Locally Linear Models on Face Appearance Manifolds with Application to Dual-Subspace Based Classification[*]

Wei Fan & Dit-Yan Yeung
Department of Computer Science, Hong Kong University of Science and Technology
{fwkevin,dyyeung}@cs.ust.hk

Abstract

Recently, there has been a flurry of research on face recognition based on multiple images or shots from either a video sequence or an image set. This paper is also such an attempt in multiple-shot face recognition. Specifically, we propose a novel nonparametric method that first extracts discriminating local models via clustering. We apply a hierarchical distance-based clustering procedure according to some distance measure on the appearance manifold to cluster similar face images together. Based on the local models extracted, we then construct the intrapersonal and extrapersonal subspaces. Given a new test image, the angle between the projections of the image onto the two subspaces is used as a distance measure for classification. Since a test example contains multiple face images in multiple-shot face recognition, the final classification combines the classification decisions of all individual test images via a majority voting scheme. We compare our method empirically with some previous methods based on a database of video sequences of human faces, showing that out method significantly outperforms other methods.

1 Introduction

Over the past decade or so, the computer vision community has witnessed an increasing trend in performing automatic face recognition [16] based on either a video sequence or an image set. While traditional face recognition methods based on single-shot still images can achieve a certain level of success under restricted conditions, their performance is generally unsatisfactory under more realistic conditions with significantly larger illumination and pose variations, as commonly encountered in applications such as visual surveillance and video retrieval. In this paper, we consider multiple-shot face recognition by extracting locally linear models from nonlinear face appearance manifolds.

Some recent psychological and neural studies [10] show that information useful for identifying a human face can be found both in the invariant structure of features and in idiosyncratic movements and gestures. However, most existing face recognition methods which take into account both cue types simply combine them in a somewhat *ad hoc* manner. Moreover, continuous extraction of face regions from every video frame is generally assumed, posing a formidable challenge even to many state-of-the-art face detection systems. This provides a possible explanation to the finding that the video sequences available in the Face Recognition Vendor Test (FRVT) 2002 could not lead to performance improvement in terms of the recognition rate. In this paper, instead of requiring that face extraction be performed consecutively on all input video frames, we assume independence between all the test images which are assumed to be drawn from some fixed but unknown distribution on the underlying face appearance manifold. This relaxed assumption allows our method to be applicable even to sparse or unordered images, rather than images from continuous video sequences.

Human faces are complex visual patterns embedded in high-dimensional image spaces. One possible approach to the characterization of nonlinear face manifolds is to extract discriminating information locally, and then make use of these local models to characterize the structural variability of the face appearance. A major contribution of this paper is the introduction of a new method for automatically extracting local representative models from the face manifold of each individual, where the representative models on the manifold are locally linear patches that characterize some discriminating information about the individual. Motivated by the Isomap algorithm [13] for nonlinear dimensionality reduction, we use a graph-based method to approximate the geodesic distances between face images in the image space. A hierarchical agglomerative clustering (HAC) algorithm [3] based on geodesic distance is then applied to the face images for each individual separately to form local

[*] This research has been supported by Competitive Earmarked Research Grant HKUST621305 from the Research Grants Council of the Hong Kong Special Administrative Region, China.

clusters of similar face images. These clusters thus formed are used to build local models which will play an important role in the subsequent classification problem. For each local model, we construct two subspaces to characterize intrapersonal and extrapersonal variations. Given a new test image, the angle between the projections of the image onto the two subspaces is used as a distance measure for classification. Since a test example contains multiple face images in multiple-shot face recognition, the final classification combines the classification decisions of all individual test images via a majority voting scheme. We compare our method empirically with some previous methods based on a database of video sequences of human faces. Experimental results show that our method significantly outperforms other competing methods.

2 Previous Work

Methods based on video sequences typically make use of both spatial and temporal information simultaneously. For example, Zhou *et al.* [17] use a probabilistic framework to characterize the kinematics and identity using a motion vector and an identity variable, respectively. The sequential importance sampling (SIS) algorithm is developed to estimate the joint posterior distribution, and marginalization over the motion vector yields a robust estimate of the posterior distribution of the identity variable. Recently, hidden Markov models (HMM) [8] and probabilistic appearance manifolds [7] are both used to learn the transition probabilities among several viewing states embedded in the observation space.

