Learning Bundle Manifold by Double Neighborhood Graphs

Chun-guang Li, Jun Guo, and Hong-gang Zhang

PRIS lab., School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications,
100876 Beijing, China
{lichunguang,guojun,zhhg}@bupt.edu.cn

Abstract. In this paper, instead of the ordinary manifold assumption, we introduced the bundle manifold assumption that imagines data points lie on a bundle manifold. Under this assumption, we proposed an unsupervised algorithm, named as *Bundle Manifold Embedding (BME)*, to embed the bundle manifold into low dimensional space. In BME, we construct two neighborhood graphs that one is used to model the global metric structure in local neighborhood and the other is used to provide the information of subtle structure, and then apply the spectral graph method to obtain the low-dimensional embedding. Incorporating some prior information, it is possible to find the subtle structures on bundle manifold in an unsupervised manner. Experiments conducted on benchmark datasets demonstrated the feasibility of the proposed BME algorithm, and the difference compared with ISOMAP, LLE and Laplacian Eigenmaps.

1 Introduction

In the past decade manifold learning has attracted a surge of interest in machine learning and a number of algorithms are proposed, including ISOMAP[1], LLE[2,3], Laplacian Eigenmap[4], Hessian LLE[5], Charting[6], LTSA[7], MVU[8], Diffusion Map[9] and etc. All these focus on finding a nonlinear low dimensional embedding of high dimensional data. So far, these methods have mostly been used for the task of exploratory data analysis such as data visualization, and also been successfully applied to semi-supervised classification problem[10,11].

Under the assumption that the data points lie close to a low dimensional manifold embedded in high dimensional Euclidean space, manifold learning algorithms learn the low dimensional embedding by constructing a weighted graph to capture local structure of data. Let $G(V, E, W)$ be the weighted graph, where the vertex set V corresponds to data points in data set, the edge set E denotes neighborhood relationships between vertices, and W is the weight matrix. The different methods to choose the weight matrix W will lead to different algorithms.

In the case of multiple manifolds co-exist in dataset, one should tune neighborhood size parameter (i.e. k in $k-nn$ rule, ϵ in ϵ -rule) to guarantee the constructed neighborhood graph to be connected, or to deal with each connected

H. Zha, R.-i. Taniguchi, and S. Maybank (Eds.): ACCV 2009, Part III, LNCS 5996, pp. 321–330, 2010.

components individually. Neighborhood graph can be viewed as discretized representation of the underlying manifolds and the neighborhood size parameter controls the fitness to the true manifolds. Constructing neighborhood graph by a larger neighborhood size parameter will result in losing 'the details' of the underlying manifolds, due to the occurrence of 'short-cut' connections between different manifolds or the different parts of the same manifold (when the manifold with high curvature). On the other hand, treating each connected component individually will lose the information of correspondence between different components so that the panorama of the global geometry of observation data set cannot be obtained.

In this paper, instead of the global manifold assumption, we suggest that it is more appropriate to imagine data points to lie on a bundle manifold, when one faces the task to visualize the image dataset in which may consists of multiple classes. For example, facial images of a certain person, under the pose changing and expression variation, in which the different expressions relate to different classes, span a bundle manifold. Under the bundle manifold assumption, to make the embedding faithful, it is needed to preserve the subtle local metric structures. We propose a naive way to visualize the subtle structure of bundle manifold, named as Bundle Manifold Embedding (BME). By incorporating the intrinsic dimension as apriori information, BME can discover the subtle substructure in an unsupervised manner. Experiments conducted on benchmark datasets validated the existence of subtle structure, and demonstrated the feasibility and difference of the proposed BME algorithm compared with ISOMAP, LLE and Laplacian Eigenmaps.

1.1 The Related Work

In computer vision, object images with continuous pose variations can be imagined to lie close to a Lie group[12]. Ham et al.[13] reported that different databases may share the same low-dimensional manifold and the correspondence relationship between different data sets can be learned by constructing unit decomposition. Lim et al.[14] presented a geometric framework on the basis of quotient space for clustering images with continuous pose variation. Recently, Ham and Lee[15] assumed a global principal bundle as the model of face appearance manifold for facial image data set with different expressions and pose variation. And they also proposed a factorized isometric embedding algorithm to extract the substructure of facial expressions and the pose change separately.

