Bayesian Face Recognition Based on Gaussian Mixture Models

Xiaogang Wang and Xiaoou Tang

Department of Information Engineering
The Chinese University of Hong Kong
{xgwang1, xtang}@ie.cuhk.edu.hk

Abstract

Bayesian analysis is a popular subspace based face recognition method. It casts the face recognition task into a binary classification problem with each of the two classes, intrapersonal variation and extrapersonal variation, modeled as a Gaussian distribution. However, with the existence of significant transformations, such as large illumination and pose changes, the intrapersonal facial variation cannot be modeled as a single Gaussian distribution, and the global linear subspace often fails to deliver good performance on the complex non-convex data set. In this paper, we extend the Bayesian face recognition into Gaussian mixture models. The complex intrapersonal variation manifold is learnt by a set of local linear intrapersonal subspaces and thus can be effectively reduced. The effectiveness of the novel method is demonstrated by experiments on the data set from AR face database containing 2340 face images.

1. Introduction

Face recognition has drawn much attention in recent years. However, it remains a complicated problem far from being completely solved. There are two major difficulties. The first is due to the significant intrapersonal variation. The face images for the same person may have very different appearance under different conditions. The existence of expression, lighting and pose changes leads to a complex distribution of face set. Second, there are usually not enough reference samples for each face class to capture all kinds of variations. Sometimes, only one reference image is available for each person, while the probe image may be taken under a very different condition. So the design of a face recognition system should focus on how to reduce intrapersonal variation using limited training data.

A family of subspace methods such as PCA and LDA are developed to extract low dimensional features from the raw face data for recognition [1][2][4][5][6][7]. Bayesian analysis [3] is another successful subspace face recognition method. Instead of classifying the probe face image into L classes for L individuals, it casts the face recognition task into solving a binary pattern recognition problem with each of the two classes, intrapersonal variation and extrapersonal variation, modeled as a Gaussian distribution. The success of Bayesian face recognition is based on the fact that it separates the intrapersonal transformation difference, caused by expression, lighting, and pose changes etc., from the intrinsic difference discriminating individual identity, and effectively reduces it under a probabilistic measure [4][5]. However, when the transformation difference is significant, the intrapersonal variation manifold will become highly non-convex and complex. A global linear subspace based on a single Gaussian model often fails to deliver good performance. In this paper, we extend the Bayesian face recognition into Gaussian mixture models. The complex intrapersonal variation manifold is decomposed into several clusters with simple distribution, and learnt by multiple local intrapersonal subspaces.

Several mixture linear subspace methods have been proposed in previous work. However, their clustering procedures are all based on face images or face classes instead of intrapersonal face difference as proposed in this paper. In view-based PCA subspaces [8] and two-stage LDA face recognition [9], the face images are partitioned into several subsets according to different views and a PCA or LDA subspace is trained for each view. The pose of the probe face image is first recognized and then the face class is recognized among a subset of reference images with the same view. This approach requires the face class has at least one reference sample for each cluster and this condition is difficult to meet in many applications. In [10], Gross et al. employ Gaussian mixture models to characterize each face class. It also requires a large number of training samples for each person. In [11][12], the face classes are clustered into several subsets based on class centers. However, each subset still contains all kinds of intrapersonal variations. Even though there are only a small number of face classes in each subset, the face data distribution still may be too complex to be linearly separable.

The main advantage of our mixture Bayesian face recognition method over these conventional mixture subspaces methods is that it focuses on intrapersonal variation, the most significant factor deteriorating face recognition performance and causing complex data distribution. Moreover, it can be accomplished even if there is only one reference sample for each face class. Experiments on the data set from AR face database [13]

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

containing 2340 face images show the superiority of the new method.

2. Bayesian Face Recognition

Face recognition can be essentially considered as determining whether two face vectors are from the same individual (intrapersonal variation Ω_I) or different individuals (extrapersonal variation Ω_E). We decompose the difference (Δ) between two face vectors into three components: intrinsic difference ($\widetilde{I}$), which discriminates different individuals; transformation difference ($\widetilde{T}$), caused by all kinds of transformations such as varying expressions, illuminations, and views; and noise ($\widetilde{N}$) [4][5]. $\widetilde{T}$ and $\widetilde{N}$ are the two components deteriorating the recognition performance. Normally, $\widetilde{N}$ is of small energy. The main difficulty for face recognition comes from $\widetilde{T}$, which can change the face appearance substantially. A successful face recognition algorithm should reduce the energy of $\widetilde{T}$ as much as possible without sacrificing much of $\widetilde{I}$.