Although facial dynamics, if properly modeled, are tolerant of appearance variations caused by changes in head pose orientation and facial expression, they are not stable and discriminating enough for use in real-world face recognition systems. In this paper, we are interested in a more general scenario in which the multiple images in the image set may come from independent observations that are not necessarily collected over consecutive time steps. As a consequence, the images may have very different viewing conditions. For such isolated observations, it is usually difficult to exploit the temporal relationships between images. Two previous approaches to this problem are the mutual subspace method (MSM) [15] and the probabilistic modeling method [12]. Both methods are based on very simplistic modeling of face pattern variations using a single Gaussian distribution on the face space. Apparently this restriction is unsatisfactory for modeling face variations in real images such as video sequences. The manifold density divergence method [1] relaxes the single Gaussian assumption by using Gaussian mixture models instead. The dissimilarity between estimated probability densities is measured in terms of the Kullback-Leibler divergence. While the methods above are all parametric, Hadid *et al.* [4] use a nonparametric approach which embeds the face manifold using the locally linear embedding (LLE) [11] algorithm and applies k-means clustering in the embedding space. The cluster centers then serve as local models for representing the face manifold locally.

Our work bears resemblance to [8, 7, 4] in that all these video-based face recognition methods make use of local manifold models. However, our method does not explicitly embed the training images to a lower-dimensional space before performing clustering, for two main reasons. First, we want to avoid the computational requirements of the embedding step which typically involves solving an eigendecomposition problem. Second, and perhaps more importantly, embedding to a lower-dimensional space may lead to information loss which can affect the subsequent face recognition results. Instead, we make use of the estimated geodesic distances between face images directly in the distance-based clustering procedure. Our classification method has been inspired by the dual subspace method [9], which differentiates between two classes of face variation: intrapersonal and extrapersonal. The local discriminating information is well captured by the two corresponding subspaces as their dominating orientations are quite different. In the next three sections, we will formulate the problem more exactly and then present details of different components of our method.

3 Problem Setting

Given a face database with $C \geq 2$ subjects, where each subject $c\,(c = 1, 2, \ldots, C)$ has a set F_c of n_c images:

$$F_c = \{f_{c,1}, f_{c,2}, \ldots, f_{c,n_c}\}. \tag{1}$$

For each subject c, we construct a set of local models:

$$P_c = \{p_{c,1}, p_{c,2}, \ldots, p_{c,m_c}\}, \tag{2}$$

where typically $m_c \ll n_c$.

Due to within-class variations in illumination, pose, facial expression and other factors, each subject is better represented by a collection of local models rather than a single global model. The local models are obtained by first performing clustering within the data set for each subject to obtain local clusters which summarize representative latent states of the face variations, and then applying linear fitting to each local cluster.

Given a test set $T = \{t_1, t_2, \ldots, t_l\}$ containing l images of a subject whose identity is one of the C subjects in the training database. We compare each test image t_i with all the gallery local models $p_{c,j}\,(c = 1, \ldots, C; j = 1, \ldots, n_c)$ obtained from the training data. The identity k_i of test image t_i is determined as:

$$k_i = \arg\max_{1 \leq c \leq C} \left\{ \max_{1 \leq j \leq n_c} \mathcal{S}(t_i, p_{c,j}) \right\}, \tag{3}$$

where $\mathcal{S}(\cdot,\cdot)$ denotes the probability that a test image lies on a local patch of the face manifold.

To determine the identity of the entire test set T, we apply a majority voting scheme to the set of individual decisions $\{k_1, k_2, \ldots, k_l\}$ to obtain a single combined decision.

4 Local Model Construction

As is typically the case, we represent each face image of size $r \times s$ as a point in an rs-dimensional space. If we obtain a set of face images for one subject from a video sequence, the different face images are expected to be highly correlated and hence they typically lie on or close to a low-dimensional manifold. The degrees of freedom of the manifold correspond to within-class variations of the face images. Figure 1 shows a training set (blue dots) and a test set (red stars) of images for the moving face of a subject from two short video clips. The three leading principal components of the data are shown. Although the two data sets are obtained from two different video clips, we can see that there is significant overlap between the two manifolds. Moreover, the manifolds have relatively few independent degrees of freedom.

