Discriminant Analysis on Embedded Manifold*

Shuicheng Yan[2], Hongjiang Zhang[1], Yuxiao Hu[1], Benyu Zhang[1],
and Qiansheng Cheng[2]

[1] Microsoft Research Asia, Beijing, P.R. China
[2] School of Mathematical Sciences, Peking University, Beijing 100871, P.R. China

Abstract. Previous manifold learning algorithms mainly focus on uncovering the low dimensional geometry structure from a set of samples that lie on or nearly on a manifold in an unsupervised manner. However, the representations from unsupervised learning are not always optimal in discriminating capability. In this paper, a novel algorithm is introduced to conduct discriminant analysis in term of the embedded manifold structure. We propose a novel clustering algorithm, called Intra-Cluster Balanced K-Means (ICBKM), which ensures that there are balanced samples for the classes in a cluster; and the local discriminative features for all clusters are simultaneously calculated by following the global Fisher criterion. Compared to the traditional linear/kernel discriminant analysis algorithms, ours has the following characteristics: 1) it is approximately a locally linear yet globally nonlinear discriminant analyzer; 2) it can be considered a special Kernel-DA with geometry-adaptive-kernel, in contrast to traditional KDA whose kernel is independent to the samples; and 3) its computation and memory cost are reduced a great deal compared to traditional KDA, especially for the cases with large number of samples. It does not need to store the original samples for computing the low dimensional representation for new data. The evaluation on toy problem shows that it is effective in deriving discriminative representations for the problem with nonlinear classification hyperplane. When applied to the face recognition problem, it is shown that, compared with LDA and traditional KDA on YALE and PIE databases, the proposed algorithm significantly outperforms LDA and Mixture LDA, has better accuracy than Kernel-DA with Gaussian Kernel.

1 Introduction

Previous works on manifold learning [2][6][7][9] focus on uncovering the compact, low dimensional representations of the observed high dimensional unorganized data that lie on or nearly on a manifold in an unsupervised manner. These algorithms can be divided into two classes: 1) algorithms with mapping function only for sample data. The sample points are represented in a low dimensional space by preserving the local or global properties of a manifold, like ISOMAP [16], LLE [12], Laplacian Eigenmap [1]; 2) algorithms with mapping function for the whole data space. Roweis [13] presented an algorithm that automatically aligns a mixture of local dimensionality reducers into a single global representation of the data throughout

* This work was performed at Microsoft Research Asia.

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3021, pp. 121–132, 2004.

space; Brand [3] presented a similar work to merge local representations and construct a global nonlinear mapping function for the whole data space. He [8]proposed the simple locality preserving projections to approximate the Laplacian Eigenmap algorithm. All these algorithms are unsupervised and most of them are only evaluated on toy problems.

In this paper, we propose an algorithm to utilize the class information for discriminant analysis in term of the manifold structure and applications in general classification problems such as face recognition. It is motivated by the following observations: First, previous works on manifold learning focus on exploring the optimal low dimensional representations that best preserve some characteristics of a manifold, while the best representative features are not always the best discriminant features for general classification task. On the other hand, the meaningful information may be lost in the dimensionality reduction, which in turn will degrade the posterior discriminant analysis based on the low dimensional data. Second, Linear Discriminant Analysis can only handle the linear classification problem and Kernel Discriminant Analysis [10] suffers from its heavy computation and memory cost although it can handle nonlinear cases in principle. The proposed algorithm is an efficient, low time and memory cost one for discriminant analysis based on the manifold structure.

For a curved manifold, the globally linearly inseparable manifold may be easily separable locally. The intuition of this work is to place some local Linear Discriminant Analyzers on a curved manifold, then merge these local analyzers into a global discriminant analyzer via global Fisher criterion. In the first step, the traditional methods such as Mixture Factor Analysis (MFA) [1] can not be directly applied, since they can not guarantee that there are balanced samples for the classes in a cluster and it's impossible to conduct Local discriminant analysis with only one class of samples in a cluster. In this work, we formulate this task as a special clustering problem and propose a novel clustering approach, called Intra-Cluster Balanced K-Means (ICBKM), to ensure that there are balanced samples for the classes in a cluster.

