^2$ to ∞. We can use Newton's method to find the root λ. A good initial value for λ is zero, and the objective function vanishes to zero very fast.

Notice that the RNC of one data point can be efficiently computed in a local neighborhood, not involving any global optimization.

6 Experimental Results

First we test our dimension estimation method on four data sets [9, 11]: Swiss roll data, Isomap face data, LLE face data, and ORL face data. The number of the nearest neighbors, K, is set to 7, 8, 12, and 12, respectively. Table 1 shows the numbers of simplices in each dimension. Recall that the dimension of a complex is the maximal dimension of its simplices. For instance, the complex generated from Swiss roll data is composed of 1357 2D simplices, while no 3D simplices are contained in this complex. Therefore, the estimated dimension for Swiss roll

Table 1. Numbers of simplices in each dimension

Dim.	0	1	2	3	4	5	6	7	8	9	10
Swiss roll	1000	1800	1357	0							
Isomap	698	2337	5072	3782	751	0					
LLE	1965	6177	22082	40500	40384	19726	2820	0			
ORL	400	3011	11048	30602	91575	304923	932544	2261383	3674580	2835000	0

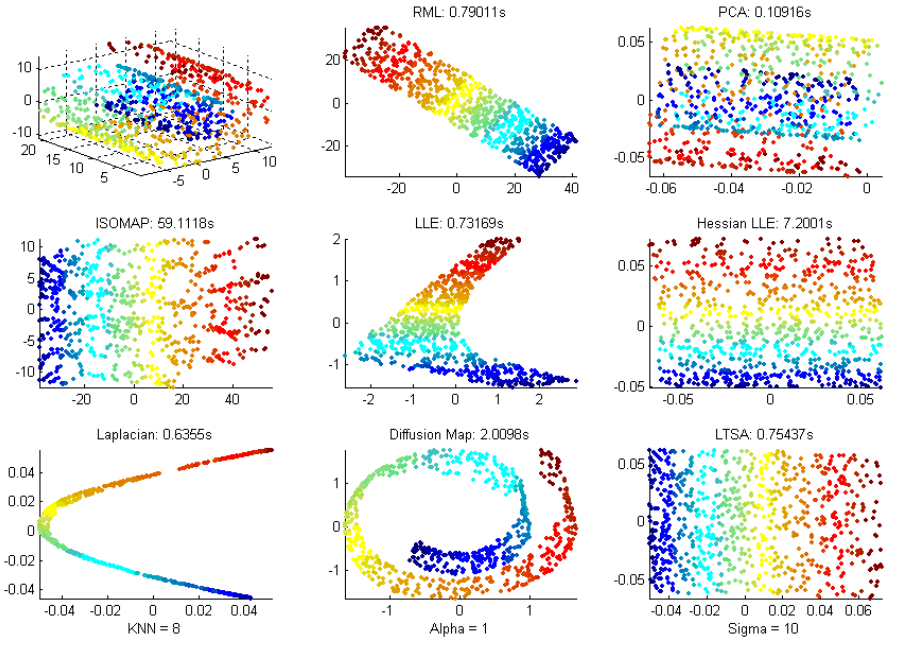

Fig. 4. Comparison results of Swiss Roll

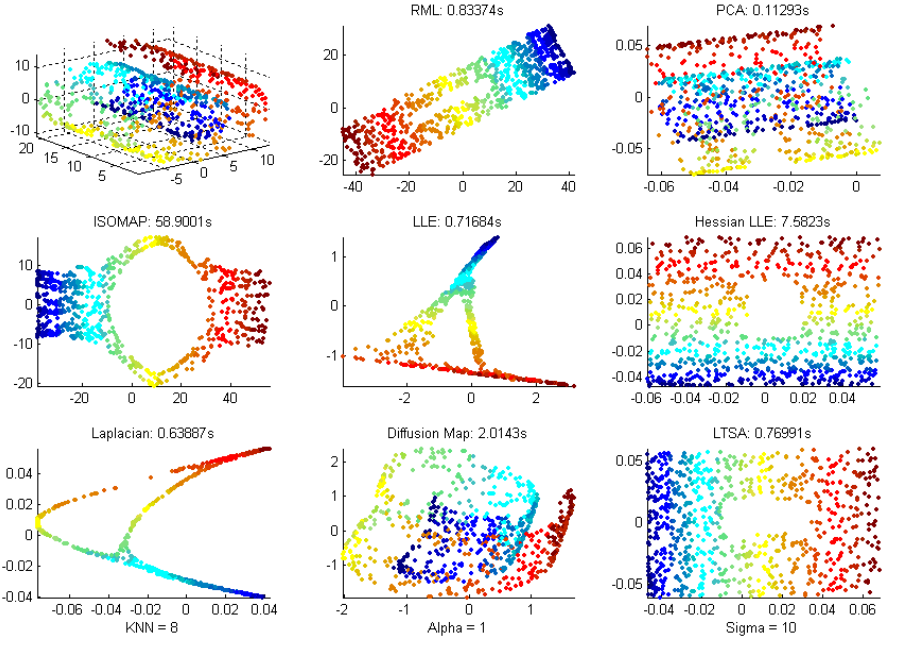