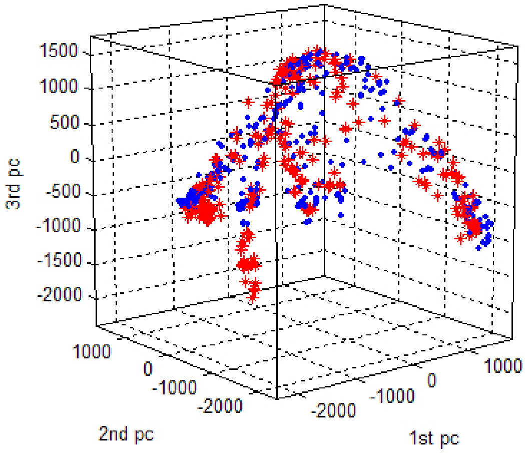

Figure 1. The first three principal components of a training set (blue dots) and a test set (red stars) of images for one moving face, which are automatically detected from two short video clips.

4.1 Graph-Based Approximation of Geodesic Distances

Unlike traditional linear dimensionality reduction methods (e.g., PCA and LDA) which often overestimate the intrinsic dimensionality of the face data set, recently proposed nonlinear dimensionality reduction methods (e.g., Isomap [13] and LLE [11]) can effectively discover a low-dimensional embedding of the manifold.

In this paper, we use the graph-based method of Isomap for approximating the geodesic (or shortest-path) distances between images in a face manifold. We briefly review the method here. More details can be found in [13].

For each subject, we first construct an undirected neighborhood graph with n nodes corresponding to the n images in a face data set for the subject. Each node is connected to its nearest neighbors determined with respect to the Euclidean distance in the image space. There exist different alternatives. In ϵ-Isomap, there is an edge between nodes i and j if the Euclidean distance between them is smaller than some threshold ϵ. Another alternative is the k-Isomap, which connects nodes i and j if node i is among the k nearest neighbors of node j, or vice versa. The edge weight is equal to the Euclidean distance between the two corresponding nodes in the image space.

The shortest path between any two nodes in the graph can be computed using different shortest path algorithms. For example, the Floyd's algorithm has $O(n^3)$ time complexity. There also exist more efficient algorithms, such as Dijkstra's algorithm (with Fibonacci heaps) which has $O(kn^2 \log n)$ time complexity. As a result, we obtain an $n \times n$ matrix of geodesic distances. Instead of using this matrix to embed the n points to a lower-dimensional space as in Isomap, here we use the distance matrix directly in a distance-based clustering procedure.

4.2 Hierarchical Agglomerative Clustering

Hierarchical clustering is a way to investigate grouping in a data set, simultaneously over a variety of scales, by creating a cluster tree called dendrogram. The tree does not represent a single set of clusters, but rather a multi-level hierarchy where clusters at one level are joined together as clusters at the next higher level. This allows one to decide what level or scale of clustering is most appropriate to the specific application at hand.

To perform HAC on a face data set based on the geodesic distances $d_G(\cdot,\cdot)$ as computed above, we apply the following procedure to obtain K clusters:

1. Each data point is initialized as a singleton cluster C_i.

2. Find the nearest pair of clusters according to the following distance measure between clusters C_i and C_j:

$$d_{avg}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{x \in C_i} \sum_{x' \in C_j} d_G(x, x'), \qquad (4)$$

where n_i and n_j are the numbers of images in C_i and C_j, respectively. The two nearest clusters are then merged together to form a new cluster, and hence the total number of clusters is reduced by one.

3. This merging procedure continues until the pre-specified number of clusters (K) is reached.

The data points in each cluster are used to form a local model. Local model construction will be discussed next.