Taking the advantage of the clustering results of ICBKM, the sample data are reset as clusters, and local discriminant analysis can be conducted in each cluster. The traditional way to recognize a new data using these local analyzers is to conduct the classification using the nearest local analyzer. In this work, the local analyzers are dependent in both learning and inferring stage, and the optimal discriminative features for each cluster are computed simultaneously. First, PCA is conducted in each cluster; then the posterior probability of each cluster for a given data, i.e. $p(c|x)$ can be obtained. The optimal discriminative features for each cluster are computed by maximizing the global Fisher criterion, i.e. maximizing the ratio of the weighted global inter- and intra- scatters, where the scatters are computed based on the $p(c|x)$ weighted representations for the samples. In the inferring stage, the low dimensional representation for new data is derived as the $p(c|x)$ weighted sum of the projections from different clusters and the classification can be conducted using Nearest Neighbor (NN) algorithm based on the low dimensional representations. This algorithm can be justified in two different perspectives: 1) it automatically merges the local linear discriminant analyzers; and 2) it can be considered as a special kernel discriminant analysis algorithm with geometry-adaptive-kernel, in contrast to traditional KDA whose kernel is independent to the samples.

The rest of the paper is structured as follows. The intra-cluster balanced K-Means clustering method and global discriminant analysis based on the clustering results are introduced in section 2. In section 3, we present our justifications for the proposed algorithm in two different perspectives. The toy problem and the face recognition experiments compared with traditional LDA and KDA on YALE and PIE database are illustrated in section 4. Finally, we give the conclusion remarks in section 5.

2 Discriminant Analysis on Embedded Manifold

Suppose $X = \{x_1, x_2, \cdots x_N\}$ be a set of sample points that lie on or nearly on a low dimensional manifold embedded in a high dimensional observed space. For each sample $x_i \in \mathbb{R}^D$, a class label is given as $l_i \in \{1, 2, \cdots L\}$. Previous works on manifold learning are unsupervised and mainly focus on finding the optimal low dimensional embedding, i.e. the best low dimensional representations that preserve some characteristics of a manifold. However, the class information is not efficiently utilized in these algorithms and the derived representations are not always optimal for general classification task.

Here we show how to utilize the class information to conduct nonlinear discriminant analysis in term of the manifold structure. A continuous manifold can be considered as a combination set of a series of open sets; and for the discrete sample data on it, they can be considered as the combination of a series of clusters. On the other hand, the globally linearly inseparable manifold may be easily separable on these local open sets. It motivates us to conduct local discriminant analysis in these local clusters, and then merge these local analyzers into a global discriminant analyzer. Following this idea, we first segment the sample data into clusters. The traditional clustering algorithms like K-means [11] and Normalized Cut [14] can not be applied to the problem we concern here since there may be only single class in some clusters, which makes the local discriminant analysis impossible. To address this, we have proposed a clustering algorithm called *Intra-Cluster Balanced K-Means* to ensure that the sample numbers for the classes in a cluster are balanced. Secondly, we search for the local optimal features in each cluster by following the global Fisher Criterion in which we maximize the ratio of the cluster weighted inter- and intra-class scatters. In the following subsections, we will introduce the two steps of our algorithm in detail, respectively.

2.1 Intra-cluster Balanced K-Means Clustering

K-means clustering algorithm aims at putting the more similar samples in the same cluster. It is unsupervised, thus it can not guarantee that there is a balanced number of samples for the classes in a cluster. Compared to the general clustering algorithms, the clustering problem we concern here has some special characteristics: 1) the class label for each sample is presented and it can be supervised; 2) its purpose is not only to put the similar samples in the same cluster, but also to ensure that the samples for the classes in each cluster should be balanced since the local discriminant analysis will be conducted in each cluster. Cheung [4] proposed a variation K-Means approach called

---

ICBKM: Given the class label set $S_l = \{1, 2, ..., L\}$, the data set χ, the class label l_i for each sample x_i in χ and the cluster number K.