Fig. 5. Comparison results of Swiss Hole

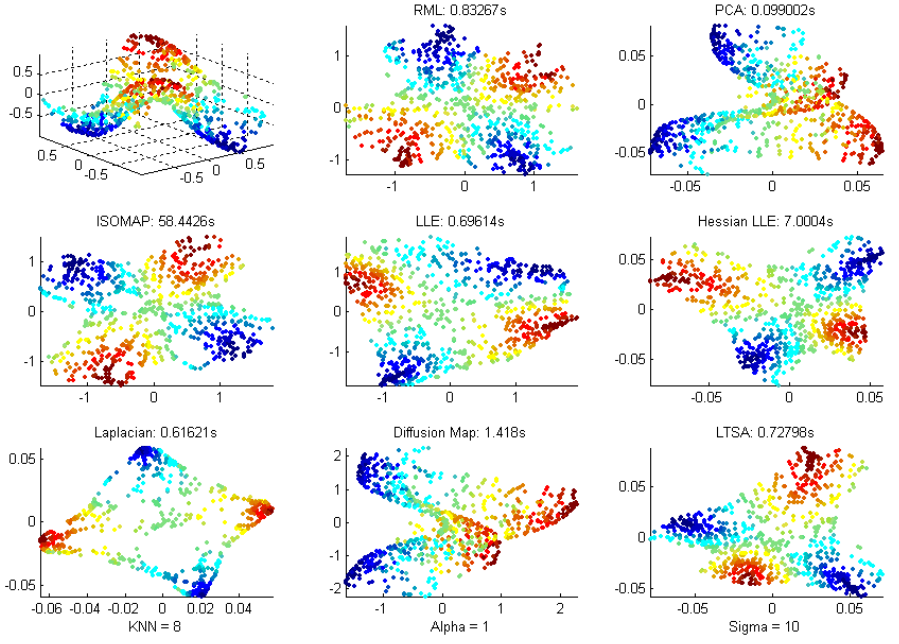

Fig. 6. Comparison results of Twin Peaks

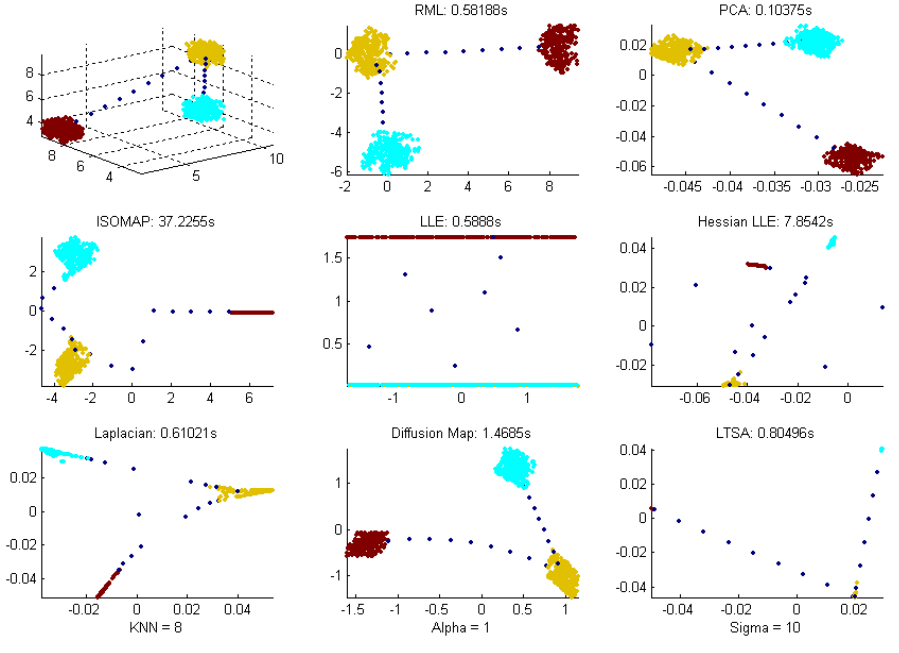

Fig. 7. Comparison results of 3D Clusters

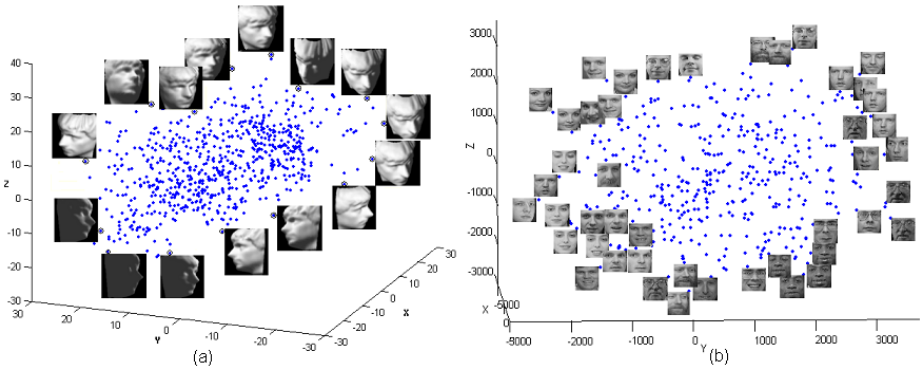

Fig. 8. Three-dimensional embedding: (a) Isomap face data; (b) ORL face data

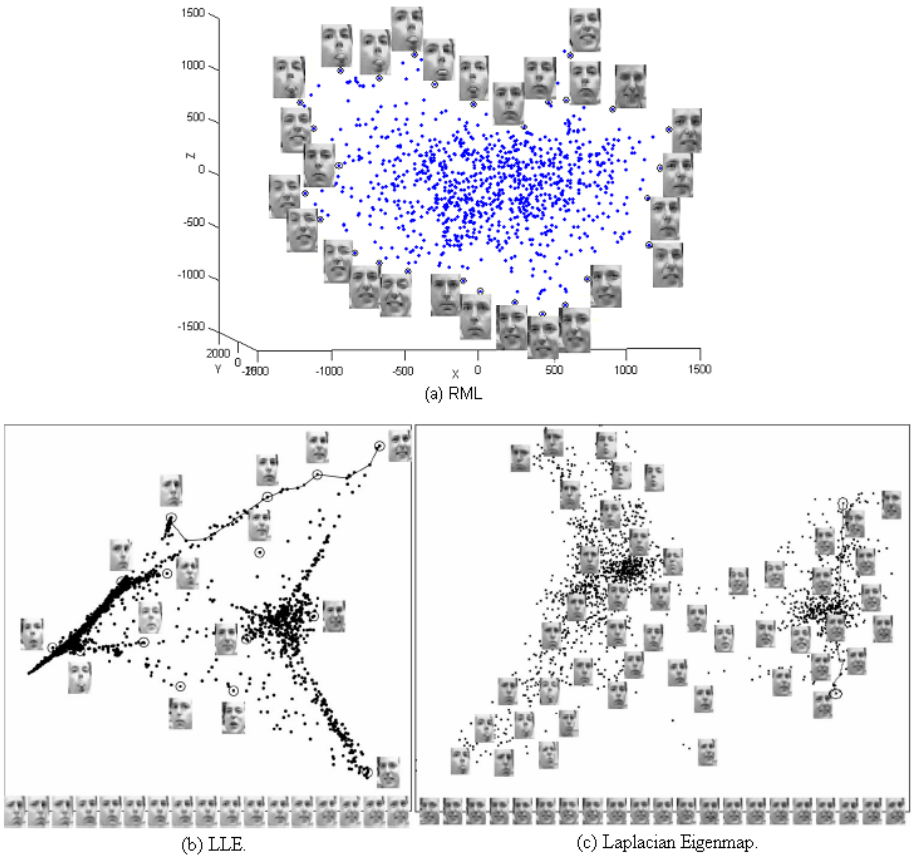

Fig. 9. Comparison results of LLE face data

is 2. Notice that our estimation for the Isomap face dataset is 4, though it is rendered with 3 parameters (one for lighting and two for pose). Several other methods [26] reported similar estimates of about 4 dimension.