4.3 Representation of Local Models

Given a training data set $F_c = \{f_{c,1}, f_{c,2}, \ldots, f_{c,n_c}\}$ for some subject $c\,(c = 1, 2, \ldots, C)$, the hierarchical clustering procedure described above starts with n_c singleton clusters and terminates when it reaches the number of clusters m_c specified beforehand depending on the length of the video sequence or the number of images n_c in the data set. In our experiments, we set $m_c = 5 \sim 9$ depending on the actual number of detected faces in the video clips (ranging from 250 to 800). For each cluster formed, the cluster mean (which usually does not correspond to a real image in the data set) or the image nearest to the cluster mean may be used as a representative exemplar. Thus subject c uses the set of exemplars $E_c = \{e_{c,1}, e_{c,2}, \ldots, e_{c,m_c}\}$ as a set of local models to represent the original data set F_c. Figure 2 shows five exemplars extracted from a set of 250 training images for one subject based on the above two strategies (see Figure 3 for the original images and Figure 1 (blue dots) for the low-dimensional embedding). They seem to represent different head poses in the data set. Hadid *et al.* [4] extract these exemplars in a similar way (LLE + k-means) and then perform template matching in an appearance-based face recognition system.

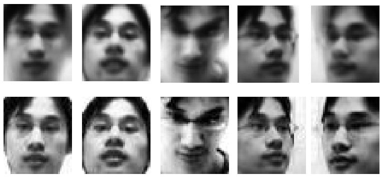

Figure 2. Five exemplars extracted from the set of 250 training images for one subject in Figure 1 (blue dots) based on the cluster means (first row) or the images nearest to their respective cluster means (second row).

While this exemplar-based representation is simple, using a single exemplar for each cluster may not fully characterize the variability of the image data. We argue for representing each local model not just in terms of its cluster mean but also some discriminating information of the entire cluster which is a local patch on the face manifold. For instance, the mixture models for local dimensionality reduction in [6] construct an affine subspace for each cluster (i.e., translation of a linear subspace to the corresponding cluster mean).

Figure 3. Original images from the set of 250 training images in Figure 1 (blue dots).

5 Dual-Subspace Discriminative Classification Method

It is commonly observed that variations between local models for the same subject due to changes in illumination and viewing direction are almost always larger than those due to changes in face identity. When extracting discriminating information from a local model, a reasonable and efficient way is to consider how to distinguish the model from the most confusing local models only within a small neighborhood. Figure 4 shows the cluster centers of one local model p and its six nearest local models that do not belong to the same subject class as p. In such a small spatial scale, the nearest neighbors seem to characterize faces viewed under the same condition, e.g., $+45\,^\circ$ profile faces. The subtle differences between these nearby models are more crucial to classification than those between the faraway ones. Thus, in the following analysis, we only consider discriminating information in the neighborhood of each local model, aiming to concentrate on the "hard" examples only.

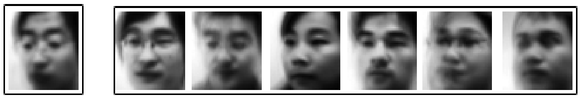

Figure 4. Cluster centers of one local model (left) and its nearest local models that correspond to different subjects (right).

5.1 Distance Measure for Local Models

Note that we have not described how to measure the distance between any two local models p_i and p_j. Let us denote the data matrices for p_i and p_j, respectively, as

$$\begin{aligned} X &= [I_{1,i}, I_{2,i}, \cdots, I_{k_i,i}] && (5) \\ Y &= [I_{1,j}, I_{2,j}, \cdots, I_{k_j,j}] && (6) \end{aligned}$$

where each column of X or Y corresponds to one image in the corresponding cluster (local model). The columns of X and Y define two linear subspaces $\mathcal{X} = \mathrm{span}(X)$ and $\mathcal{Y} = \mathrm{span}(Y)$ in the image space. A distance measure for linear subspaces is the projection $\mathcal{L}_2$-norm:

$$dist_{\mathcal{L}_2}(\mathcal{X}, \mathcal{Y}) = \|P_{\mathcal{X}} - P_{\mathcal{Y}}\|_2, \tag{7}$$

where $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ are the orthogonal projection matrices onto $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $\|.\|_2$ denotes the matrix $\mathcal{L}_2$-norm. The projection $\mathcal{L}_2$-norm is related to the largest canonical angle (or principal angle) between two subspaces. If the maximum canonical angle is small, the subspaces are close to each other. In [5], Hotelling recursively defined the canonical angles $\theta_1, \cdots, \theta_r \in [0, \pi/2]$ between $\mathcal{X}$ and $\mathcal{Y}$ as

$$\cos(\theta_r) = \max_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} x^T y = x_r^T y_r, \tag{8}$$

subject to $\|x\| = \|y\| = 1$, $x^T x_i = 0$, $y^T y_i = 0$, $i = 1, \cdots, r-1$, where $r = \min(\mathrm{rank}(X), \mathrm{rank}(Y))$.