1. **Initialization:** Compute the standard deviation δ of the data set χ; randomly select $\overline{x}_1, \overline{x}_2, \cdots, \overline{x}_K$ as the initial cluster centers, then assign each x_i to the cluster whose center is the nearest to x_i;
2. **Reset Cluster Centers:** For each cluster C^k, reset the center as the average of all the samples assigned to cluster C^k;
3. **Assignment Optimization:** For each $x_i \in \chi$, assign it to the cluster that makes the objective function smallest and the result satisfies the constraint in (1).
4. **Exchange Optimization:** For each cluster, exchange the cluster labels for the sample in C^k that is the farthest to cluster center and the sample$\notin C^k$ that is the nearest to cluster center. If no improvement, keep the previous labels.
5. **Evaluation:** If current step has no improvement, return the final clustering results $\left\{C^1, C^2, \cdots, C^K\right\}$; else, go step 2;

Fig. 1. Intra-Cluster Balanced K-Means Algorithm

Cluster Balanced K-Means (CBKM), in which the concept cluster balance was proposed. However, it only ensures that the sample number in each cluster is balanced and does not take into account the class label information. To provide a solution to the special clustering problem, we propose a novel clustering approach named *Intra-Cluster Balanced K-Means* (**ICBKM**). ICBKM satisfies the requirement that there are balanced samples for classes in each cluster by adding an extra regularization term to constrain the sample number variation for the classes in each cluster.

Formally, the objective function of ICBKM can be represented as:

$$\arg \min_{K_i \in \{1,2,\cdots K\}} \sum_{i}^{N} \frac{\left|x_i - \overline{x}^{K_i}\right|^2}{\delta^2} + \alpha \sum_{k=1}^{K} | N^k - \overline{N} |^2 + \beta \sum_{k=1}^{K} \sum_{c=1}^{c_k} \left|N_c^k - \overline{N}^k\right|^2 \tag{1}$$

subject to: $c_k \geq 2 \qquad (k = 1, 2, ..., K)$

where $\overline{x}^k$ is the average of the samples in cluster *k*; N^k is the sample number in cluster *k*; $\overline{N}$ is the average sample number for each cluster; N_c^k is the sample number of the *c*-th class in cluster *k*; $\overline{N}^k$ is the average sample number for each class in cluster *k*; c_k is the class number in cluster *k*; α and β are the weighting coefficients for the last two terms.

