Spectral Clustering on Multiple Manifolds

Yong Wang, Yuan Jiang, Yi Wu, and Zhi-Hua Zhou, *Senior Member, IEEE*

***Abstract*— Spectral clustering (SC) is a large family of grouping methods that partition data using eigenvectors of an affinity matrix derived from the data. Though SC methods have been successfully applied to a large number of challenging clustering scenarios, it is noteworthy that they will fail when there are significant intersections among different clusters. In this paper, based on the analysis that SC methods are able to work well when the affinity values of the points belonging to different clusters are relatively low, we propose a new method, called spectral multi-manifold clustering (SMMC), which is able to handle intersections. In our model, the data are assumed to lie on or close to multiple smooth low-dimensional manifolds, where some data manifolds are separated but some are intersecting. Then, local geometric information of the sampled data is incorporated to construct a suitable affinity matrix. Finally, spectral method is applied to this affinity matrix to group the data. Extensive experiments on synthetic as well as real datasets demonstrate the promising performance of SMMC.**

***Index Terms*— Clustering, local tangent space, manifold clustering, spectral clustering.**

I. Introduction

IN MACHINE learning and pattern recognition, an important research direction is to group internally "similar" objects into the same cluster and "dissimilar" objects into different clusters, which is known as *cluster analysis* [1].

Traditional central grouping techniques, e.g., K-means [2], proceed by comparing all the data points to a small number of cluster prototypes or centroids. A major disadvantage of these methods is that they cannot be used to separate clusters that are very long or nonlinearly separable.

During the past decade, spectral clustering methods (abbreviated as SC) [3]–[7] have emerged as a principled relaxation of the NP-hard normalized cut clustering problem, and have been successfully applied to a number of challenging clustering scenarios. Essentially, spectral methods group data by using eigenvectors of an affinity matrix which is derived from pairwise similarities between points of the original data.

Manuscript received September 5, 2010; revised April 9, 2011; accepted April 15, 2011. Date of publication June 17, 2011; date of current version July 7, 2011. This work was supported in part by the National Fundamental Research Program of China under Grant 2010CB327903, the National Science Foundation of China under Grant 60975043 and Grant 60975038, and the Jiangsu 333 High-Level Talent Cultivation Program.

Y. Wang is with the Department of Mathematics and Systems Science, National University of Defense Technology, Changsha 410073, China. He is also with the National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (e-mail: yongwang82@gmail.com).

Y. Jiang and Z.-H. Zhou are with the National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (e-mail: jiangy@lamda.nju.edu.cn; zhouzh@lamda.nju.edu.cn).

Y. Wu is with the Department of Mathematics and Systems Science, National University of Defense Technology, Changsha 410073, China (e-mail: wuyi_work@sina.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNN.2011.2147798

However, a critical issue is that the promising success of SC relies on the situation that the samples are generated from approximately well-separated clusters, where each cluster can be considered as a connected component to a certain extent [3], [6]. They will fail when there are significant intersections among different clusters. The reason is that the performance of SC heavily relies on the constructed undirected graph or the affinity matrix, while different clusters near the intersections will easily be connected by the undirected graph, and thus the affinity matrix will be corrupted with poor pairwise affinity values and misleading information will diffuse across different clusters [8]. Thus, the way for extending the scope of SC to the intersecting case is a challenging and interesting work. Note that it is hard for any clustering method to give a reasonable performance if there is no restriction on the distribution of the samples. Therefore, in this paper, we focus our attention on the clustering of unlabeled data observations that lie on or close to multiple smooth low-dimensional manifolds, some of which possibly intersecting with each other (see Fig. 1). This restriction is reasonable since a large number of recent research efforts have shown that the perceptually meaningful structure of many real-world data possibly resides on a low-dimensional manifold [9]–[12].

Another line of research that motivates us to extend SC methods to group multiple manifolds with possible intersections is *manifold clustering* [13], [14]. Manifold clustering, which regards a cluster as a group of points around a compact low-dimensional manifold, has been realized as a reasonable and promising generalization of traditional centroid-based clustering methods. Although this field is fairly new, a considerable amount of work has recently been reported. When different manifolds have different intrinsic dimensions and densities, we can divide the points into different clusters according to both dimensionality and density, e.g., the translation Poisson mixture model (TPMM) [15]. However, it becomes infeasible when different manifolds have the same dimensionality and density. Thus, more effective methods are required. When the samples can be well approximated by a mixture of linear manifolds (linear or affine subspaces), a large number of elegant linear manifold clustering methods can be utilized for representing the linear manifolds and clustering the data points. For example, generalized principal component analysis (GPCA) [16] represents the underlying linear manifolds by using a set of homogeneous polynomials, local subspace affinity (LSA) [17] computes an affinity for any pair of points based on principal subspace angles between different linear manifolds, and spectral curvature clustering (SCC) [24] computes polar curvature for each fixed-size subset of the data. Alternatively, K-planes [18], [19] addresses linear manifold clustering by iterating between assigning data to manifolds and estimating a manifold to each cluster. However,

1045–9227/$26.00 © 2011 IEEE

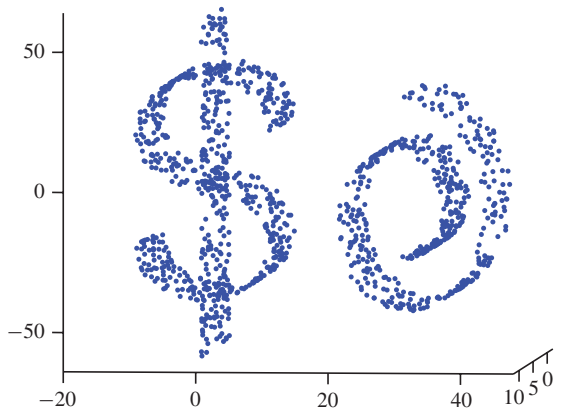

Fig. 1. Data points drawn from multiple manifolds where some data manifolds are separated but some are intersecting.

all the linear methods fail to deliver good performance in the presence of nonlinear structures.

Since nonlinear methods can be naturally applied to linear and/or nonlinear manifolds, more attention should be paid to nonlinear manifold clustering methods. As far as we know, K-manifolds [13], which starts by estimating geodesic distances between points, is the first method to classify unorganized data nearly lying on multiple intersecting nonlinear manifolds. Unfortunately, this method is limited to dealing with intersecting manifolds since the estimation of geodesic distances will fail when there are widely separated clusters. On the contrary, existing SC methods [3]–[7] are well suited to group samples generated from separated manifolds but have difficulty in dealing with points near the intersections of different manifolds. Therefore, the way of addressing the problem of *hybrid nonlinear manifold clustering* [14], which is a more general framework where some data manifolds are separated but some are intersecting is an interesting research direction.

Recently, a "divide and conquer" method, named mumCluster [14], has been proposed to deal with hybrid nonlinear manifold clustering. MumCluster first divides the complicated hybrid modeling into single manifolds and intersecting manifolds. Then, each intersecting manifold is further divided into intersection areas and non-intersection areas. Finally, a more faithful undirected graph is constructed to reveal different clusters. Empirical studies have shown the effectiveness of mumCluster, however, it is quite heuristic. Moreover, mumCluster heavily relies on the correct local dimension estimation, which will have difficulty when faced with noisy real-world data.

In this paper, we extend the scope of SC methods to give a more principled method for the grouping of multiple smooth low-dimensional manifolds. Our basic idea is based on the analysis that spectral methods are able to work well when the affinity values for pairwise points belonging to different clusters are relatively low. However, in traditional SC methods which are based on a radial distance between the pairwise points, the affinity matrix will be corrupted with poor affinity values, i.e., the affinity value between the pairwise points belonging to different clusters but near the intersection is high. Thus, misleading information will diffuse across different clusters and spectral methods cannot give reasonable performances [8]. Accordingly, adaptations are required when the samples are lying on or close to multiple smooth manifolds with possible intersections. Fortunately, in this case, some natural geometric information of the sampled data on the manifold can be exploited to supervise the construction of suitable pairwise affinity values. Our affinity definition, which is a function of both the proximity via Euclidean distance and the "similarity" between local estimated tangent subspaces, will improve the reliability of the affinity value between the pairwise points. Then, we can apply spectral methods to achieve a better performance.

It is worthwhile to highlight several aspects of our contributions here.

1) Though it is not a new finding that a poor affinity matrix leads to poor SC performance, as far as we know, we are the first to relate SC to a classical manifold learning method, i.e., laplacian eigenmaps (LEM) [20], to explain its poor performance when there are significant intersections among different clusters. Our analysis reveals that the performance of SC relies on the constructed affinity matrix, and spectral methods are able to work well when the pairwise points belonging to different clusters have a relatively low affinity value.
2) When the unlabeled data points are lying on or close to multiple smooth low-dimensional manifolds, we take advantage of additional geometric information presented in the sampled data to give a general framework and a concrete criterion to construct the expected affinity matrix and then find the correct clusters.
3) Our proposed method can be successfully applied to the general case of manifold clustering. Specifically, it is able to handle situations where the manifolds on which the data points lie are: (a) linear and/or nonlinear, and (b) intersecting and/or not intersecting.