A numerically stable algorithm to compute the canonical angles was proposed by Bjork and Golub [2] based on QR factorization of the data matrices X, Y and singular value decomposition (SVD), as follows.

Let $X = Q_X R_X$ and $Y = Q_Y R_Y$, where Q denotes an orthonormal basis of the respective subspace and R is an upper-diagonal matrix with the Gram-Schmidt coefficients representing the columns of the original matrix in the new orthonormal basis. The singular values $s_1, \cdots, s_r$ of the matrix $Q_X^T Q_Y$ are the cosines of the principal angles:

$$\cos(\theta_i) = s_i, \quad i = 1, \cdots, r. \tag{9}$$

5.2 Dual Subspaces

Using the distance measure defined above, we keep the k nearest local models whose identities are different from that of the local model p being considered. Given a training face database with C subjects, where each subject $c\,(c = 1, 2, \cdots, C)$ has m_c local models, we essentially obtain an adjacency graph with $\sum_{c=1}^{C} m_c$ nodes, one for each local model, and a set of edges connecting nodes p_i and p_j if they are close enough (with respect to the k-neighborhood criterion) and belong to different subjects. In the sequel, we focus on each node and its neighbors in the graph since they together contribute to the local discriminating information that is useful for recognition. This mechanism is based on the intuition that, for recognition, the salient features of a local model are those that best distinguish it from all other most confusing local models of recognition interest.

Motivated by the dual subspace representation proposed in [9], we consider a feature space of Δ vectors representing the differences between two images ($\Delta = I_j - I_k$). One can define two distinct and mutually exclusive classes of facial image variations: intrapersonal variations Ω_I (corresponding, for example, to different facial expressions and illuminations of the same individual) and extrapersonal variations Ω_E (corresponding to variations between different individuals). In our implementation (see Figure 5), all the difference vectors for the local model p_i are calculated based on the cluster center $\mu_i = \frac{1}{k_i} \sum_{k=1}^{k_i} I_{k,i}$, i.e.,

$$\begin{aligned} \Omega_I(i) &= \{\Delta \mid \Delta = I_{k,i} - \mu_i, \forall I_{k,i} \in p_i\} && (10) \\ \Omega_E(i) &= \{\Delta \mid \Delta = I_{k,j} - \mu_i, \forall I_{k,j} \in p_j, j \neq i\}. && (11) \end{aligned}$$

We empirically find that the largest canonical angle between the two subspaces is around 65 ° on the average, indicating that their dominant orientations are quite different.

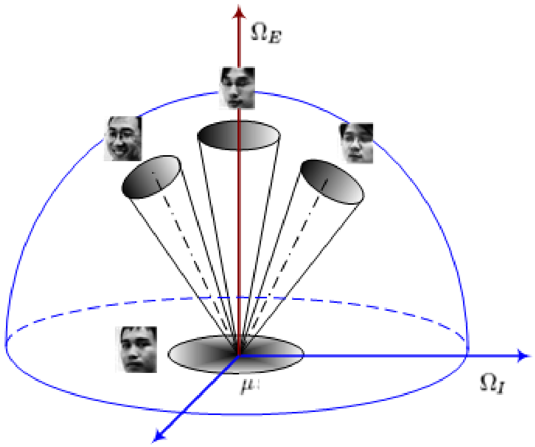

Figure 5. Illustration of the intrapersonal and extrapersonal subspaces in a feature space of the difference vectors Δ.