In the Bayesian algorithm, the similarity between two images can be measured as the intrapersonal likelihood $P(\Delta \mid \Omega_I)$. Principal component analysis (PCA) is applied on the intrapersonal difference set $\{\Delta \mid \Delta \in \Omega_I\}$ to compute the intrapersonal principal subspace F and its complementary subspace $\overline{F}$. Assuming that Ω_I has a Gaussian distribution, $P(\Delta \mid \Omega_I)$ is estimated as the product of two independent marginal Gaussian densities in F and $\overline{F}$,

$$P(\Delta \mid \Omega_I) = \left[\frac{\exp\left(-\frac{1}{2} d_F(\Delta)\right)}{(2\pi)^{M/2} \prod_{i=1}^{M} \lambda_i^{1/2}}\right] \left[\frac{\exp\left(-\varepsilon^2(\Delta)/2\rho\right)}{(2\pi\rho)^{(N-M)/2}}\right]. \quad (1)$$

In Eq.(1), $d_F(\Delta)$ is a Mahalanobis distance in F , referred as "distance-in-feature-space" (DIFS),

$$d_F(\Delta) = \sum_{i=1}^{M} \frac{y_i^2}{\lambda_i}, \quad (2)$$

where y_i is the principal component of Δ projected into F and λ_i is the eigenvalue. $\varepsilon^2(\Delta)$ is defined as "distance-from-feature-space" (DFFS), equivalent to PCA residual error in $\overline{F}$.

Ω_I contains $\widetilde{T}$ and $\widetilde{N}$ only, since it comes from the same individual. PCA on Ω_I produces a set of principal axes dominated by $\widetilde{T}$. When a face difference Δ is projected onto the intrapersonal subspace, its $\widetilde{T}$

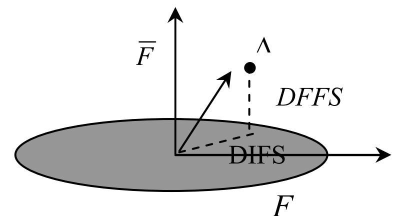

Figure 1. Compute DIFS and DFFS of Δ in the intrapersonal subspace.

component is therefore compacted onto a small number of largest eigenvectors in F. The energy of $\widetilde{I}$ is mainly concentrated in $\overline{F}$. In such a way, $\widetilde{T}$ and $\widetilde{I}$ are decoupled. Since λ_i explicitly describes the energy distribution of $\widetilde{T}$, $\widetilde{T}$ can be effectively reduced by the inverse weighting of eigenvalues in DIFS. DFFS is a also distinctive component for recognition. It throws away most $\widetilde{T}$ on large eigenvectors, while keeps most of $\widetilde{I}$.

3. Bayesian Face Recognition Based on Gaussian Mixture Models

However, when the intrapersonal difference in the face data set is too large, the intrapersonal variation manifold will be too complex to be modeled as a single Gaussian distribution. The derived intrapersonal subspace cannot effectively separate $\widetilde{T}$ from $\widetilde{I}$. We project the intrapersonal differences of samples from AR database onto the first two eigenvectors of intrapersonal subspace and plot them in Figure 2. Apparently it is not a Gaussian distribution. A better choice is to decompose the complex manifold into K simpler clusters and use multiple intrapersonal subspaces to model local regions. An algorithm deriving the local intrapersonal subspaces is proposed as following.

(1) Randomly choose the initial cluster assignment $\ell(\Delta) \in \{1,2,\ldots,K\}$ for each intrapersonal difference sample Δ in the training set.
(2) Perform PCA separately on each cluster $C_k = \{\Delta \mid \ell(\Delta) = k\}$ and compute the local intrapersonal subspace with eigenvector matrix U^k and cluster center m_k .
(3) Project the training intrapersonal difference example Δ into each local intrapersonal subspace spanned by r largest eigenvectors $W^k = \left[\vec{u}_1^k, \ldots, \vec{u}_r^k\right]$ and compute the reconstruction error as

$$\varepsilon_k^2 = \left\| (\Delta - m_k) - W^k \left(W^k\right)^T (\Delta - m_k) \right\|^2 . \quad (3)$$

Reassign Δ to the cluster with the minimum reconstruction error,

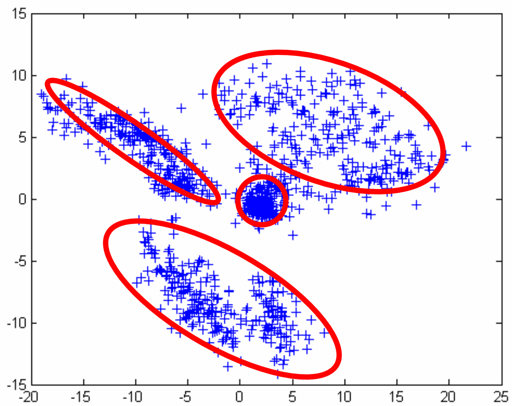

Figure 2. Project the samples of intrapersonal difference from the AR database to the first two eigenvectors of intrapersonal subspace. The complex distribution can be better modeled as several local Gaussian models.