The rest of this paper is organized as follows. Section II gives a brief review and analysis of classical SC. Section III presents the spectral multi-manifold clustering (SMMC) method. In Section IV, we report the experiments. Finally, we conclude and raise several future issues in Section V.

II. Review and Analysis on Classical SC

Given a set of unlabeled data points $X = \{x_i \in \Re^D, i = 1, \ldots, N\}$, the aim of cluster analysis is to assign these points into k disjoint subsets such that data points belonging to the same cluster are "similar" and data points belonging to different clusters are "dissimilar."

A. Brief Review on Classical SC

In SC, a neighborhood graph on the data points is first constructed based on some criteria, such as the fully connected graph or the K-nearest neighbor graph [6]. Then, a weighted affinity matrix $W \in \Re^{N\times N}$ is defined, whose (i, j) element w_{ij} reflects the similarity between x_i and x_j. Specifically

$$w_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2/2\sigma^2) & i \neq j, \\ 0 & i = j \end{cases} \tag{1}$$

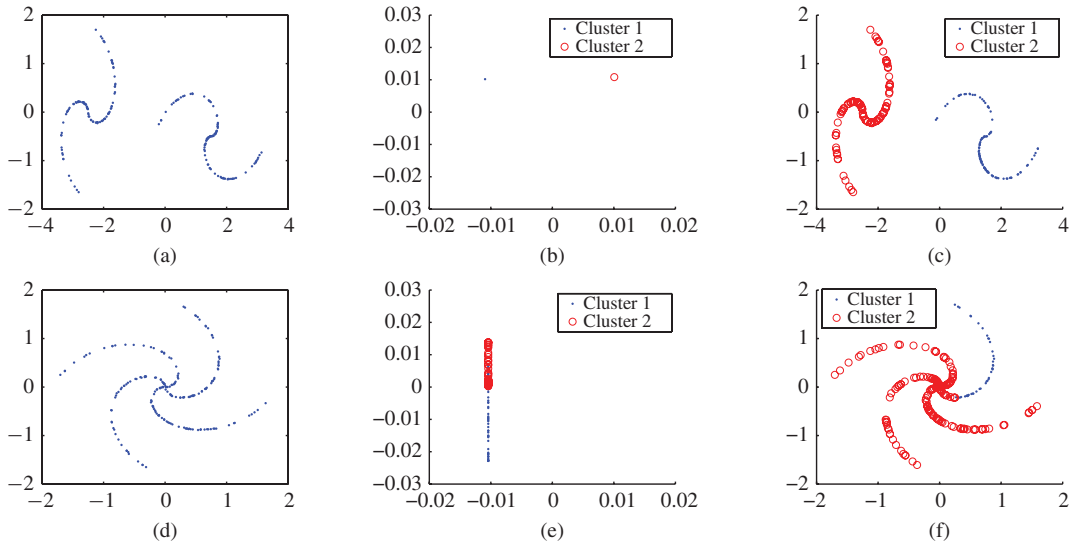

Fig. 2. Different performances of classical SC on well-separated clusters and significantly intersecting clusters. (a) Well-separated clusters. (b) Images in the embedded space. (c) Grouping results of spectral clustering. (d) Significantly intersecting clusters. (e) Images in the embedded space. (f) Grouping results of spectral clustering.

where $\|\cdot\|$ denotes the Euclidean norm. Given a neighborhood graph with affinity matrix W, a simple but efficient clustering criterion is the *normalized cut (Ncut)* [5] which is defined as

$$Ncut(X_1, \ldots, X_k) \triangleq \frac{1}{2}\sum_{i=1}^{k}\frac{W(X_i, \bar{X}_i)}{vol(X_i)} \tag{2}$$

where $X_1, \ldots, X_k$ is a partition of X ($X_1 \cup \cdots \cup X_k = X$, $X_i \cap X_j = \emptyset$, $i \neq j$ and $X_i \neq \emptyset$, $i = 1, \ldots, k$), $W(A, B) \triangleq \sum_{x_i \in A, x_j \in B} w_{ij}$, $vol(A) \triangleq \sum_{x_i \in A, j \in \{1,\ldots,N\}} w_{ij}$ and $\bar{A}$ is the complement of A. Noting that small values of $W(X_i, \bar{X}_i)$ indicate that X_i is a well-defined cluster and large values of $vol(X_i)$ indicate clusters of considerable size, it follows that lower values of Ncut correspond to better clusterings. Thus the aim is to minimize Ncut.

Following some algebraic formulations, it turns out [6] that minimizing normalized cut can be equivalently recast as

$$\begin{aligned} \min_{X_1,\ldots,X_k} \quad & Tr(H^{\mathrm{T}}(E - W)H) \\ \text{s.t.} \quad & H^{\mathrm{T}}EH = I \end{aligned} \tag{3}$$

where E is an $N \times N$ diagonal matrix with $E_{ii} = \sum_j w_{ij}$, I is the identity matrix, $H \in \Re^{N \times k}$ is a specific discrete matrix, and Tr denotes the trace of a matrix.

Unfortunately, solving the above discrete optimization problem is NP-hard. To make it tractable, an efficient relaxation is adopted so as to solve a real-valued problem instead of a discrete-valued one [5], [6]. This is done by computing the first k generalized eigenvectors $u_1, \ldots, u_k$, corresponding to the k smallest eigenvalues, of the generalized eigenproblem

$$(E - W)u = \lambda Eu. \tag{4}$$

Finally, K-means method is performed on the row vectors of $U = [u_1, \ldots, u_k] \in \Re^{N \times k}$ to obtain the clusters.

B. Further Analysis on Classical SC

In the following, we discuss how the performance of SC relies on the constructed affinity matrix W. That is, SC methods can work well when the points belonging to different clusters have (relatively) low weight.

When the k clusters are well separated, this property of the affinity matrix is easy to achieve by carefully tuning the nearest parameter K or the scaling parameter σ [6]. In the "ideal" case where all pairs of points belonging to different clusters have weight $w = 0$, each cluster degenerates to a connected component. In this case, spectral methods will map all the points in the same cluster into a single point in $\Re^k$ and there are k mutually orthogonal points in this embedded space [3], [6]. Then, the followed K-means will easily group the points to the true clusters.

On the contrary, when there are significant intersections among different clusters, the affinity matrix will be corrupted with poor pairwise affinity values, i.e., affinity value between the pair of points belonging to different clusters but near the intersection is always high since their Euclidean distance is small. For simplicity of discussion, we assume the data consist of two intersecting clusters. In this case, they look like a "tight" cluster and thus will be hard to split into two subsets. In fact, the generalized eigenproblem $(E - W)u = \lambda Eu$ performed on two intersecting clusters is the same as the optimal objective of LEM [20], which is one of the classical manifold learning methods. It is easy to show that the first two generalized eigenvectors consist of a component indicator vector with equal value and an embedding vector which optimally preserves local neighborhood information of the original data [20]. As a result, the intersecting structure of the two different clusters is preserved in the embedded space. Then, the followed K-means fails to extract useful structure.

In Fig. 2, we show different performances of classical SC on well-separated clusters and significantly intersecting clusters. The data are generated from the same two spirals. In Fig. 2(a),

they are well separated, while Fig. 2(d) they intersect at the origin. Fig. 2(b) and (e) show their images in the embedded space, respectively. For the purpose of visualization, we use the true class information to label the samples in these two subfigures. As can be seen, the two clusters collapse into two mutually orthogonal points in Fig. 2(b), while they still mix together in Fig. 2(e). Fig. 2(c) shows that classical SC gives perfect grouping to the well-separated clusters, while it is poor for the significantly intersecting case [see Fig. 2(f)].

III. SMMC

Given a set of unlabeled data points $X = \{x_i \in \Re^D, i = 1, \ldots, N\}$ that stem from $k > 1$ distinct smooth manifolds $\{\Omega_j \subseteq \Re^D, j = 1, 2, \ldots, k\}$, some of which possibly intersecting with each other [14], the objective of manifold clustering is to assign each sample to the manifold it belongs to. In the sequel, all manifolds are assumed to have the same dimension d ($0 < d < D$), which, together with their number k, is known. An example is illustrated in Fig. 1 for $k = 3$, $D = 3$, and $d = 2$.

It seems that we can address the hybrid nonlinear manifold clustering problem using the following strategy. First, run classical SC [3], [5], [6] to reveal all the connected components. Then, for each of the connected component run, e.g., K-manifolds [13], to further unravel the intersecting clusters. However, it is difficult to judge whether a connected component is composed of a single manifold or multiple intersecting manifolds, and if it is composed of multiple intersecting manifolds, how many manifolds are there in this connected component. Indeed, the difficulties encountered by this simple design partially motivate us to propose the SMMC method.

A. Affinity Matrix

As has been discussed in Section II-B, spectral-based clustering methods are able to work well when the affinity values of the points belonging to different clusters are relatively low. However, the traditional affinity matrix based on a radial distance between the points is not suitable. Thus, our basic idea to deal with hybrid nonlinear manifold clustering is to incorporate some natural and reliable geometric information of the sampled data to construct a suitable affinity matrix and then to find the correct clusters.