In this paper, following the way consistent with Ham and Lee[15], we suppose that in some multi-class dataset data points lie on bundle manifold and each class-specific manifold relates to an orbit. Under the bundle manifold assumption, we will discuss how to develop manifold learning algorithm to discover the subtle substructure of bundle manifold.

2 The Proposed Algorithm: Bundle Manifold Embedding

The principle of learning manifold is to preserve the local metric structure when obtain the low dimensional embedding. In the case of bundle manifold assumption,

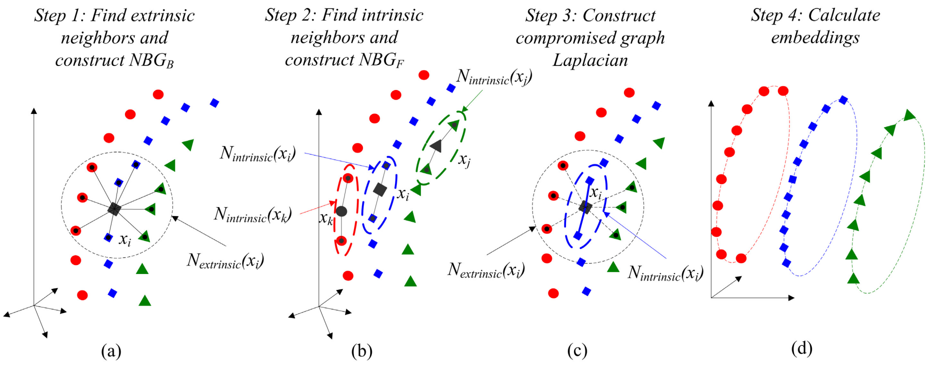

Fig. 1. Illustration for the proposed BME algorithm

however, there are extra subtle substructures, that is, the orbit structure. Therefore the local metric structures on bundle manifold consist of two aspects: (a) local metric between different orbits, and (b) local metric within each orbit. To obtain the faithful embedding of bundle manifold, two kinds of local metric need to be preserved as much as possible.

We propose to construct two neighborhood graphs: NBG_B and NBG_F, where NBG_B is used to represent the local metric structure between orbits, and NBG_F is used to capture the subtle substructure for each orbit (an orbit is likely corresponding to data points within each class, also can be viewed as fibre manifold). By means of spectral graph approach on combination of the weight matrices of two neighborhood graphs, the low dimensional embedding can be obtained.

Given data set $X = \{x_i, x_i \in R^m, i = 1, ..., n\}$, where m is the dimensionality of the observation space and n is the number of data points in data set X. First of all, we need to interpret a little of the two concepts: ***extrinsic neighbors*** and ***intrinsic neighbors***, denoted as $N_{extrinsic}$ and $N_{intrinsic}$, respectively. An illustration is given in Fig. 1 (a) and (b). For a data point $x_i \in X$, its extrinsic neighbors $N_{extrinsic}(x_i)$ can be selected by the ambient space metric, that is, Euclidean metric; whereas the intrinsic neighbors $N_{intrinsic}(x_i)$ of $x_i \in X$ are defined to be able to capture the subtle substructure of each orbit. In fact the extrinsic neighbors are the same as in the common manifold learning algorithms; whereas the intrinsic neighbors need to be selected along each orbit. The later is the key to discover the subtle structure of bundle manifold.

The proposed bundle manifold embedding algorithm consists of four steps as illustrated in Fig. 1, and will be depicted in the next subsections.