To determine whether a test face image I_t lies on any local manifold p_i, we first compute the difference image $\Delta_t = I_t - \mu_i$ and then measure the probability $\mathcal{S}(.,.)$ in (3) as follows

$$\mathcal{S}(I_t, p_i) = \frac{1}{\Lambda} \left(\frac{|\cos(\theta(\Delta_t, \Omega_I(i))) - \cos(\theta(\Delta_t, \Omega_E(i)))|}{|\cos(\theta(\Omega_I(i), \Omega_E(i)))|} \right), \tag{12}$$

where $\theta(\Delta_t, \Omega_I(i))$ (or $\theta(\Delta_t, \Omega_E(i))$) is the largest canonical angle between Δ_t and the intrapersonal (or extrapersonal) subspace with respect to p_i, and Λ is a normalization factor which ensures that $\mathcal{S}(.,.)$ is a probability measure. The denominator $|\cos(\theta(\Omega_I(i), \Omega_E(i)))|$ balances the influence of different local models on the probability measure according to their respective discriminability. Specifically, the smaller $\theta(\Omega_I(i), \Omega_E(i))$ is, the less one should rely on the measurement from model p_i. Using the canonical angle as a distance measure for comparison is reasonable since we only deal with locally linear patches on the face manifold, as opposed to global methods such as MSM [15].

6 Experiments and Discussions

We have performed extensive experiments on a 40-subject video data set, which possesses large pose variation and moderate differences in facial expression and illumination. Each subject is found in two video clips, one for training and one for testing, captured using a video camera at a rate of 30 frames per second for about 15-30 seconds. We use the 'AdaBoost + Cascade' face detector of Viola and Jones [14] to detect the face in each video frame. The faces are successfully detected in about 2/3 of the frames. All the detected faces are resized to gray-level images of size 45×40, followed by a histogram equalization step to eliminate the lighting effects. The examples shown in Figure 6 are representative of the variations in the data set.

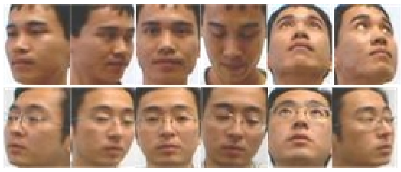

Figure 6. Representative images of two subjects from the data set used in our experiments.

We perform experiments on the following methods:

1. Nearest neighbor template matching with exemplars extracted by:
   (a) Random selection from the training set
   (b) PCA + k-means clustering
   (c) LLE + k-means clustering [4]
   (d) Geodesic distance approximation + HAC
2. Mutual subspace method (MSM) [15]
3. Kullback-Leibler divergence method [12]
4. Our dual-subspace discriminative method

For all exemplar-based methods and our dual-subspace discriminative method, we use a majority voting scheme to combine the decisions of individual frames to give the final classification result. Specifically, let us assume that there are C classes and a test video sequence contains K frames. We use variables δ_{ij} $(i = 1, \ldots, C; j = 1, \ldots, K)$ to represent the decisions of the K frames, such that $\delta_{ij} = 1$ if the jth frame is decided to belong to the ith class, and 0 otherwise. The final recognition result of the test sequence is $h = \arg\max_{i=1}^{C} \sum_j \delta_{ij}$, i.e., the test sequence belongs to the hth class.

6.1 Comparison of Exemplar-based Methods

We first compare the four exemplar-based methods. Since the focus here is the automatic extraction of local models, we simply build appearance-based face recognition systems based on performing nearest neighbor template matching either in the original image space or after applying PCA and LDA which are traditional linear dimensionality reduction methods.

The training video sequence of each subject is used to build local models using the four exemplar-based methods. Depending on the sequence length, 5-9 local models are built for each subject. During the testing stage, we perform 10 random trials for each subject with each trial performed by randomly sampling 30 frames from the corresponding test video sequence. The recognition rates shown in Table 1 are the average results over all random trials for all subjects.

Table 1. Average recognition rates (%) of exemplar-based methods.

	Original	PCA	LDA
Random selection	65.62	63.21	74.62
PCA + k-means	74.02	75.26	88.29
LLE + k-means	88.33	86.76	92.43
Geodesic + HAC	89.74	88.10	94.14

The results show that the methods based on manifold learning (LLE + k-means; geodesic + HAC) can select better exemplars (local models) than the other methods (random selection; PCA + k-means) since they yield higher recognition rates. This finding is not unexpected, as methods based on manifold learning can effectively reveal meaningful hidden structures in the nonlinear face manifolds.