Though the data are globally lying on or close to multiple smooth nonlinear manifolds, locally, each data point and its neighbors are lying on a linear patch of the manifold [9], [10]. Moreover, the local tangent space at each point provides a good low-dimensional linear approximation to the local geometric structure of the nonlinear manifold [21]. Finally, as will become clear shortly, at the intersection areas of different manifolds, the points on the same manifold have similar local tangent spaces while the points from different manifolds have dissimilar tangent spaces. Therefore, this type of local geometric information can be used to help the construction of affinity matrix.

For faraway points, it is difficult to judge whether they are in the same manifold or not by using only local geometric information, and so we focus on local regions. Intuitively, for two points in the same local area, if 1) they are close to each other and if 2) they have similar local tangent spaces, then they will have a high chance to lie on the same manifold. For nearby points, if they have different local tangent spaces, such as two points at the intersection of S-curve and the vertical affine subspace (see Fig. 1), they are very likely to come from different manifolds. Thus, we should consider two affinity functions between two points x_i and x_j, with one of them defined as a function of their corresponding local tangent spaces (named *structural similarity* p_{ij}), and the other defined via the Euclidean distance $q_{ij} = q(\|x_i - x_j\|)$ (named *local similarity*). Then, these two functions are fused together to give the final affinity value

$$w_{ij} = f(p_{ij}, q_{ij}) \tag{5}$$

where f is a suitable fusion function. It should be noted that, in order to have the expected property of the affinity matrix, f should be a monotonically decreasing function of the Euclidean distance and, at the same time, a monotonically increasing function of the similarity of two tangent spaces.

Now, we give the concrete formulation of p, q, and f used in our current method. Suppose the tangent space at x_i ($i = 1, \ldots, N$) is Θ_i, the structural similarity between the local tangent spaces of two points x_i and x_j can be defined as

$$p_{ij} = p(\Theta_i, \Theta_j) = \left(\prod_{l=1}^{d} \cos(\theta_l)\right)^{\circ}. \tag{6}$$

In (6), $o \in \mathrm{N}^+$ is an adjustable parameter. $0 \leq \theta_1 \leq, \ldots, \leq \theta_d \leq \pi/2$ are a series of principal angles [22] between two tangent spaces Θ_i and Θ_j, defined recursively as

$$\cos(\theta_1) = \max_{\substack{u_1 \in \Theta_i, v_1 \in \Theta_j \\ \|u_1\| = \|v_1\| = 1}} u_1^T v_1 \tag{7}$$

and

$$\cos(\theta_l) = \max_{\substack{u_l \in \Theta_i, v_l \in \Theta_j \\ \|u_l\| = \|v_l\| = 1}} u_l^T v_l, \quad l = 2, \ldots, d \tag{8}$$

where ${u_l}^{\mathrm{T}} u_i = 0$, ${v_l}^{\mathrm{T}} v_i = 0$, $i = 1, \ldots, l-1$.

The local similarity is simply defined as

$$q_{ij} = \begin{cases} 1, & \text{if } x_i \in Knn(x_j) \text{ or } x_j \in Knn(x_i), \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $Knn(x)$ denotes K nearest neighbors of x.

Finally, these two functions are simply multiplied together to give the affinity value

$$w_{ij} = p_{ij} q_{ij} = \begin{cases} \left(\prod_{l=1}^{d} \cos(\theta_l)\right)^{\circ}, & \text{if } x_i \in Knn(x_j) \\ & \text{or } x_j \in Knn(x_i), \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

It is easy to check that the affinity value defined in (10) has the expected property, i.e., the points belonging to different clusters/manifolds have relatively low value. This is because, when the pair of points from different manifolds are far from each other, they will have the affinity value of 0. But, when they are close to the intersection of different manifolds, they

will have dissimilar local tangent spaces which will also have a relatively low affinity when the tuning parameter o is large enough. Thus, when spectral method is applied to this matrix, a better performance is expected.

An unresolved problem in the above formulation is how to effectively approximate the local tangent space at each sample. In the following section, we will discuss this issue in detail.

B. Local Tangent Space

Generally, the tangent space at each point can be constructed from the local neighborhood of the given sample [17], [21]. Specifically, given a point x and its n closest neighbors $N(x) = \{x^1, \ldots, x^n\}$ in Euclidean space, the local geometric information around x is captured by its local sample covariance matrix $\sum_x$, which is defined as

$$\sum\nolimits_x = \frac{1}{n}\sum_{i=1}^{n}(x^i - u_x)(x^i - u_x)^{\mathrm{T}} \tag{11}$$

where $u_x = 1/n\sum_{i=1}^{n} x^i$.

Then, the local tangent space Θ_x at x is approximated by the d left singular vectors of $\sum_x$ corresponding to its d largest singular values. That is, suppose the singular value decomposition of $\sum_x$ is

$$\sum\nolimits_x = \begin{bmatrix} U_d & \tilde{U}_d \end{bmatrix} \begin{bmatrix} \sum_d & 0 \\ 0 & \tilde{\sum}_d \end{bmatrix} \begin{bmatrix} V_d & \tilde{V}_d \end{bmatrix}^{\mathrm{T}} \tag{12}$$

where $\begin{bmatrix} U_d & \tilde{U}_d \end{bmatrix} \in \Re^{D\times D}$ is an orthogonal matrix and $U_d \in \Re^{D\times d}$. Then, we have

$$\Theta_x = span(U_d). \tag{13}$$

Unfortunately, when two points x and y are very close to each other, their local tangent spaces Θ_x and Θ_y according to (13) will be very similar even if they are from different manifolds. This is because that their local neighborhoods $N(x)$ and $N(y)$ based on the Euclidean distance will strongly overlap, leading to similar local covariance matrices $\sum_x$ and $\sum_y$. Therefore, this traditional definition of local tangent space will not work well for hybrid nonlinear modeling. In the later part of this subsection, we give an efficient and effective technique to approximate the local tangent space at each point.

Our basic idea is based on the fact that global nonlinear manifolds can locally be well approximated by a series of local linear manifolds [9], [10], and principal component analyzers [23] can successfully pass across the intersecting linear manifolds. Moreover, the points approximated by the same linear analyzer usually have similar local tangent spaces which can also be well approximated by the principal subspace of the local analyzer. Therefore, we can train many local linear analyzers to approximate the underlying manifolds, then the local tangent space of a given sample is determined by the principal subspace of its corresponding local analyzer.

In this paper, we train M mixtures of probabilistic principal component analyzers (MPPCA) [23], where each analyzer is characterized by the model parameters $\theta_m = \{\mu_m, V_m, \sigma_m^2\}$, $m = 1, \ldots, M$, where $\mu_m \in \Re^D$, $V_m \in \Re^{D\times d}$ and σ_m^2 is a scalar. It should be noted that M refers to all local linear manifolds that are used to approximate all the linear or nonlinear manifolds that underlie the dataset. Under the mth analyzer, a D-dimensional observed data vector x is related to a corresponding d-dimensional vector of latent variable y as

$$x = V_m y + \mu_m + \varepsilon_m \tag{14}$$

where μ_m is a robust mean of data, and the latent variable y and the noise ε_m are Gaussian functions as $y \sim \mathcal{N}(0, I)$ and $\varepsilon_m \sim \mathcal{N}(0, \sigma_m^2 I)$, respectively. Then, the marginal distribution of x is given by

$$\begin{aligned} &p(x\,|m\,) \\ &= (2\pi)^{-D/2}|C_m|^{-1/2}\exp\left\{-\frac{1}{2}(x-\mu_m)^{\mathrm{T}}C_m^{-1}(x-\mu_m)\right\} \end{aligned} \tag{15}$$

where the model covariance is

$$C_m = \sigma_m^2 I + V_m V_m{}^{\mathrm{T}}. \tag{16}$$

We can learn all the model parameters μ_m, V_m, and σ_m^2 by using EM method to maximize the log-likelihood of observing the dataset $X = \{x_i, i = 1, \ldots, N\}$

$$\mathcal{L} = \sum_{i=1}^{N}\ln\left\{\sum_{m=1}^{M}\pi_m\, p(x_i\,|m\,)\right\} \tag{17}$$

where (15) and (16) are utilized and π_m is the mixing proportion, subject to $\pi_m \geq 0$ and $\sum_{m=1}^{M}\pi_m = 1$. Specifically, the principal steps of the EM learning method [23] are as follows.