2.1 Find Extrinsic Neighbors and Construct Extrinsic Neighborhood Graph NBG_B

For each data point $x_i \in X$, we select the extrinsic neighbors $N_{extrinsic}(x_i)$ by Euclidean distance. We prefer to choose $k - nn$ rules rather than ϵ ball rule for its simplicity to determine the local scale parameter. And then a weighted graph with n nodes, one for each of data points, and a set of edges connecting each of extrinsic neighbors, is constructed. Denoted the extrinsic neighborhood graph as $NBG_B(V, E_B, W)$, where the vertex set V correspond to data points, the undirected edges set E_B denote neighborhood relationships between the vertices and the weight W_{ij} on each edge is the similarity between the two vertexes. We make use of the Gaussian kernel $W_{ij} = e^{-\|x_i - x_j\|/\sigma_i\sigma_j}$ with adaptive scale parameter $\sigma_i = \|x_i - x_i^k\|$ and $\sigma_j = \|x_j - x_j^k\|$, where x_i^k and x_j^k are the k-th nearest neighbors of x_i and x_j, respectively.

The neighborhood graph $NBG_B(V, E_B, W)$ is constructed to offer the background information for discovering the subtle structure of orbits, instead of to directly calculate the local metric information between different orbits. Therefore the neighborhood scale parameter k need to be large enough to guarantee the connectivity of extrinsic neighborhood graph $NBG_B(V, E_B, W)$. The neighborhood relationship defined by Euclidean distance can be used to preserve the 'global structure' within local neighborhood[1], see Fig. 1(a).

2.2 Find Intrinsic Neighbors and Construct Intrinsic Neighborhood Graph NBG_F

The subtle substructure in bundle manifold is orbits (also called fibers) of the structure group. For multi-class data set, Euclidean distance cannot distinguish such a subtle local metric structure, so that it often failed to find the orbit structure. In order to select intrinsic neighbors in an unsupervised way, we need a new metric, which can discern the subtle difference between different orbits.

Under the bundle manifold assumption, each orbit is controlled by structure group (a Lie group), and it is also a smooth manifold. Therefore we can suppose that each data point and its neighbors lie on a locally linear patch of the smooth manifold within each orbit and define the intrinsic neighbors by local linear structure. Based on neighborhood graph $NBG_B(V, E_B, W)$, we characterize the local geometry of these patches by nonnegative linear coefficients that reconstruct each data point x_i from its k neighbors x_i^j where $j \in N_{extrinsic}(x_i)$, $N_{extrinsic}(x_i)$ is the index set of k extrinsic neighbors of x_i. Here we need to solve a set of quadratic programming problems: for $x_i \in X, i = 1, \ldots, n$

$$\begin{aligned}
&\min_{a_{ij}} \varepsilon(a_{ij}) = \|x_i - \textstyle\sum_{j \in N_{extrinsic}(x_i)} a_{ij} x_i^j\|^2 \\
&\quad s.t. \textstyle\sum_{j \in N_{extrinsic}(x_i)} a_{ij} = 1 \\
&\quad\quad a_{ij} \geq 0
\end{aligned} \tag{1}$$

where a_{ij} are the nonnegative local linear reconstructing coefficients. For $j \notin N_{extrinsic}(x_i)$, a_{ij} are set to zero. Notice that weight matrix A constructed in such a manner is consistent with convex LLE[3].

[1] Under the bundle manifold assumption, the local subtle structure is orbit structure. Euclidean distance cannot distinguish the subtle orbit structure. Therefore, the neighborhood computed by Euclidean distance can be used as reference information to capture the global structure of neighborhood.

Given the nonnegative reconstructing coefficients matrix A, constructing the intrinsic neighborhood graph $NBG_F(V, E_F, A)$ is straightforward. Taking the intrinsic dimension d of data set as apriori information, we keep the $d+1$ largest nonnegative reconstructing coefficients and set those minor coefficients to zero. This is our recipe to find the exact intrinsic neighbors. The nonnegative coefficients are treated as affinity measure to find the intrinsic neighbors and those dominant coefficients indicate a reliable subtle neighborhood relationship of each orbit.