Second, four sets of synthetic data from the MANI demo (http://www.math.umn.edu/∼wittman/mani/) and the above three sets of face data are used to illustrate the behavior of our manifold learning algorithm RML. For synthetic data, several other competing algorithms (PCA, Isomap, LLE, HLLE, Laplacian Eigenmaps, Diffusion maps, LTSA) are compared and the results are shown in Fig. 4-7. RML outperforms other algorithms by correctly learning the nonlinear geometry of each data set. Both RML and Isomap have metric-preserving properties, e.g. intrinsically mapping Swiss Roll data onto a 2D Rectangle region. However, Isomap fails on Swiss Hole. In general, LTSA and HLLE consistently perform better than other spectral methods, though they cannot preserve the original metrics of each data set. The running speed of RML is less than one second, which is comparable to that of LLE, Laplacian Eigenmaps, and LTSA. Often HLLE and Diffusion maps spend several seconds, while Isomap needs one minute. Fig. 8-9 show the embedding results of three sets of face data. In contrast to LLE [11] and Laplacian Eigenmaps [13], RML yields embedding results that are uniformly distributed and well organized.

7 Conclusion

We presented a RNC-based manifold learning method for nonlinear dimensionality reduction, which can learn intrinsic geometry of the underlying manifold with metric-preserving properties. Experimental results demonstrate the excellent performance of our algorithm on synthetic and real data sets. The algorithm should find a wide variety of potential applications, such as data analysis, visualization, and classification.

References

1. Donoho, D.: High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality. Mathematical Challenges of the 21st Century. American Mathematical Society, Los Angeles, CA (2000)
2. Jolliffe, I.: Principal Component Analysis. Springer-Verlag, New York (1989)
3. Cox, T., Cox, M.: Multidimensional Scaling. Chapman & Hall, London (1994)
4. Bourlard, H., Kamp, Y.: Auto-association by multilayer perceptrons and singular value decomposition. Biological Cybernetics **59** (1988) 291–294
5. Erwin, E., Obermayer, K., Schulten, K.: Self-organizing maps: Ordering, convergence properties and energy functions. Biological Cybernetics **67** (1992) 47–55
6. Durbin, R., Willshaw, D.: An analogue approach to the travelling salesman problem using an elastic net method. Nature **326** (1987) 689–691
7. Bishop, C., Svenson, M., Williams, C.: Gtm: The generative topographic mapping. Neural Computation **10** (1998) 215–234
8. Hastie, T., Stuetzle, W.: Principal curves. J. American Statistical Association **84** (1989) 502–516

9. Tenenbaum, J., Silva, V.d., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science **290** (2000) 2319–2323
10. Silva, V., Tenenbaum, J.: Global versus local methods in nonlinear dimensionality reduction. Advances in Neural Information Processing Systems, MIT Press (2003)
11. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science **290** (2000) 2323–2326
12. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation **15** (2003) 1373–1396
13. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: Face recognition using laplacianfaces. IEEE Trans. Pattern Analysis and Machine Intelligence **27** (2005) 328–340
14. Donoho, D., Grimes, C.: Hessian eigenmaps: New tools for nonlinear dimensionality reduction. Proc. National Academy of Science (2003) 5591–5596
15. Weinberger, K., Saul, L.: Unsupervised learning of image manifolds by semidefinite programming. Proc. CVPR (2004) 988–995
16. Brand, M.: Charting a manifold. Advances in Neural Information Processing Systems, MIT Press (2003)
17. Zhang, Z., Zha, H.: Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM Journal on Scientific Computing **26** (2004) 313–338
18. Nadler, B., Lafon, S., Coifman, R., Kevrekidis, I.: Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. (submitted to Journal of Applied and Computational Harmonic Analysis)
19. Seung, H., Lee, D.: The manifold ways of perception. Science **290** (2000) 2268–2269
20. Sha, F., Saul, L.: Analysis and extension of spectral methods for nonlinear dimensionality reduction. Proc. Int. Conf. Machine Learning, Germany (2005)
21. Brun, A., Westin, C.F., Herberthson, M., Knutsson, H.: Fast manifold learning based on riemannian normal coordinates. Proc. 14th Scandinavian Conf. on Image Analysis, Joensuu, Finland (2005)
22. Jost, J.: Riemannian Geometry and Geometric Analysis. 3rd edn. Springer (2002)
23. Freedman, D.: Efficient simplicial reconstructions of manifolds from their samples. IEEE Trans. Pattern Analysis and Machine Intelligence **24** (2002) 1349–1357
24. Cormen, T., Leiserson, C., Rivest, R., Stein, C.: Introduction to Algorithms. MIT Press, Cambridge (2001)
25. Golub, G., Loan, C.: Matrix Computations. 3rd edn. Jonhs Hopkins Univ. (1996)
26. Levina, E., Bickel, P.: Maximum likelihood estimation of intrinsic dimension. Advances in Neural Information Processing Systems, MIT Press (2004)