E-Step: Using the current set of parameters $\theta_m = \{\mu_m, V_m, \sigma_m^2\}$, compute

$$R_{im} = \frac{\pi_m\, p(x_i\,|m\,)}{\sum_{m=1}^{M}\pi_m\, p(x_i\,|m\,)} \tag{18}$$

$$\pi_m^{new} = \frac{1}{N}\sum_{i=1}^{N}R_{im} \tag{19}$$

$$\mu_m^{new} = \frac{\sum_{i=1}^{N}R_{im}x_i}{\sum_{i=1}^{N}R_{im}}. \tag{20}$$

M-Step: Re-estimate the parameters V_m and σ_m^2 as

$$V_m^{new} = S_m V_m(\sigma_m^2 I + T_m^{-1}V_m^T S_m V_m)^{-1} \tag{21}$$

$$(\sigma_m^2)^{new} = \frac{1}{d}tr[S_m - S_m V_m T_m^{-1}(V_m^{new})^T] \tag{22}$$

where

$$S_m = \frac{1}{\pi_m^{new}N}\sum_{i=1}^{N}R_{im}(x_i - \mu_m^{new})(x_i - \mu_m^{new})^T \tag{23}$$

$$T_m = \sigma_m^2 I + V_m^T V_m. \tag{24}$$

Note that we use K-means to initialize EM. Finally, sample x_i is grouped into the jth local analyzer subject to

$$p(x_i\,|j\,) = \max_m\ p(x_i\,|m\,) \tag{25}$$

and the local tangent space of x_i is then given by

$$\Theta_i = span(V_j). \tag{26}$$

Algorithm SMMC

Input: Data set X, number of clusters k, dimension of the manifolds d, number of mixture models M, number of neighbors K, tuning parameter o.

Process:

1: Train M d-dimensional local linear manifolds by using MPPCA to approximate the underlying manifolds;

2: Determine the local tangent space of each point;

3: Compute pairwise affinity between two local tangent spaces using (6);

4: Compute the affinity matrix $W \in \Re^{N \times N}$ using (10);

5: Compute the diagonal matrix E with $E_{ii} = \sum_j w_{ij}$;

6: Extract the first k generalized eigenvectors $u_1, \ldots, u_k$ of $(E - W)u = \lambda Eu$;

7: Apply K-means to cluster the row vectors of U in $\Re^k$.

Output: A partition of the data into k disjoint clusters.

Fig. 3. Pseudocode of SMMC.

We can estimate the reconstruction error of using M local linear analyzers to approximate the underlying manifolds as

$$error(M) = \sum_{j=1}^{M} \sum_{l=1}^{N_j} \left(x_l^j - \mu_j\right)^{\mathrm{T}} \left(I - V_j V_j^{\mathrm{T}}\right) \left(x_l^j - \mu_j\right) \tag{27}$$

where $x_l^j, l = 1, \ldots, N_j$ are the N_j ($\sum_{j=1}^{M} N_j = N$) points which are grouped into the jth local analyzer.

C. Final Algorithm and Complexity Analysis

After the local tangent space of each point has been estimated, we can compute the affinity matrix W as described in Section III-A. Then, spectral method is used on this matrix to give the final clusters. The pseudocode of our method, i.e., SMMC, is shown in Fig. 3.

Let us further analyze the computational complexity of SMMC. The complexity of SMMC is composed of three parts: estimating the local tangent space of each sample, computing the affinity matrix W, and applying spectral method on W. The N local tangent spaces Θ_i, $i = 1, \ldots, N$ are estimated using EM on M MPPCA, with K-means to initialize the model parameters. This procedure has complexity of $O(NDM(t_1 + dt_2))$, where t_1 and t_2 are the number of iterations before K-means and EM to convergence, respectively. In the second part, the complexity of computing the affinity values between any two local tangent spaces is $O(N^2Dd^2)$. For the K nearest neighbor search for each sample, the complexity is $O((D + K)N^2)$. The third part uses spectral method on W to project the data onto a k-dimensional embedded space, and then K-means is used to group the data into k clusters. The complexity of the generalized eigenvector problem is $O((N + k)N^2)$, and the complexity of K-means on the k-dimensional embedded space for t_3 iterations is $O(Nk^2t_3)$. Thus, the total time complexity of the SMMC method is

$$O(N^3 + N^2(Dd^2 + K + k) + N(DM(t_1 + dt_2) + k^2t_3)). \tag{28}$$

Since the number of iterations before the convergence of K-means and EM are usually small (less than 50), $d < D$, $K \ll N$, $k \ll N$, and $M \ll N$, the complexity of SMMC is primarily determined by N and D.

IV. Experiments and Applications

In this section, we test the performance of SMMC using a series of synthetic and real-world datasets. We also compare our method with several state-of-the-art techniques. The codes and data are available online at http://lamda.nju.edu.cn/code_SMMC.ashx.

A. Experimental Setting and Evaluation Metric

All of our experiments are performed on a PC configured with an Intel Dual Core-based system with $4 * 2$ GHz CPU and 8 GB RAM memory under MATLAB platform.

In all the following experiments, we use the results of K-means as the baseline for comparison. Among different versions of classical SC, the unsymmetrical normalized SC [5] is used following von Luxburg's suggestion [6]. As in [3], the scaling parameter σ^2 in SC is searched over a large range from 10^{-5} to 10^5 at $10^{0.5}$ step, and the best result is reported.

Clustering accuracy is used as the evaluation criterion to assess the performance of clustering, which is computed by using the true class labels associated with each of the datasets. Clearly, higher accuracy implies better performance. Clustering accuracy is defined as the maximum classification accuracy among all possible alignments

$$\text{clustering accuracy} = \max_{align} \sum_{i=1}^{N} \frac{\delta(t_i = c_i)}{N} \tag{29}$$

where t_i is the true label and c_i is the obtained cluster label of x_i, $\delta(\cdot)$ is the delta function.

B. Simulations on Synthetic Data

In this section, we conduct a series of simulations on synthetic datasets with controllable structures to examine the effectiveness of the SMMC method and to compare our results with those obtained by using state-of-the-art methods.

1) Effectiveness of MPPCA on Handling Intersection Regions: We first show how reliably the proposed MPPCA can handle the intersection regions, i.e., effectively separating the different sides of the manifolds near the intersection into different probabilistic analyzers. We conduct experiments on the hybrid data (see Fig. 1) and the three-line data [see Fig. 4(e)] and the results are shown in Fig. 4. Different local linear patches corresponding to the M probabilistic analyzers are shown in Fig. 4(a) and (f) through different colors and symbols. To see more clearly how MPPCA separates the points near the intersection into different local linear patches, we magnify the intersection regions in Fig. 4(b)–(d) and (g) and (h), respectively. As we can see from this figure, MPPCA obtains a collection of good local patches (each ideally from one manifold) to break the intersection regions.

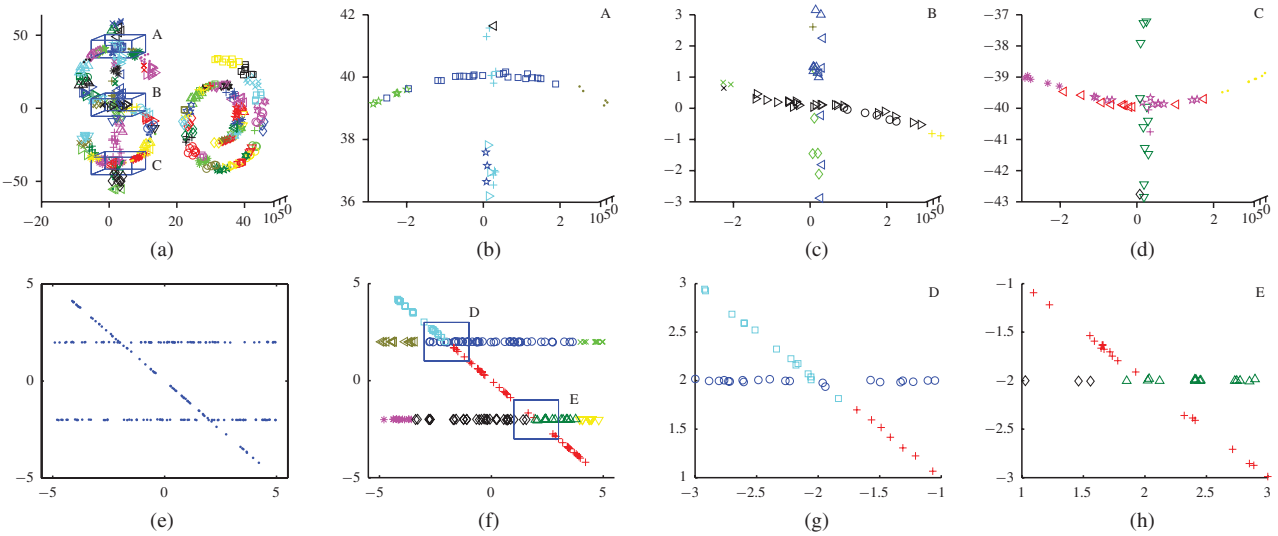

Fig. 4. Illustration of the effectiveness of MPPCA on handling intersection regions. (a) Applying MPPCA on the hybrid data with $M = 110$. (b)–(d) Magnified images of the intersection regions A, B, and C, respectively. (e) Three lines in $\Re^2$. (f) Applying MPPCA on the three-line data with $M = 9$. (g) and (h) Magnified images of the intersection regions D and E, respectively.