In fact, the intrinsic neighborhood graph $NBG_F(V, E_F, A)$ is refinement of the extrinsic neighborhood graph and can be derived from $NBG_B(V, E_B, W)$ by removing those edges related to minor reconstructing coefficients a_{ij}. The remaining $d+1$ dominant positive coefficients will span simplex with dimension d. These simplex can reveal the subtle intrinsic structure hiding in bundle manifold, that is, help us to reveal those orbit structures or class-specific manifolds. In the virtue of this recipe, the obtained intrinsic neighborhood graph can distinguish the subtle local structures.

2.3 Construct the Generalized Graph Laplacian

We denote L_B as the normalized graph Laplacian of $NBG_B(V, E_B, W)$, where $L_B = D^{-1/2}(D - W)D^{-1/2}$, D is diagonal matrix with entries $D_{ii} = \sum_j W_{ij}$. On the other hand, we denote $L_A = A^T A$, in which L_A is a nonnegative symmetric matrix for capturing the subtle substructure from $NBG_F(V, E_F, A)$. To obtain the faithful embedding of bundle manifold, both the local metric within each orbit, and the local metric between different orbits need to be preserved. Therefore we need to make use of both information from $NBG_B(V, E_B, W)$ and $NBG_F(V, E_F, A)$ to form an affinity relationship matrix. For simplicity, we linearly combine the normalized graph Laplacian matrix L_B and the nonnegative symmetric matrix L_A as follows:

$$U = (1 - \gamma)L_B + \gamma L_A \tag{2}$$

where $\gamma (0 \leq \gamma < 1)$ is a tradeoff parameter. In the extreme case, $\gamma = 0$, it turned out to be Laplacian Eigenmap[4], in which the subtle substructure information from orbit is ignored. The more γ tends to 1, the more the subtle substructure is emphasized.

The matrix U is nonnegative, symmetric[2], and carries two aspects of local metric information. We treat U as affinity weight matrix to construct a generalized graph Laplacian $\tilde{L}$ for embedding the data points into low dimensional space, in which $\tilde{L} = U - \tilde{D}$ and $\tilde{D}$ is diagonal matrix with entries $\tilde{D}_{ii} = \sum_j U_{ij}$.

2.4 Embed into Low Dimensional Space

Under the assumption that data points lie on bundle manifold, mapping data points from high-dimensional space into low dimensional space must keep the

[2] The weight matrix W need to be symmetric.

local metric both on orbits and between orbits as much as possible. The constructed generalized graph Laplacian $\tilde{L}$ above is ready for such a purpose.

Suppose that the embedding is given by the $n \times q$ matrix $Y = [y_1, y_2, \ldots, y_n]^T$ where the i-th row provides the embedding coordinates of the i-th data point. As in Laplacian Eigenmap[4], we formulated a quadratic optimization problem as following:

$$\begin{aligned} &\min \varepsilon(Y) = \frac{1}{2} \sum_{i,j} \{U_{ij} \|y_i - y_j\|^2\} = trace(Y^T \tilde{L} Y) \\ &s.t.\ \ Y^T \tilde{D} Y = I \\ &\qquad Y^T \tilde{D} \mathbf{1} = \mathbf{0} \end{aligned} \tag{3}$$

where $y_i \in R^q$ is q-dimensional row vector, I is identity matrix, $\mathbf{1} = [1, 1, \ldots, 1]^T$ and $\mathbf{0} = [0, 0, \ldots, 0]^T$. Apparently, the optimization problem in (3) is a generalized eigenvector problem as:

$$\tilde{L}\mathbf{f} = \lambda \tilde{D} \mathbf{f} \tag{4}$$

Let $\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_q$ be the solution of equation (4), and ordered them according to their eigenvalues in ascending order:

$$\begin{aligned} \tilde{L}\mathbf{f}_0 &= \lambda_0 \tilde{D} \mathbf{f}_0 \\ \tilde{L}\mathbf{f}_1 &= \lambda_1 \tilde{D} \mathbf{f}_1 \\ &\cdots \\ \tilde{L}\mathbf{f}_q &= \lambda_q \tilde{D} \mathbf{f}_q \end{aligned} \tag{5}$$

where $0 = \lambda_0 \le \lambda_1 \le \ldots \le \lambda_q$. We remove the eigenvector $\mathbf{f}_0$ corresponding to eigenvalue 0 and employ the next q eigenvectors to obtain the embedding Y as $Y = [y_1, y_2, \ldots, y_n]^T = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_q]$ in q-dimensional Euclidean space.