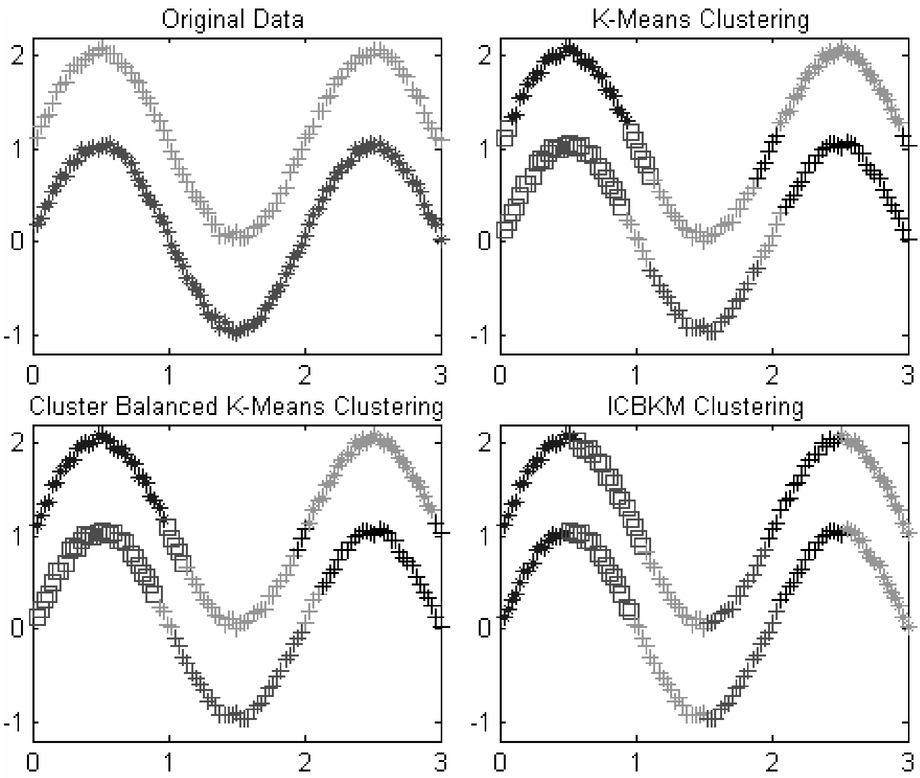

Fig. 2. Toy problem on synthesized data ($\alpha = 0.15, \beta = 0.1$)

In the objective function, minimizing the third term is to ensure that the classes have similar number of samples in a given cluster and the first two terms are the same as in CBKM. The objective function is not trivial and we can not obtain the close form solution directly. Here, we apply an iterative approach as traditional K-Means does. The Pseudo-code is listed in Figure 1. We present a new step for optimization called *Exchange Optimization* in ICBKM, in which the first term can be optimized while the last two terms are kept constant and the optimization conflict between the first term and last two terms that the assignment optimization step may face can be avoided.

The comparison experiments on the synthesized data are conducted and Figure 2 shows the results. It demonstrates that ICBKM presents intra-cluster balanced clustering results; and there is only one class in some clusters in the clustering results of K-Means and CBKM, which makes the local discriminant analysis impossible.

2.2 Global Discriminant Analysis by Merging Local Analyzers

Taking the advantage of the proposed Intra-Cluster Balanced K-Means approach, the sample data are segmented into clusters with balanced samples for different class. The traditional way to utilize these clustering results is to conduct discriminant analysis in each cluster, then determine the class label for new data according to its nearest discriminant analyzer. In this way, the local analyzers is independent and the final classification use only part of the available information. We propose to utilize the global Fisher criterion to combine the local discriminant analyzers into a globally nonlinear discriminant analyzer. The global Fisher criterion maximizes the ratio of the weighed inter- and intra- cluster scatters. The entire algorithm has three steps and they are introduced in detail as follows:

PCA Projections: In each cluster, Principal Components Analysis (PCA) is conducted for dimensionality reduction; moreover, like in Fisher-faces, PCA step can prevent the algorithm from suffering from the singular problem when the sample number is less than the feature number. In all our experiments, we retain 98% of the energy in term of reconstruction error. Thus, in each cluster, each data $x_i \in \chi$ can be presented as a low dimensionality feature vector:

$$z_i^k = (W_{pca}^k)^T (x_i - \overline{x}^k) \qquad k = 1, \cdots, K \tag{2}$$

in which W_{pca}^k is the leading eigenvectors. The conditional probability for cluster k given the data x, $p(C^k \mid x)$, can be obtained using a simple formulation [13]:

$$p(C^k \mid x) = p^k(x) / \sum_{j=1}^{K} p^j(x) \tag{3}$$

where $p^k(x) = \exp\{-\alpha^k(x)\}$ and $\alpha^k(x)$ is the *activity signal* of the data for cluster k. In our experiments, $\alpha^k(x)$ is set as the Mahalanobis Distance of the data in the PCA space of cluster k.