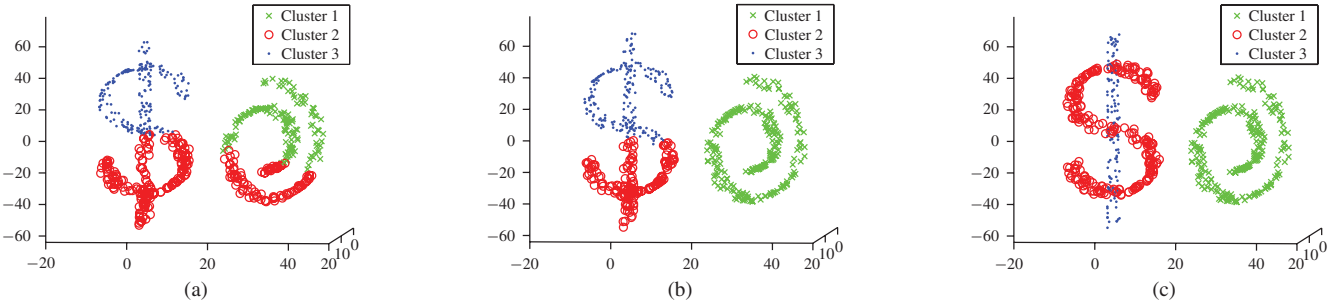

Fig. 5. Grouping results using different methods on the hybrid data. Different clusters are shown via different colors and symbols. (a) K-means. (b) Spectral clustering. (c) SMMC.

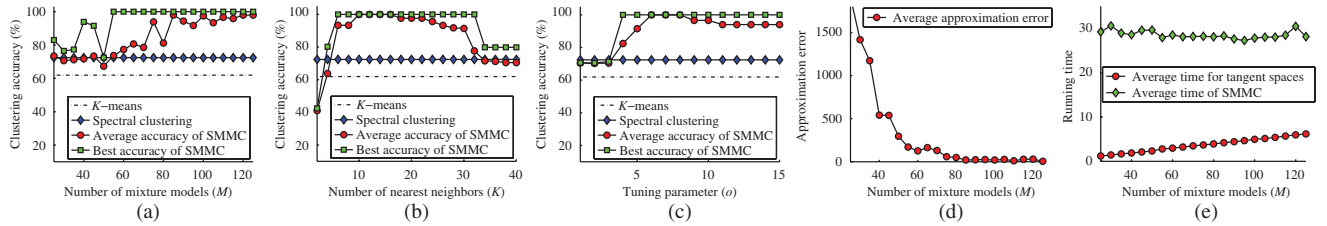

Fig. 6. Influence of parameters for SMMC. (a) Different M values. (b) Different neighbors. (c) Different o values. (d) Reconstruction error. (e) Running time.

2) Visual Comparison on Synthetic Hybrid Data: In this section, we visually compare the performance of SMMC with K-means and SC to the task of detection of multiple hybrid manifolds. The results performed on the hybrid data of Fig. 1 are shown in Fig. 5. As we can see from Fig. 5(c), SMMC reliably finds clusters consistent with different manifolds. Traditional SC works better than K-means, which can correctly partition the separated Swiss roll from the other two manifolds. However, it still confuses points from the two intersecting manifolds.

3) Model Selection: There are three adjustable parameters in the SMMC method, i.e., M, K, and o. In this section, we study how these parameters influence the performance of SMMC and then give some guidelines for their selection. The results performed on the hybrid manifolds of Fig. 1 are plotted in Fig. 6. The results are shown in two manners: 1) the average clustering accuracy of SMMC (which is averaged over 10 independent trials) is shown to analyze the clustering performance from the viewpoint of statistics and 2) the best clustering accuracy of SMMC is compared with the best results of K-means and classical SC, with the aim to show the efficacy of our method in dealing with hybrid nonlinear manifold clustering.

From these results, we make several interesting observations. 1) The performance of SMMC depends more on the number of mixture models. However, it can be seen clearly from Fig. 6(a) that the more the number of mixture models, the higher the clustering accuracy. This phenomenon can be

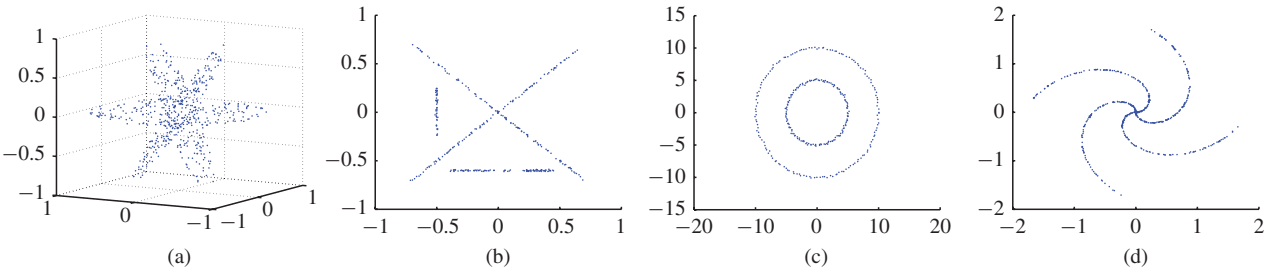

Fig. 7. Data samples with a wide variety of structures. (a) Three planes (each with 400 samples). (b) Five affine subspaces, where there are two subspaces lying approximately in the same affine line but are well separated ($N = 700$). (c) Two circles (each with 300 points). (d) Two spirals (each with 500 samples).

explained from Fig. 6(d), as the number of mixture models increases, the average approximation error decreases, which means that the approximation to the local linear patches of the manifolds is more faithful. Thus, the estimation for the local tangent space of each sample is more reliable, leading to better performance of SMMC which relies on the estimation of these tangent spaces. 2) The performance of SMMC is robust for a range of values of K as long as it does not include neither too small nor too large values, which is consistent with many existing observations (see [10], [12]). The reason is that there may be many disconnected subclusters when K is too small, while the local restriction will be lost when it is too large. 3) SMMC works well when the tuning parameter o is large enough. The reason is that the larger o, the better separability of points from different manifolds since, for $x < 1$, x° gets closer to zero for larger values of o. 4) To complete our theoretical analysis of computational complexity, we also show the average running time of SMMC, as well as the average time for estimating the tangent spaces. As we can see from Fig. 6(e), the time for estimating the tangent spaces increases nearly linearly with the number of analyzers M, which is consistent with our theoretical analysis. However, the total running time of SMMC seems to be independent of the number of mixture models. This observation is reasonable, since the time complexity of SMMC is dominated by computing the affinity matrix W and performing spectral analysis on W, which are independent of the number of mixture models.

From the above observations, we can give some rules of thumb to set up these three parameters. As a general recommendation, we suggest to work with $M = \lceil N/(10d) \rceil$, $K = 2\lceil \log(N) \rceil$[1] and $o = 8$. When all the manifolds are approximately linear, we can work with a relative small M, e.g., $M = 3k$. Moreover, we recommend a search for the optimal values of these parameters as $M \in [\lceil N/(10d) \rceil, \lceil N/(2d) \rceil]$, $K \in \left[\lceil \log(N) \rceil, 3\lceil \log(N) \rceil\right]$, and $o \in [4, 12]$. However, it should be noted that, for general datasets, the optimal values of these parameters should be selected according to noise level, the distribution of the samples, etc.

4) Comparison with State-of-the-Art Methods: Since a considerable amount of work has been done on manifold clustering, it is interesting and meaningful to compare our proposed SMMC with these state-of-the-art methods.

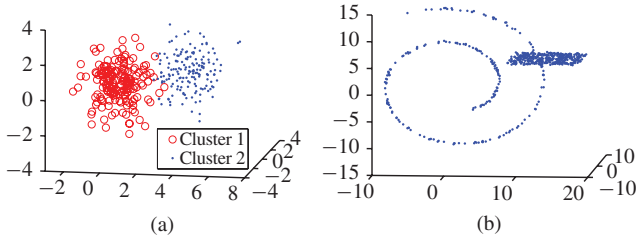

Fig. 8. Data samples drawn from (a) two Gaussian "cloud" clusters of significant overlap (each cluster has 500 points) and (b) a spiral (400 samples) and a plane (800 samples).

We compare SMMC with the following methods: K-means [2], GPCA [16], K-planes [18], [19], LSA [17], SCC [24], K-manifolds [13], SC [3], [5], and mumCluster[2] [14]. We compare all these methods on five synthetic datasets with different characteristics and complexities [see Fig. 7(a)–(d) and 1]. It should be noted that different levels of noise have been added in all these synthetic datasets. So, all the points lie around the underlying manifolds. It is expected that some methods cannot work well in some scenarios. For example, K-means, SC, and linear manifold clustering methods cannot work well on intersecting nonlinear manifolds. However, since these methods are quite famous, and in order to confirm that our proposed method outperforms them, we also include them in the comparison.

The average accuracy over 30 independent runs and the corresponding standard deviations followed by the highest clustering accuracy of different methods are tabulated in Table I. The average running time of these methods are also shown in Table I, which give a rough indication of their time complexity.

These experiments reveal a number of interesting facts. 1) Linear manifold clustering methods such as GPCA, K-planes, LSA, and SCC work well on multiple linear manifolds (e.g., the three-planes dataset) but may deteriorate on clustering affine subspaces, which is in line with the observations reported in [24]. Moreover, they fail to deliver good performance when faced with inherent nonlinear problems, such as the two-spirals dataset and the hybrid dataset.