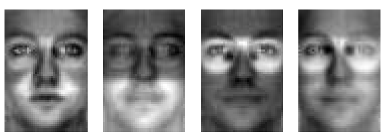

Figure 3. Cluster centers for different local intrapersonal subspaces.

$$\ell(\Delta) = \arg\min_{k}(\varepsilon_k^2)\,. \qquad (4)$$

(4) Stop if no training examples has changed cluster, otherwise return to step (2).

In Eq. (4), we use the "distance-from-feature-space" (DFFS) instead of Euclid distance to cluster center to assign the cluster to training samples. A small DFFS means the local subspace could well capture the major $\widetilde{T}$ component, which is critically important to Bayesian face recognition. The cluster centers of some local intrapersonal subspaces are shown in Figure 3.

When a probe face image is input, we compare its difference with the reference image and estimate the intrapersonal likelihood $P_k(\Delta\,|\,\Omega_I)$ for each local intrapersonal subspace. The local intrapersonal subspace with the maximum $P_k(\Delta\,|\,\Omega_I)$ is the best to model Δ, so $P(\Delta\,|\,\Omega_I)$ is estimated as

$$P(\Delta\,|\,\Omega_I) = \max\{P_k(\Delta)\}_{k=1}^{K}\,. \qquad (5)$$

$P_k(\Delta\,|\,\Omega_I)$ can be estimated using Eq. (1). However, it is sometimes unstable when there are only a small number of training samples falling into some local cluster and the eigenvalue spectrum cannot be well estimated. Here, we estimate $P_k(\Delta\,|\,\Omega_I)$ using a simpler but more stable version,

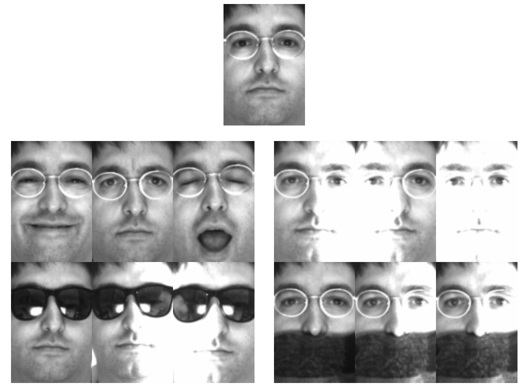

Figure 4. Face image examples taken in the same session for one subject in the AR database.

$$P_k(\Delta\,|\,\Omega_I) = \exp\left(-\varepsilon_k^2(\Delta)\right) \qquad (6)$$

where $\varepsilon_k^2(\Delta)$ is DFFS to the *kth* intrapersonal subspace.

4. Experiments

We conduct experiments on a data set from the AR database. It contains 90 subjects and each subject has 26 face images taken in two sessions. For each session, there are 13 face images under different kinds of transformations. Some examples are shown in Figure 4. The 1170 face images taken in the first sessions are used for training set to compute the intrapersonal subspace. In testing, one neural face image taken in the first session for each subject is used as reference, and the 1170 face images taken in the second session are used as probe. In preprocessing, all the images are normalized for scaling, translation, and rotation, such that the eye centers are in fixed positions. A rectangle mask is used to remove the background and most of the hair region.

In Figure 5, we compare the accumulative recognition accuracies of Bayesian face recognition based on Gaussian mixture models with several linear subspace methods based on uniform model, PCA, LDA, and Bayes. PCA is a baseline for evaluation, since it captures all kinds of major facial variation including both *T* and *I*, and does not take effort to reduce the intrapersonal variation. The recognition accuracy is extremely low, with about 20% on top one match. Bayes performs much better than PCA, since it reduces the transformation difference to some extent. Bayesian face recognition based on four local intrapersonal subspaces has more than 10% improvement on top one match to the best linear subspace method as shown in Table 1. This clearly shows that the Gaussian mixture models can better deal with the complex data distribution. We also try the mixture Bayes based on different number of clusters. It is found that the performance only has a slight variation as shown in Fig. 6. The mixture linear subspaces methods proposed in [8][9]