Nonlinear Dimensionality Reduction by Following Global Fisher Criterion: as previously mentioned, the linear discriminant analysis can not handle the nonlinear classification problem; and the KDA suffers from the heavy computation and memory cost in the classification stage. Here, we propose a novel discriminant analysis algorithm to conduct nonlinear discriminant analysis while need not to store samples for the feature extraction of the new data. The optimal features for all the clusters are simultaneously computed in a closed form and the local discriminant analyzers are automatically merged into a globally nonlinear discriminant analyzer by following the global Fisher criterion. For each sample $x_i \in \chi$, it can be represented in cluster k as a low dimensional vector z_i^k. The purpose of the algorithm is to find the optimal feature directions w_f^k and the translations w_0^k for each cluster that minimizes the global Fisher criterion. Let $w^k = ((w_f^k)^T, (w_0^k)^T)^T$, the optimal representation for x_i is a weighted sum of the projections from different cluster:

$$\begin{aligned}\Gamma(x_i) &= \sum_{k=1}^{K} P(C^k \mid x_i)((W_{pca}^k)^T (x_i - \overline{x}^k) \cdot w_f^k + w_0^k) \\ &= \sum_{k=1}^{K} P_i^k (z_i^k \cdot w_f^k + w_0^k) = z_i \cdot w\end{aligned} \tag{4}$$

where $w^T = ((w^1)^T, (w^2)^T, \cdots, (w^K)^T)$ and $z_i^T = ((p_i^1 z_i^1)^T, p_i^1, \cdots, (p_i^K z_i^K)^T, p_i^K)$. And the global intra-class and inter-class scatter can be represented as:

$$S_w = \sum_{i=1}^{N} (\Gamma(x_i) - \overline{\Gamma}^{l_i})(\Gamma(x_i) - \overline{\Gamma}^{l_i})^T = w^T \sum_{i=1}^{N} (z_i - \overline{z}^{l_i})(z_i - \overline{z}^{l_i})^T w = w^T M_w w \tag{5}$$

$$S_b = \sum_{l=1}^{L} N_l (\overline{\Gamma}^{l_i} - \overline{\Gamma})(\overline{\Gamma}^{l_i} - \overline{\Gamma})^T = w^T \sum_{l=1}^{L} N_l (\overline{z}^{l_i} - \overline{z})(\overline{z}^{l_i} - \overline{z})^T w = w^T M_b w \tag{6}$$

where $\overline{\Gamma}^l$ is the mean of $\Gamma(x)$ belonging to class l and $\overline{\Gamma} = \frac{1}{N}\sum_{i=1}^{N} \Gamma(x_i)$. The global Fisher criterion is to maximize the cluster weighted inter-class scatter while minimize the cluster weighted intra-class scatter, *i.e.*

$$w^* = \arg\max_w \frac{|w^T M_b w|}{|w^T M_w w|} \tag{7}$$

It has close form solution and can be directly computed out using generalized eigen-decomposition algorithm [5].

Nonlinear Dimensionality Reduction for Classification: For a new data, the posterior probabilities for each cluster can be computed according to Eqn (3) and its low dimensional representation is obtained via the following nonlinear mapping functions in term of the derived local features in each cluster:

$$M(x) = \sum_{k=1}^{K} P(C^k \mid x)(W_{pca}^k)^T (x - \overline{x}^k) \cdot w_f^{*k} + w_0^{*k}) \tag{8}$$

It is an explicit nonlinear mapping function from the data space to the low dimensional space. The consequent classification can be conducted based on these low dimensional representations using the traditional approaches like Nearest Neighbor (NN) or Nearest Feature Line (NFL). In all our experiments, we used the NN for final classification.

3 Justifications

Our proposed algorithm for discriminant analysis on embedded manifold (Daemon) consists of two steps: 1) separate the samples into class balanced clusters; and 2) merge the local discriminant analyzers into a global nonlinear discriminant analyzer by following the global Fisher criterion. It supervises the local analyzers and automatically decides the responsibility for each analyzer, which is somewhat like the background procedure named Daemon in UNIX system, thus this algorithm is referred as Daemon in the following. The intuition of Daemon is to merge the local discriminant analyzers into a unified framework; while it can be understood from a different perspective: it is a special kind of Kernel Discriminant Analysis algorithm, in which the kernel is data adaptive and geometry dependent; unlike other kernel machines that are independent to the samples they will analysis. In the following, we will discuss these two points in detail.