3 Experimental Results

In this section we will demonstrate data visualization experiments on the benchmark data sets. The proposed BME algorithm is compared with ISOMAP, LLE, and Laplacian Eigenmaps. The parameter k is chose $k = 20$ for all algorithms in all experiments and the intrinsic dimension d used in BME is set to $d = 1$.

The first set of experiments are conducted on a selected data subset COIL-4, which consists of image samples from four classes (i. e. object-3, object-5,object-6 and object-19) of COIL-20[16]. The four classes of objects are the most similar four classes of object images in COIL-20. The manifold embedded in each class of COIL-20 is homeomorphism to circle (i.e. S^1).

As can be seen from Fig. 2 in panels (a), (b) and (c), ISOMAP, LLE, and Laplacian Eigenmap all failed to find the subtle class-specific manifold substructures. The results obtained by the proposed BME algorithm are given in Fig. 2 panels (d), (e) and (f) with different tradeoff parameter $\gamma = 0.90, 0.95, 0.99$. It is obvious that the subtle class-specific manifold substructures are discovered by BME algorithm.

In Fig. 2 panel (d) and (e), however, it can be observed that data points of the three classes of object-3, object-6 and object-19 are still piled together. We gathered those piled data points of the three classes (object-3, object-6 and

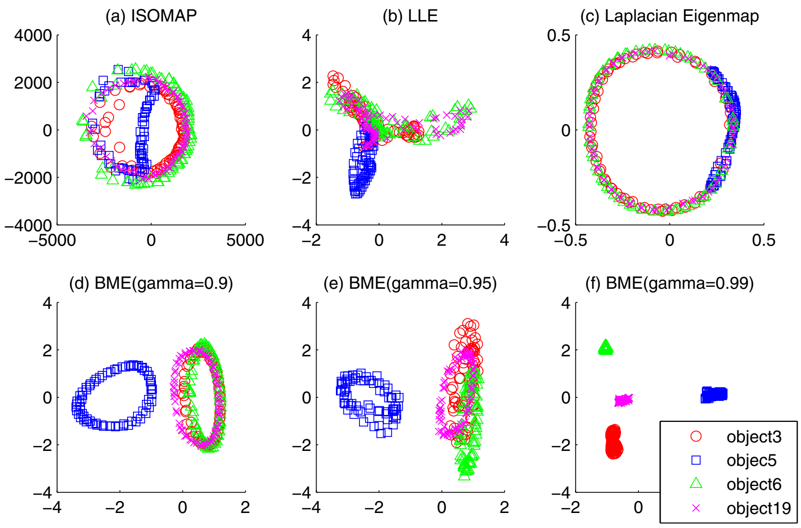

Fig. 2. Data visualization results that compared the proposed BME with ISOMAP, LLE, Laplacian Eigenmap on COIL-4 data set (where k=20 is used for all algorithms)