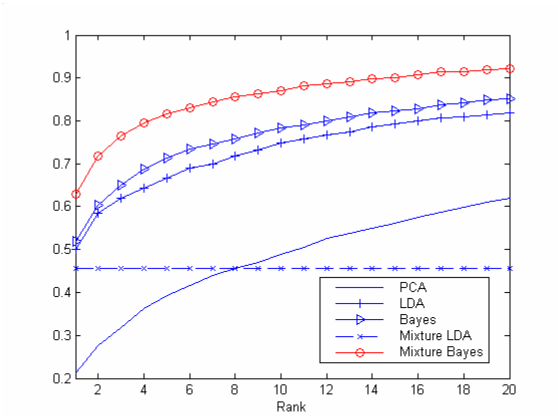

Figure 5. Accumulative recognition accuracies of PCA, LDA, Bayes, and mixture Bayes. Mixture LDA only has the recognition accuracy of top one match.

Table 1. Rank 1 recognition accuracies.

PCA	LDA	Bayes	Mixture LDA	Mixture Bayes
0.2137	0.5009	0.5188	0.4550	0.6385

cannot be used here, since there is only one reference sample for each class. We test the mixture LDA method proposed in [12] by grouping the face classes into four clusters. This method only has top one recognition accuracy (45% as shown in Fig. 5). It fails to improve the performance, because the complex distribution of this data set is caused by large intrapersonal variations, which still exist in each cluster of that method.

5. Conclusion

In this paper, we propose a novel Bayesian face recognition approach based on Gaussian mixture models. Under this framework, the complex intrapersonal manifold is characterized by several local intrapersonal subspaces. It shows a significant improvement to conventional linear subspace methods based on uniform model.

Acknowledgement

The work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project no. CUHK 4190/01E, CUHK4224/03E, and CUHK409/03).

References

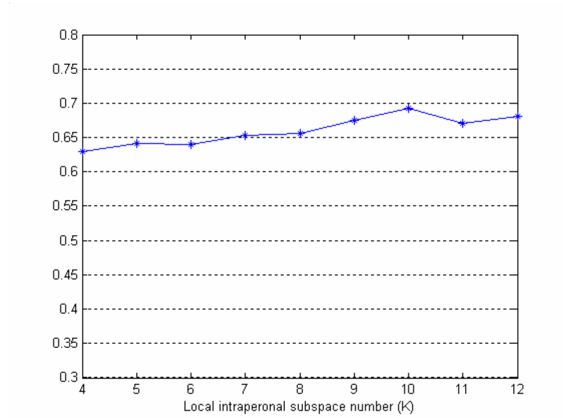

Figure 6. Recognition accuracies of mixture Bayes on top one match based on different number of local intrapersonal subspaces.

[1] M. Turk and A. Pentland, "Face recognition using eigenfaces," *in Proceedings of CVPR*, pp. 586-591, Hawaii, June, 1991.

[2] V. Belhumeur, J. Hespanda, and D. Kiregeman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. on PAMI,* Vol. 19, No. 7, pp. 711-720, July 1997.

[3] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian Face Recognition," *Pattern Recognition*, Vol. 33, pp. 1771-1782, 2000.

[4] X. Wang and X. Tang, "Unified Subspace Analysis for Face Recognition," *in Proceedings of ICCV*, pp. 679-686, Nice, France, Oct. 2003.

[5] X. Wang and X. Tang, "A Unified Framework for Subspace Face Recognition," *IEEE Trans. on PAMI*, Vol. 26, No. 9, Sept. 2004.

[6] X. Wang and X. Tang, "Random Sampling LDA for Face Recognition," *in Proceedings of CVPR*, 2004.

[7] X. Wang and X. Tang, "Dual-Space Linear Discriminant Analysis for Face Recognition," *in Proceedings of CVPR*, 2003.

[8] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," *in Proceedings of CVPR*, 1994.

[9] E. Demir, L. Akarun, and E. Alpaydin, "Two-Stage Approach for Pose Invariant Face Recognition," in *Proceedings of ICASSP*, Vol. 4, pp. 2342-2344, 2000.

[10] R. Gross, J. Yang and A. Waibel, "Growing Gaussian Mixture Models for Pose Invariant Face Recognition," in *Proceedings of ICPR,* pp. 1088-1091, 2000.

[11] H. Kim, D. Kim, and S. Y. Bang, "Face Recognition Using LDA Mixture Model," in *Proceedings of ICPR*, pp. 486-489, 2002.

[12] J. Lu and K. N. Plataniotis, "Boosting face recognition on a large-scale database," in *Proceedings of ICIP*, Vol. 2, pp. 109-112, 2002.

[13] A.M. Martinez and R. Benavente, "The AR face database," *CVC Tech. Report #24*, 1998.