Automatically Merging Local Discriminant Analyzers: In the first step, Daemon uses ICBKM for clustering and ICBKM ensures that the derived cluster has balanced samples for the classes, thus local discriminant analysis can be conducted in each cluster. Daemon merges these local discriminant analyzers by following global Fisher criterion. The local optimal directions in each cluster are dependent and computed out simultaneously, which is different from the traditional way to utilize clustering results in which local analyzers are independent. In the classification stage, these local analyzers are also dependent and the final representation is a weighted sum of the outputs from these local analyzers. The Eqn (8) can be also presented as:

$$M(x) = \sum_{k=1}^{K} P(C^k \mid x) FE_k(x) \tag{9}$$

in which $FE_k(x)$ is the feature extractor in cluster *k*. As shown in Fig 3, the local analyzer has different optimal feature direction and it is locally discriminative; moreover, they can be merged and result in a globally nonlinear discriminant analyzer.

Special Kernel Discriminant Analysis: Daemon follows the global Fisher criterion in the learning stage and intrinsically is a discriminant analysis algorithm; on the other hand, it is a special kind of Kernel Discriminant Analysis algorithm with geometry-adaptive-kernel. The traditional kernel machine is manually defined and independent to the sample data. As shown in Eqn (4), Daemon can be considered a process in which the training data is mapped into another data space $\{z\}$, and then LDA are conducted in the new feature space. Therefore, Daemon can be considered a special Kernel Discriminant Analysis algorithm and the kernel is:

$$k(x, y) = \phi(x) \cdot \phi(y) \tag{10}$$

where $\phi(x) = z(x) = ((p^1 z^1)^T, p^1, \cdots, (p^K z^K)^T, p^K)^T$ in which $z(x)$ is defined as in Eqn (4). The kernel has the following characteristics: 1) it has explicit mapping function from the input space to another feature space as the polynomial kernel does;

and 2) the kernel is dependent on the training samples and adaptive to the geometry structure. It can be solved just like a traditional Kernel Discriminant Analysis algorithm. Let the sample matrix $X_k = (\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_N))$ and the optimal feature direction $\varphi = X_k v$, the problem is changed as follows:

$$w^* = \arg\max_w \frac{|v^T KMKv|}{|v^T KNKV|} \tag{11}$$

where $K_{ij} = k(x_i, x_j)$, $M = \sum_{l=1}^{L} P_l - P, N = I - \sum_{l=1}^{L} P_l$; and

$P_l = \frac{1}{N_l} e_l e_l^T$, $P = \frac{1}{N} ee^T$, e is a vector with ones, $e_j(i) = \delta_{l_i, j}$.

It can be solved using generalized eigen-decomposition algorithm and the projection of a new data onto the discriminant direction is:

$$M_k(x) = (\varphi, \phi(x)) = \sum_{i=1}^{N} v_i k(x_i, x) \tag{12}$$

It's obvious that $X_k v == w$ since Eqn (7) is equal to Eqn (11) when w is replaced by $X_k v$. Consequently, $M(x)$ in Eqn (8) and $M_k(x)$ in Eqn (12) are also equal. In other words, daemon is a Kernel Discriminant Analysis algorithm with explicit nonlinear mapping function from the input space to another feature space; moreover, as it is designed to be adaptive to the special data geometry structure, it should have strong ability to cope with the nonlinearly distributed data.

4 Experiments

In this section, we present two types of experiments to evaluate the Daemon algorithm. The experiment on toy problem of the synthesized data demonstrates the effective of ICBKM to derive discriminative feature in nonlinear classification problem; and the face recognition results on YALE and CMU PIE database shows that Daemon significantly outperforms LDA and has slightly better accuracy than traditional KDA.