[1]Generally, M should be set to make sure that the approximation to the local linear patches of the manifolds is faithful enough. It seems that we can set $M = \lceil N/(d+1) \rceil$ since $d+1$ points define the position of a d-dimensional plane. However, this setting is sensitive to noise. K should be set to make sure that the similarity graph of the same manifold is connected, which should be chosen on the order of $\log(N)$ as shown in [6].

[2]In our current execution, we slightly modify the original mumCluster code to have a given number of clusters.

TABLE I

COMPARISON OF THE CLUSTERING ACCURACY (MEAN ± STD. FOLLOWED BY THE HIGHEST ACCURACY IN THE PARENTHESES) AND THE AVERAGE COMPUTATION TIME (IN SECONDS) ON FIVE SYNTHETIC DATASETS. THE HIGHEST CLUSTERING ACCURACY ON EACH DATA SET IS MARKED BY A DIAMOND. THE BEST PERFORMANCE IS HIGHLIGHTED BY A STAR AND THE PERFORMANCES WITHOUT SIGNIFICANT DIFFERENCE WITH THE BEST PERFORMANCE ARE BOLDFACED (PAIRED t-TEST AT 95% SIGNIFICANCE LEVEL)

Dataset	Three-planes		Five-affine-subspaces		Two-circles		Two-spirals		Hybrid data	
	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time
K-means	0.366 ± 0.012 (0.391)	0.02	0.580 ± 0.022 (0.606)	0.01	0.501 ± 0.001 (0.505)	0.01	0.538 ± 0.003 (0.582)	0.01	0.563 ± 0.042 (0.619)	0.02
GPCA	0.983 ± 0.000 (0.983)	0.01	0.787 ± 0.000 (0.787)	0.01	0.500 ± 0.000 (0.500)	0.01	0.524 ± 0.000 (0.524)	0.01	0.458 ± 0.000 (0.458)	0.02
K-planes	**0.944 ± 0.120** (0.983)	0.01	0.789 ± 0.095 (0.859)	0.01	0.502 ± 0.002 (0.505)	0.01	0.549 ± 0.001 (0.552)	0.01	0.355 ± 0.010 (0.394)	0.01
LSA	0.969 ± 0.000 (0.969)	71.76	0.593 ± 0.000 (0.593)	21.10	0.505 ± 0.000 (0.505)	15.11	0.500 ± 0.000 (0.500)	45.52	0.574 ± 0.000 (0.574)	378.87
SCC	0.984 ± 0.001 (0.987)	1.95	0.782 ± 0.069 (0.949)	3.06	0.502 ± 0.002 (0.507)	0.56	0.549 ± 0.021 (0.596)	0.71	0.527 ± 0.096 (0.648)	4.77
SC	0.408 ± 0.000 (0.408)	3.29	0.779 ± 0.090 (0.831)	2.39	**1.000 ± 0.000*** (1.000)$^\diamond$	1.38	0.588 ± 0.000 (0.588)	4.29	0.724 ± 0.000 (0.724)	3.41
K-manifolds	0.746 ± 0.159 (0.953)	837.21	0.469 ± 0.066 (0.590)	144.60	0.519 ± 0.022 (0.593)	59.39	0.765 ± 0.193 (0.968)	261.39	0.411 ± 0.021 (0.460)	4405.11
mumCluster	0.889 ± 0.161 (0.984)	6.80	0.830 ± 0.058 (0.991)	4.51	**1.000 ± 0.000*** (1.000)$^\diamond$	1.48	**0.885 ± 0.000*** (0.885)	5.97	**0.986 ± 0.000** (0.986)	42.73
SMMC	**0.986 ± 0.003*** (0.993)$^\diamond$	3.50	**0.945 ± 0.089*** (0.994)$^\diamond$	1.73	**1.000 ± 0.000*** (1.000)$^\diamond$	2.04	**0.859 ± 0.127** (0.996)$^\diamond$	5.41	**0.993 ± 0.037*** (1.000)$^\diamond$	25.11

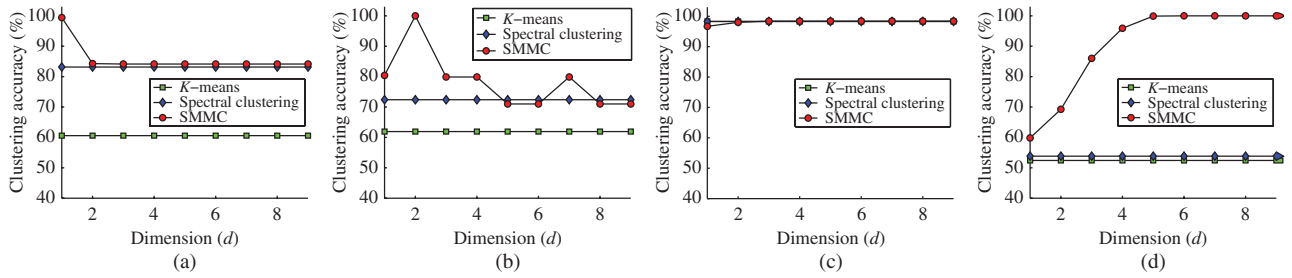

Fig. 9. Clustering performance with different dimensions. (a) Five-affine-subspaces data. (b) Hybrid data. (c) Two "cloud" clusters. (d) Two 6-D subspaces in $\Re^{10}$.

2) Though much work has shown that classical SC methods give promising results to multiple well-separated clusters [3], [5], [6] (e.g., the two-circles dataset), the experiments here reveal that SC cannot work well on intersecting clusters. 3) K-manifold delivers good performance to group samples generated from intersecting manifolds (i.e., the three-planes data and the two-spirals data). However, it works poorly when there are well-separated clusters (e.g., the hybrid data). The reason is that the estimation of geodesic distances is unfaithful in this case, leading to incorrect clustering. 4) The heuristic mumCluster and our SMMC not only give reasonable performance to linear manifolds and multiple intersecting nonlinear manifolds, but also work well on the more general hybrid nonlinear manifold clustering problem. Moreover, the proposed SMMC achieves the highest clustering accuracy and the best average accuracy almost in all cases. 5) Generally, as we expected, the running time of nonlinear methods is higher than that of the linear methods, while the cost of SMMC is comparable to that of the other nonlinear methods.

5) Further Studies: The proposed SMMC method starts by assuming that the intrinsic or true dimension d of the low-dimensional manifolds is known and all the underlying manifolds have the same dimension. It is interesting to see the performance of SMMC when a value of d other than its true value is used and when the underlying manifolds have mixed dimensions. First, we conduct a series of experiments on the five-affine-subspaces dataset, the hybrid dataset and two 3-D overlapped "cloud" clusters without significant manifold structure [see Fig. 8(a)]. In our experiments, we shift them to $\Re^{10}$ by padding zeros, and we also generate two 6-D linear subspaces in $\Re^{10}$. The clustering performance of our SMMC with different dimensions on these datasets is shown in Fig. 9. As can be seen, SMMC works better around the true dimension and it degenerates to regular SC when the given dimension d is far away from the ground-truth dimensionality. The reason is that the performance of SMMC relies on the estimation of the local tangent space of each sample, which is more reliable around the true dimension. These observations

TABLE II
PERFORMANCE OF DIFFERENT METHODS WHEN CLUSTERING DATA SAMPLED FROM MIXED DIMENSIONS. THE BEST ARE BOLDFACED

Dataset	Spiral-plane data	$(1, 2, 3) \in \Re^5$
K-means	88.67%	46.50%
GPCA	98.92%	96.08%
K-planes	67.58%	99.91%
LSA	69.08%	99.58%
SCC	99.00%	99.91%
SC	92.75%	80.58%
K-manifolds	73.83%	99.75%
mumCluster	86.17%	87.58%
SMMC	**99.75%**	**100.00%**
TPMM	98.58%	75.92%

suggest that it is better to estimate the true dimension of each manifold before using SMMC, which is left for future investigation.

We next compare SMMC with the other methods on clustering manifolds of mixed dimensions in order to further evaluate their performances. TPMM [15], which groups manifolds according to both dimensionality and density, is also tested. Two datasets are used in these experiments: one is the spiral-plane data used in [15] [see Fig. 8(b)], and the other [denoted by $(1, 2, 3) \in \Re^5$] is sampled from three linear subspaces in $\Re^5$ with dimension 1, 2, 3, respectively. Following Chen and Lerman [24], we set d as the maximum dimension for all the methods that can not be directly used to mixed dimensions (such as LSA, SCC, and our SMMC). Table II shows the clustering accuracy of these methods. As can be seen, SMMC is comparable to the other methods in these mixed cases.