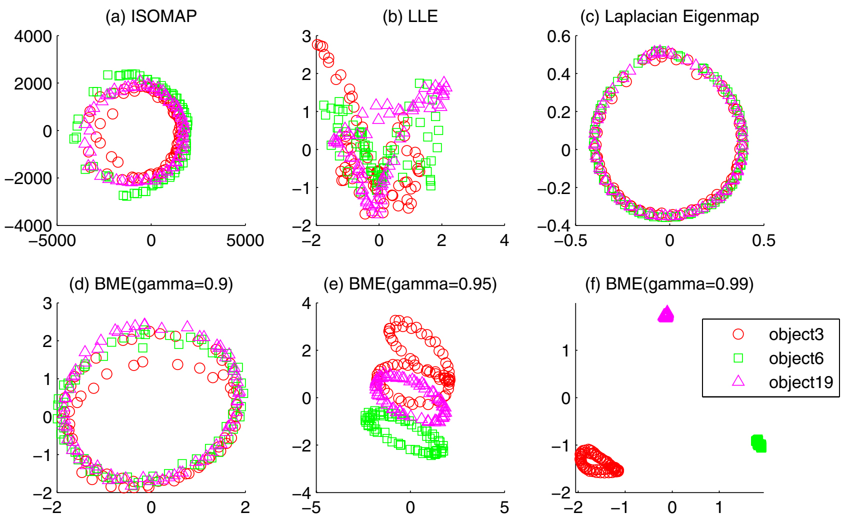

Fig. 3. Data visualization results that compared the proposed BME with ISOMAP, LLE, Laplacian Eigenmap on COIL-3 (where k=20 is used for all algorithms)

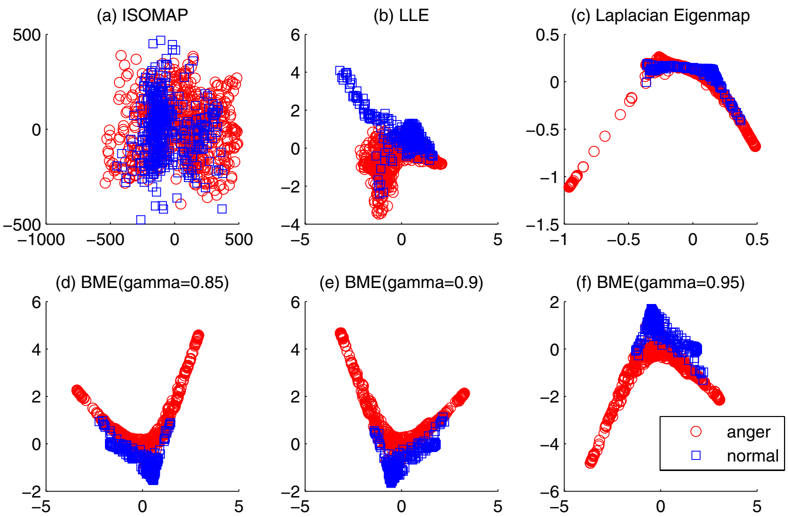

Fig. 4. Data visualization results that compared the proposed BME with ISOMAP, LLE, Laplacian Eigenmap on FreyfaceExpression-2 data set (where k=20 is used for all algorithms)

object-19) in Fig. 2 panel (d) and (e) to form data set COIL-3. It is interesting to further visualize the data in COIL-3. Therefore we conduct the second set of experiments on dataset COIL-3. The results are presented in Fig. 3. The ISOMAP, LLE and Laplacian Eigenmap still failed; whereas the proposed BME algorithm can reveal the subtle substructures distinctly.

From the data visualization results above, we can draw the conclusion that the tradeoff parameter γ controls the clearness of the discovered subtle substructure. The larger the parameter γ is, the more consideration is taken into the intrinsic neighborhood graph, and it will lead to focusing on subtle substructure much more. The smaller parameter γ corresponds to take into consideration more the background information, that is, from the extrinsic neighborhood graph. An over-large γ, however, will degrade the corresponding relationship between different class-specific manifolds, and will result in losing the faithfulness of the obtained subtle substructure. Strictly speaking, the global geometric structure hiding in the COIL-4 and COIL-3 data sets may not be exactly a principal bundle manifold, but then using such an assumption will remind us to capture the true geometric structure carefully and help us to explore the 'real feature' of data set.

The third set of experiments are conducted on Frey face data set[3]. We manually sorted the Frey face dataset into five expressions categories ('anger', 'happy', 'sad', 'tongue_out' and 'normal') and choose the two most similar classes of expressions ('anger' and 'normal') as the FreyfaceExpression-2 dataset for data

[3] http://www.cs.toronto.edu/~roweis/data.html, B. Frey and S. Roweis

visualization. There are continuous pose variation in both 'anger' and 'normal' expressions. Therefore the two expression manifolds will share the similar one-dimensional subtle structure. The experimental results are displayed in Fig. 4. We can find that the BME is superior to the other algorithms that it can reveal the subtle substructure.