4.1 Toy Problem

As shown in the upper-left image of Figure 2, the original data is composed of two classes of samples and they can not be separated linearly. They are synthesized according to the following function:

$$\begin{cases} x_i^k = 0.03 * i + \delta \\ y_i^k = \sin(\pi x_i) + k + \delta \end{cases} \qquad \text{k=0, 1, } \delta \sim N(0,0.1) \tag{13}$$

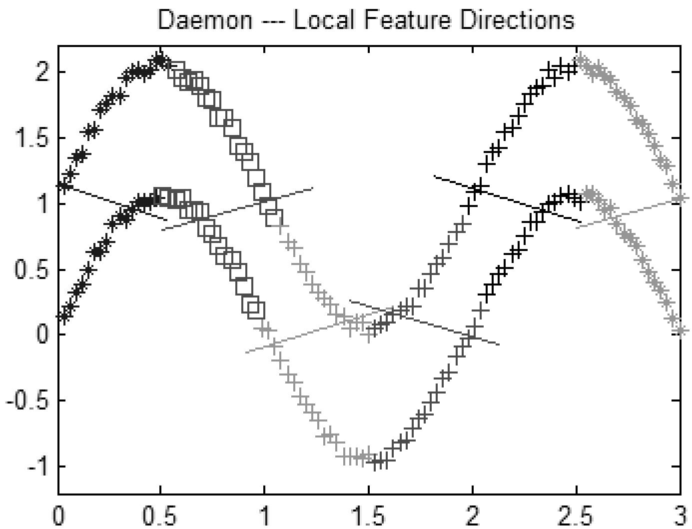

Fig. 3. The derived local feature directions using Daemon

We have systematically compared the clustering results of three K-Means-like algorithms. The original K-Means algorithm produced clustering results that aim at least sums of intra-cluster variances. As shown in the up-right image of Figure 2, the sample numbers for the classes in a cluster is not balanced and some clusters have only one class of samples. It makes the consequent local discriminant analysis impossible in these clusters. The cluster-balanced K-means algorithm produced similar result as that of the original K-means algorithm, yet, the sample number for each cluster is balanced.

As shown in the down-right image of Figure 2, the clustering result from our proposed ICBKM algorithm has the following properties: 1) the sample numbers for the classes in a cluster are balanced, which fascinates the local discriminant analysis in each cluster; and 2) the two classes of samples in each cluster can be easily separated. It is obvious that our proposed ICBKM algorithm produced more useful clustering result than the other two methods and Intra-cluster balanced clustering result presents proper structure representation for the following analysis. The computed local feature direction in each cluster is illustrated in Figure 3. It shows that the local feature direction is approximately optimal for the samples in a cluster.

4.2 Face Recognition

In this subsection, the YALE [17] and PIE [15] databases are used for face recognition experiment. In both experiments, the face image is normalized by fixing the eyes in the same position and each pose uses a different position. Yale face database is constructed at the Yale Center for Computational Vision and Control. It contains 165 grayscale images of 15 individuals. For each individual, six faces are used for training, and the other five are used for testing. Table 1 illustrated the face

Table 1. Comparison between LDAs, Kernel-DA and Daemon on Yale database

Algorithm	Fisher-faces	Kernel-DA	Mixture-LDA	Daemon(K=2)
Accuracy	80%	84%	82.7%	88%

Table 2. Comparison between Fisher-faces, Kernel-DA and Daemon on PIE database

Algorithm	Fisher-faces	Kernel-DA	Daemon(K=5)
Accuracy(67 Dim)	63.63%	67.79%	71.12%

recognition results of the Daemon and LDA, KDA and Mixture LDA that trains different LDA model for each cluster from ICBKM. It shows that Daemon significant outperforms LDA and Mixture LDA, has better results than traditional KDA with Gaussian Kernel.