Similar to SMMC, LSA [17] also uses principal angles for defining pairwise affinities, but in a different way: $p_{ij} = e^{-\sum_{l=1}^{d} \sin^2(\theta_l)}$. It would be interesting to compare the affinity in (6) with the other affinity definitions that use principal angles. Therefore, we replace the affinity in (6) with $p_{ij} = \sum_{l=1}^{d} \cos^2(\theta_l)$, $p_{ij} = e^{-\sum_{l=1}^{d} \sin^2(\theta_l)}$ in LSA and $p_{ij} = (e^{-\sum_{l=1}^{d} \sin^2(\theta_l)})^{o}$, which adds a tuning parameter o in LSA, to define the similarity of the tangent spaces, denoted as SMMC-cosine, SMMC-LSA, and SMMC-LSA2, respectively. Fig. 10 shows their clustering performance on five synthetic datasets. It can be observed that both SMMC and SMMC-LSA2 outperform the other methods, though SMMC is better than SMMC-LSA2 on three-planes dataset. Compared to SMMC-LSA, the good performance of SMMC-LSA2 suggests that it is beneficial to have an additional tuning parameter o to make the points belonging to different clusters have (relatively) low weights, which is expected for the success of SC as we have discussed in Section II.

C. Experiments on Real Data

In this section, we further test the performance of SMMC on several real datasets. We show that the presented SMMC method provides an effective tool and a generic solution for a broad range of practical problems.

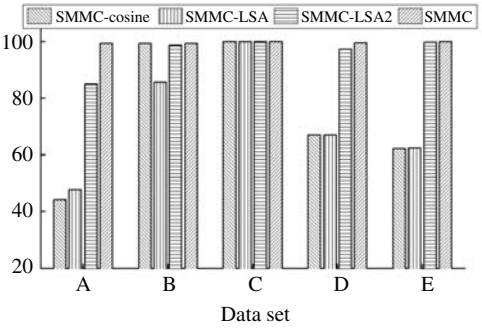

Fig. 10. Clustering performance with different definitions of subspace affinity on five synthetic datasets: (A) three-planes, (B) five-affine-subspaces, (C) two-circles, (D) two-spirals, and (E) hybrid data.

Fig. 11. Twenty images of 20 objects from the COIL-20 database.

TABLE III
COMPARISON OF CLUSTERING ACCURACY (MEAN ± STD.) ON THE COIL-20 DATABASE. THE BEST PERFORMANCE ON EACH DATASET IS MARKED BY A STAR AND THE PERFORMANCES WITHOUT SIGNIFICANT DIFFERENCE WITH THE BEST PERFORMANCE ARE BOLDFACED (PAIRED t-TEST AT 95% SIGNIFICANCE LEVEL)

Dataset	COIL-20	COIL-three-cars
K-means	0.338 ± 0.021	0.348 ± 0.005
GPCA	0.332 ± 0.000	0.375 ± 0.000
K-planes	0.531 ± 0.044	0.350 ± 0.020
LSA	0.643 ± 0.024	0.353 ± 0.004
SCC	0.671 ± 0.040	0.349 ± 0.007
SC	**0.693 ± 0.015**	0.366 ± 0.017
K-manifolds	0.299 ± 0.019	0.425 ± 0.027
mumCluster	0.523 ± 0.040	0.508 ± 0.029
SMMC	**0.707 ± 0.051***	**0.700 ± 0.056***

1) Clustering of 2-D Image Data: In the following experiments, the processed COIL-20 database[3] [25] is used. COIL-20 is a database with 1440 normalized gray-scale images of 20 objects, each object with 72 images. As can be seen from Fig. 11, the objects have a wide variety of complex geometric appearance and reflectance characteristics. The original resolution of these images is 32×32. Here, for computational efficiency, we project them onto the first 10 principal components.

Several studies (e.g., [26]) have shown that images of the same object reside on a 1-D nonlinear manifold. Moreover, the estimation of geodesic distances on the whole database fails even when we increase the number of neighbors to 60, which indicates that there are widely separated clusters in this database. Therefore, grouping different objects of this database can be seen as a hybrid nonlinear manifold clustering problem.

[3]Available at http://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php.

TABLE IV
CLUSTERING ACCURACY (%) OF DIFFERENT METHODS ON THE HOPKINS 155 MOTION SEGMENTATION DATABASE. THE BEST ARE BOLDFACED

Method	Accuracy	Two motions				Three motions				All
		Articulated (11)	Checkerboard (78)	Traffic (31)	All (120)	Articulated (2)	Checkerboard (26)	Traffic (35)	All (35)	All (155)
GPCA	Mean	93.06	92.72	98.01	94.12	79.65	73.56	90.53	77.30	90.32
	Median	**100.00**	97.22	**100.00**	98.84	79.65	71.39	97.56	76.69	96.89
K-planes	Mean	92.95	87.62	91.04	88.99	69.11	85.74	96.83	87.01	88.55
	Median	**100.00**	89.66	**100.00**	95.92	69.11	84.16	**100.00**	89.84	94.67
LSA	Mean	84.92	76.45	79.42	77.99	80.51	62.56	68.61	64.80	75.01
	Median	87.90	75.97	78.50	78.71	80.51	62.39	64.78	64.33	74.47
SCC	Mean	98.96	99.26	99.95	99.41	97.34	**97.15**	99.93	**97.71**	**99.03**
	Median	**100.00**	**100.00**	**100.00**	**100.00**	97.34	**99.89**	**100.00**	**100.00**	**100.00**
SMMC	Mean	**99.30**	**99.34**	**99.99**	**99.51**	**98.40**	96.59	**100.00**	97.38	**99.03**
	Median	**100.00**	**100.00**	**100.00**	**100.00**	**98.40**	98.97	**100.00**	99.39	**100.00**

In the first experiment, the whole database is used. To execute K-manifolds in this case, we set the number of neighbors to 80 for estimating geodesic distances. Moreover, GPCA fails to return any result within a reasonable response time (i.e., 2 days), because of its high computational complexity. Thus, following the authors' suggestion [16], GPCA is performed on the first five principal components. The experimental results of different methods are tabulated in the second column of Table III. In the second experiment, three objects of the same topic, i.e., three different cars, are selected from the total of 20 objects. K-manifolds can work in this case since eight neighbors are enough to connect all the samples, which reveals that these three objects are very close to each other. The corresponding clustering accuracy of different methods are shown in the third column of Table III.

The results on this real database reveal a number of interesting observations. 1) There are several observations consistent with the experiments on the synthetic datasets. For example, linear manifold clustering methods, i.e., GPCA, K-planes, LSA, and SCC, cannot work well on this nonlinear database, and K-manifolds performs poorly due to the presence of separated clusters. 2) Inconsistent with the synthetic experiments, mumCluster does not perform as well as SMMC. The reason is that mumCluster is more heuristic and heavily relies on the correct local dimension estimation, so it will have difficulty when there is much noise. 3) It is easy to see that SMMC is superior to the other methods.

2) 3-D Motion Segmentation: In this section, we apply SMMC to motion segmentation [27], i.e., segment a video sequence into multiple spatiotemporal regions corresponding to different rigid-body motion objects.

A benchmark motion segmentation database, i.e., the Hopkins 155 motion database,[4] is used in this experiment. The database includes 155 motion sequences of indoor and outdoors scenes containing two or three motions, which can be divided into three main categories: articulated, checkerboard, and traffic sequences [27]. For each sequence, a set of feature points were tracked automatically with a tracker, and outliers in tracking were manually removed. The articulated sequences contain multiple objects moving dependently in 3-D space, while both checkerboard and traffic sequences contain multiple objects moving independently. The motion trajectories of these sequences lie in dependent or independent affine subspaces of dimension two or three. In this case, solving the motion segmentation problem is equivalent to the hybrid linear manifold clustering problem [16].

The proposed SMMC is compared with the linear manifold clustering methods (i.e., GPCA, K-planes, LSA, and SCC) which are specially effective for 3-D motion segmentation [16], [17], [24]. Since GPCA requires a low-dimensional space to work properly, we strictly follow the authors' suggestion [16] to reduce the dimension of the data $D = 2F$ to a 5-D ambient space. All the other methods are executed in the original D-dimensional space. Table IV reports the average and median clustering accuracy of these methods. It can be seen that the performance of the proposed SMMC is highly competitive to SCC, which is the best among the linear methods, while the mean of SMMC is the best.

3) Statistical Testing: To further investigate the clustering results, we conduct paired t-test and paired Wilcoxon rank sum test at 95% significance level using the COIL-20 database in Section IV-C.1, as well as five synthetic datasets used in Section IV-B.4. Thus, there are totally seven different datasets with a wide variety of geometric characteristics. The win/tie/loss counts of SMMC versus the other methods are summarized in Table V.

It can be observed from Table V that SMMC rarely loses to the other methods. Furthermore, though the counts of win/tie/loss can indicate whether SMMC is better than another method by checking if the number of wins is larger than the number of losses, we also conduct the sign test at 95% significant level on the corresponding t-test results and Wilcoxon rank sum test results to further investigate whether SMMC is "significantly" better than any other of the previous methods. The results indicate that SMMC is indeed significantly better than all the other compared methods.

In summary, this statistical testing validates the efficiency of SMMC on both synthetic datasets and real-world problems.

D. Discussion

In Sections IV-B and IV-C, several experiments on a large number of synthetic datasets with controllable structures and real-world problems were systematically performed to show

[4]Available at http://www.vision.jhu.edu/data/hopkins155.