Finally we need to mention about the parameters k, d and γ in BME. The neighborhood scale parameter k need to be large enough to guarantee the connectivity of extrinsic neighborhood graph to provide reference information. For the explorative data analysis task, one can try the intrinsic dimension d from one to the estimated intrinsic dimension by those significant nonnegative local linear reconstruction coefficients. In addition, the parameter γ may be selected from some typical values, such as 0.85, 0.90, 0.95, 0.99 and etc.

4 Concluding Remarks and Discussion

In this paper we suggest that the true global geometric structure of some datasets is likely bundle manifold, not a single manifold, and also presented a naive unsupervised algorithm, BME, for visualizing the subtle structure. Experiments on benchmark data sets demonstrated the feasibility of the proposed BME algorithm. We believe that the principal bundle manifold assumption and the proposed bundle manifold embedding algorithm are beneficial to deeply understand the global intrinsic geometry of some image datasets.

Under bundle manifold assumption, an interesting question arose that what is the exact meaning of the estimated intrinsic dimensionality. Perhaps we need redefine the task of intrinsic dimension estimation due to the exitance of orbit structures. In the proposed algorithm, however, the information used is only the local metric at each of orbits and the locality on bundle manifold. The sharing intrinsic structure among orbits has not been employed yet. Therefore the more sophisticated means to learning the bundle manifold will be investigated in future.

Acknowledgments. This work was partially supported by National High-Tech Research and Development Plan of China under Grant No. 2007AA01Z417 and the 111 project under grand No.B08004.

References

1. Tenenbaum, J.B., Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)
2. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
3. Saul, L.K., Roweis, S.T.: Think globally, fit locally: unsupervised learning of low dimensional manifolds. Journal of Machine Learning Research 4, 119–155 (2003)
4. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15(6), 1373–1396 (2003)

5. Donoho, D.L., Grimes, C.: Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. Proc. Natl. Acad. Sci. USA 100(10), 5591–5596 (2003)
6. Brand, M.: Charting a manifold. In: NIPS, vol. 15. MIT Press, Cambridge (2003)
7. Zhang, Z., Zha, H.: Principal manifolds and nonlinear dimensionality reduction by local tangent space alignment. SIAM Journal of Scientific Computing 26(1), 313–338 (2004)
8. Weinberger, K., Packer, B., Saul, L.: Unsupervised learning of image manifolds by semidefinite programming. In: CVPR 2004, vol. 2, pp. 988–995 (2004)
9. Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., Zucker, S.W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proc. of the Natl. Academy of Sciences 102, 7426–7431 (2005)
10. Belkin, M., Niyogi, P.: Semi-supervised learning on riemannian manifolds. Machine Learning 56(1-3), 209–239 (2004)
11. Wang, F., Zhang, C.: Label propagation through linear neighborhoods. IEEE Transactions on Knowledge and Data Engineering 20(1), 55–67 (2008)
12. Rao, R., Ruderman, D.: Learning lie groups for invariant visual perception. In: NIPS, vol. 11. MIT Press, Cambridge (1999)
13. Ham, J., Lee, D., Saul, L.: Semisupervised alignment of manifolds. In: AI & STAT, pp. 120–127 (2005)
14. Lim, J., Ho, J., Yang, M.-H., Lee, K.-C., Kriegman, D.J.: Image clustering with metric, local linear structure and affine symmetry. In: Pajdla, T., Matas, J.G. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 456–468. Springer, Heidelberg (2004)
15. Hamm, J., Lee, D.D.: Separating pose and expression in face images: A manifold learning approach. Neural Information Processing – Reviews and Letters 11, 91–100 (2007)
16. Nene, S.A., Nayar, S.K., Murase, H.: Columbia object image library (coil-20). Technical Report CUCS-005-96, Columbia University (1996)