We have also conducted the multi-view face recognition on the PIE database. We used the face images of pose 02, 37, 05, 27, 29 and 11 with out-plane view variation from -45^{o} to 45^{o} in our experiments. We averaged the results over 10 random splits. The experimental results illustrated in Table 2 again show that Daemon outperforms the other two algorithms. It is demonstrated again that Daemon has strong capability to handle nonlinear classification problems and can improve the accuracy in the general classification problems compared with LDA.

5 Discussions and Future Directions

We have presented a novel algorithm called Daemon for general nonlinear classification problem. Daemon is a nonlinear discriminant analysis algorithm in term of the embedded manifold structure. In this work, the discrete sample data on a manifold is clustered by our proposed Intra-Cluster Balanced K-Means algorithm such that the sample numbers for the classes in a cluster are balanced; and then the local optimal discriminant features are simultaneously derived by following the global Fisher Criterion. It is solved via general Eigen-decomposition algorithm. Daemon can be justified as an automatic merger of the local discriminant analyzers by following the global Fisher criterion; and it can be also justified as a special kernel discriminant analysis algorithm with geometry-adaptive-kernel.

To the best of our knowledge, it is the first work to conduct discriminant analysis while explicitly considering the embedded geometry structure. In this work, we have only utilized the basic property of manifold that a manifold can be covered by a series of open sets; how to combine the other topology properties of a manifold with discriminant analysis for general classification problem is the future direction of our work, and we are considering it in theory and applications.

References

[1] M. Belkin and P. Niyogi, “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering”, *Advances in Neural Information Processing System* 15, Vancouver, British Columbia, Canada, 2001.

[2] C. Bregler, S.M. Omohundro, “Nonlinear manifold learning for visual speech recognition,” *Fifth International Conference on Computer Vision*, June 20-23, 1995.

[3] M. Brand, “Charting a manifold”, *Advances in Neural Information Processing Systems 15, 2002.*

[4] D.W. Cheung, S.D. Lee and Y. Xiao, “Effect of Data Skewness and Workload Balance in Parallel Data Mining”. IEEE Transaction on Knowledge and Data Engineering, V.14 N.3, pp. 498–513, May 2002.

[5] F.R. K. Chung, “Spectral Graph Theory”, *Regional Conferences Series in Mathematics*, number 92, 1997.

[6] D. Freedman. “Efficient simplicial reconstructions of manifolds from their samples”, in *IEEE Trans. On PAM,* Vol: 24 Issue: 10 , Oct 2002, Page(s): 1349–1357.

[7] J. Gomes, and A. Mojsilovic, "A variational approach to recovering a manifold from sample points", *Proc. European Conf. Computer Vision*, Copenhagen, May 2002.

[8] Xiaofei He, Partha Niyogi. "Locality Preserving Projections (LPP)". TR-2002-09, 2002.

[9] G. Hinton and S. T. Roweis. Stochastic Neighbor Embedding. *Advances in Neural Information Processing Systems 15, 2002.*

[10] Qingshan Liu, Rui Huang, Hanqing Lu, Songde Ma. "Face Recognition Using Kernel Based Fisher Discriminant Analysis", in Proceeding of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition Page 197: May 20-21, 2002.

[11] J. MacQueen. "On convergence of k-means and partitions with minimum average variance", Ann. Math. Statist., 36:, 1965.

[12] S. T. Roweis, and L. K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding”, *Science*, vol 290, 22 December 2000.

[13] S. Roweis, L. Saul and G. Hinton, “Global Coordination of Local Linear Models”, *Advances in Neural Information Processing System* 14, 2001.

[14] J. Shi and J. Malik, “Normalized Cuts and Image Segmentation”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22 (2000), 888–905.

[15] T. Sim, S. Baker, and M. Bsat, “The CMU Pose, Illumination, and Expression (PIE) Database”, in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition,* May, 2002.

[16] J. B. Tenenbaum, Vin de Silva, and J. C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction”, *Science*, vol 290, 22 December 2000.

[17] Yale Univ. Face Database, http://cvc.yale.edu/projects/yalefaces/yalefaces.html, 2002.