TABLE V
WIN/TIE/LOSS COUNTS OF SMMC VERSUS THE OTHER METHODS, AFTER PAIRED t-TEST AND PAIRED WILCOXON RANK SUM TEST AT 95% SIGNIFICANCE LEVEL

	K-means	GPCA	K-planes	LSA	SCC	SC	K-manifolds	mumCluster
t-test	7/0/0	7/0/0	6/1/0	7/0/0	7/0/0	5/2/0	7/0/0	4/3/0
Wilcoxon rank sum test	7/0/0	7/0/0	7/0/0	7/0/0	7/0/0	6/1/0	7/0/0	5/2/0

the efficiency of our proposed SMMC. These experiments reveal several interesting points.

1) Generally, all the manifold clustering methods perform better than the baseline K-means, which is a classical centroid-based clustering method. This reveals that clusters usually have complicated structure and cannot be separated by hyperplanes. Also, these experiments indicate that it is a reasonable and promising generalization to regard a cluster as a group of points centered on a compact low-dimensional manifold, at least in some applications.
2) Linear methods, such as GPCA, K-planes, LSA, and SCC, deliver promising performance when the samples lie on linear manifolds but fail when faced with inherent nonlinear problems. Traditional SC methods based on pairwise distance cannot work well when there are intersecting clusters. On the contrary, K-manifolds is limited to intersecting manifolds because of the estimation of geodesic distances.
3) Both mumCluster and SMMC give promising results to the general problem of hybrid nonlinear manifold clustering. However, mumCluster is quite heuristic and relies heavily on the correct estimation of intrinsic dimension. Thus, mumCluster works well on the synthetic datasets with controllable structures, but has difficulty when there is much noise in the data, such as the COIL-20 database. On the other hand, the proposed SMMC performs the best almost in all cases.
4) Apart from GPCA which is a noniterative method that does not require initialization (this fact has been confirmed in our experiments since all the standard deviations of GPCA are zero), all the other methods need multiple starts in order to obtain good solution. The experiments on the average clustering accuracy and the corresponding standard deviations of different methods show that there is no method that is consistently better than the others under all circumstances. This phenomenon suggests that their performances mostly depend on how the data distribute on the manifolds. Thus, we should try to learn more about the dataset at hand, such as noise level, the distribution of the samples, and so on, before adopting the most appropriate method. However, the experiments have shown that the proposed SMMC always has the highest clustering accuracy and the best average accuracy.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed the SMMC method for hybrid nonlinear manifold clustering, where some data manifolds are separated but some are intersecting. Our method is based on the analysis that SC is able to work well when the affinity values of the points belonging to different clusters are relatively low. Under the assumption that the unlabeled data observations are lying on or close to multiple smooth low-dimensional manifolds, we took advantage of some natural local geometric information of the sampled data, i.e., local tangent space at each sample, to construct an affinity matrix with the expected property. Spectral method was then applied to this affinity matrix to find the clusters. Extensive experiments have shown that SMMC achieves good performance over a broad range of parameter settings and is highly competitive with state-of-the-art methods.

There are several interesting future issues.

1) To improve the robustness: Many of the current methods can provide stable solutions to manifold clustering when the samples lie on or close to multiple "clean" manifolds, however, they will fail when the data are contaminated by noise, especially when faced with outliers. It is interesting to improve the robustness of SMMC.
2) To determine the number of clusters and their dimensionality: Most existing manifold clustering methods, including SMMC, require the user to provide the number of clusters and their intrinsic dimensions. However, such information is often unavailable in real practice. We thus need to develop techniques and criteria to automatically determine the optimal value of these parameters.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and anonymous reviewers for helpful comments and suggestions, and G. Haro for providing the translation Poisson mixture model codes. The research in this paper was accomplished when the first author was visiting the LAMDA Group, Nanjing University, Nanjing, China.

REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2000.
[2] J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," *Appl. Stat.*, vol. 28, no. 1, pp. 100–108, 1979.
[3] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst. 14*, 2001, pp. 849–856.
[4] P. Ren, R. C. Wilson, and E. R. Hancock, "Graph characterization via Ihara coefficients," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 233–245, Feb. 2011.
[5] J. B. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[6] U. Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
[7] K. Zhang and J. T. Kwok, "Clustered Nyström method for large scale manifold learning and dimension reduction," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1576–1587, Oct. 2010.

[8] A. Goldberg, X. Zhu, A. Singh, Z. Xu, and R. Nowak, "Multi-manifold semi-supervised learning," in *Proc. 12th Int. Conf. Artif. Intell. Stat.*, 2009, pp. 169–176.
[9] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[10] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.*, vol. 4, no. 2, pp. 119–155, 2004.
[11] H. S. Seung and D. D. Lee, "Cognition - The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.
[12] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
[13] R. Souvenir and R. Pless, "Manifold clustering," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, vol. 1. Beijing, China, Oct. 2005, pp. 648–653.
[14] Y. Wang, Y. Jiang, Y. Wu, and Z.-H. Zhou, "Multi-manifold clustering," in *Proc. 11th Pacif. Rim Int. Conf. Artif. Intell.*, 2010, pp. 280–291.
[15] G. Haro, G. Randall, and G. Sapiro, "Translated Poisson mixture model for stratification learning," *Int. J. Comput. Vis.*, vol. 80, no. 3, pp. 358–374, 2008.
[16] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, Dec. 2005.
[17] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate," in *Proc. 9th Eur. Conf. Comput. Vis.*, 2006, pp. 94–106.
[18] P. S. Bradley and O. L. Mangasarian, "k-plane clustering," *J. Global Optim.*, vol. 16, no. 1, pp. 23–32, 2000.
[19] R. Cappelli and D. Maltoni, "Multispace KL for pattern representation and classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 977–996, Sep. 2001.
[20] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.
[21] Z. Y. Zhang and H. Y. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2005.
[22] G. Golub and C. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[23] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Comput.*, vol. 11, no. 2, pp. 443–482, Feb. 1999.
[24] G. L. Chen and G. Lerman, "Spectral curvature clustering (SCC)," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 317–330, Mar. 2009.
[25] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (COIL-20)," Dept. Comput. Sci., Columbia Univ., New York, Tech. Rep. CUCS-005-96, Feb. 1996.
[26] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[27] R. Tron and R. Vidal, "A benchmark for the comparison of 3-D motion segmentation algorithms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, Jun. 2007, pp. 41–48.

Yong Wang received the B.S. degree in mathematics from Sichuan University, Chengdu, China, in 2005, and the M.S. degree in applied mathematics from National University of Defense Technology, Changsha, in 2008. He is currently pursuing the Ph.D. degree in the Department of Mathematics and Systems Science, National University of Defense Technology.

His current research interests include machine learning, statistical learning theory, and pattern recognition.

Mr. Wang won the Best Paper Award of the Pacific Rim International Conference on Artificial Intelligence in 2010.

Yuan Jiang received the Ph.D. degree in computer science from Nanjing University, Nanjing, China, in 2004.

She joined the Department of Computer Science and Technology, Nanjing University, where she is now an Associate Professor. She has published over 30 technical papers and holds three patents. Her current research interests include machine learning and information retrieval.

Yi Wu received the B.S. and M.S. degrees in applied mathematics from National University of Defense Technology, Changsha, China, in 1981 and 1988, respectively.

He is a Professor in the Department of Mathematics and Systems Science, National University of Defense Technology. He worked as a Visiting Researcher in the State University of New York at Albany, in 1999. His current research interests include data mining, statistical machine learning, and related applications.

Zhi-Hua Zhou (S'00–M'01–SM'06) received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Nanjing University, Nanjing, China, in 1996, 1998, and 2000, respectively, all with the highest honors.

He joined the Department of Computer Science and Technology, Nanjing University, as an Assistant Professor in 2001, and is currently Cheung Kong Professor and Director of the LAMDA Data Group, Nanjing University. He has published over 80 papers in leading international journals or conference proceedings, and holds 11 patents. His current research interests include artificial intelligence, machine learning, data mining, and pattern recognition.

Dr. Zhou has won various awards/honors including the National Science and Technology Award for Young Scholars of China, the Award of the National Science Fund for Distinguished Young Scholars of China, the Fok Ying Tung Young Professorship First-Grade Award, and the Microsoft Young Professorship Award. He is an Associate Editor-in-Chief of the *Chinese Science Bulletin*, Associate Editor of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING and the *ACM Transactions on Intelligent Systems and Technology*, and is on the editorial boards of various other journals. He is the Founding Steering Committee Co-Chair of the Asian Conference on Machine Learning (ACML) and Steering Committee Member of Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) Conference and Pacific Rim International Conference on Artificial Intelligence (PRICAI). He has served as a Program Committee Chair/Co-Chair for PAKDD in 2007, PRICAI in 2008, and ACML in 2009; Vice Chair or Area Chair or Senior Program Chair of various conferences Senior Program Committee Member of various conferences such as KDD, ICDM, SDM, ECMLPKDD, AAAI, IJCAI, CIKM, ICPR, etc., and has chaired various native conferences in China. He is the Chair of the Machine Learning Society of the Chinese Association of Artificial Intelligence, Vice Chair of the Artificial Intelligence and Pattern Recognition Society of the China Computer Federation, and the Chair of the IEEE Computer Society, Nanjing Chapter. He is a Senior Member of ACM and Senior Member of IEEE.