Another interesting finding is that our method slightly outperforms LLE which is based on explicit embedding of the data. For the purpose of clustering (exemplar selection), in fact there is no need to perform the last step (embedding) in LLE or Isomap. Doing so will not only require solving an eigendecomposition problem which is expensive for large data sets, but it can also lead to a certain degree of information loss. The reason we prefer a global embedding method (Isomap) to its local alternatives (e.g., LLE, Laplacian eigenmap) lies in its appealing property of explicitly preserving the global structure of a data set within a single coordinate system. As proved in the original paper, the approximated graph-based distances in Isomap asymptotically converge to the true geodesic distances of the manifold given sufficient data. Moreover, unlike k-means clustering which is sensitive to the initial seeds and may get trapped in local minima, the HAC algorithm in our method is more stable to the input data.

6.2 Generative vs. Discriminative Methods

We also compare our dual subspace method, which is a discriminative method, with two previous generative methods for multiple-shot face recognition, namely, the MSM method [15] and the Kullback-Leibler divergence method [12].

For both methods, a single global model is built for each subject. For our method, we set $k = 6$, i.e., six nearest local models that belong to a different subject class are identified for each local model.

Table 2. Average recognition rates (%) of two previous generative methods and our discriminative method.

Method	MSM	KL	Dual
Recogition rate	87.12	92.91	95.62

From the results in Table 2, we can see that our dual-subspace discriminative method outperforms the two generative methods. While generative methods typically use positive examples only during model training, discriminative methods make use of both positive and negative examples and hence they can explicitly exploit the discriminating information to achieve higher recognition rates. Specifically, discriminating information is represented in both the intrapersonal and extrapersonal subspaces in our method. Moreover, the MSM and KL divergence methods use rather simplistic Gaussian modeling that cannot model well the complex face pattern variations.

7 Concluding Remarks

We have presented a novel approach for multiple-shot face recognition that is based on extracting local models from the appearance manifolds and a discriminative recognition method using two subspaces representing intrapersonal and extrapersonal variations. Empirical results show that our method gives very promising results when compared with previous methods.

Currently we simply apply majority voting to obtain the final classification, assuming that the examples in the test set are i.i.d. However, for sequential data, even though there is no guarantee that the faces can be successfully detected from all images in the video sequence, the data set still contains meaningful temporal relationships between images. In our future work, we will consider modeling some inherent dynamics in the "fragmented" data which hopefully can further enhance the classification performance.

References

[1] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. In *Proceedings of the CVPR*, volume 1, pages 581–588, 2005.

[2] A. Björck and G. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973.

[3] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, 2000.

[4] A. Hadid and M. Pietikainen. From still image to video-based face recognition: an experimental analysis. In *Proc. of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 17–19, 2004.

[5] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–372, 1936.

[6] N. Kambhatla and T. Leen. Dimension reduction by local principal component analysis. *Neural Computation*, 9:1493–1516, 1997.

[7] K. Lee, J. Ho, M. Yang, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. In *Proceedings of the CVPR*, pages 313–320, 2003.

[8] X. Liu and T. Chen. Video-based face recognition using adaptive hidden Markov models. In *Proceedings of the CVPR*, pages 340–345, 2003.

[9] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33:1771–1782, 2000.

[10] A. O'Toole, D. Roark, and H. Abdi. Recognizing moving faces: A psychological and neural synthesis. *Journal of Vision*, 2:604a, July 2002.

[11] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.

[12] G. Shakhnarovich, J. Fisher, and T. Darrell. Face recognition from long-term observations. In *Proceedings of the ECCV*, volume 3, pages 851 – 868, 2002.

[13] J. Tenenbaum, V. Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[14] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the CVPR*, volume 1, pages 511–518, 2001.

[15] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *Proc. of IEEE Internation Conf. on Automatic Face and Gesture Recognition*, pages 318–323, 1998.

[16] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.

[17] S. Zhou and R. Chellappa. Probabilistic human recognition from video. In *Proceedings of the ECCV*, volume 3, pages 681–697, 2002.