AN IMPROVED BAYESIAN FACE RECOGNITION ALGORITHM IN PCA SUBSPACE

Xiaogang Wang and Xiaoou Tang

Department of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
(Email: xgwang1, xtang@ie.cuhk.edu.hk)

ABSTRACT

Through modeling the difference between two face images by three components, intrinsic difference (*I*), transformation difference (*T*), and random noise (*N*), we show that the Bayesian algorithm can successfully separate the main disturbing component *T* from the discriminating component *I*, however at a cost of magnified noise *N*. To control the noise, we apply PCA on the original image, then carry out the Bayesian analysis in the reduced PCA space. The new method is shown to be more effective than the standard Bayesian algorithm in experiments using 2000+ face images from the Feret database.

1. INTRODUCTION

Face recognition has been studied extensively in recent years. Among the existing face recognition techniques, subspace methods are widely used to extract low dimensional features [4]. Principle Component Analysis (PCA) method [2] is one of the most popular subspace methods. It uses Karhunen-Loeve Transform (KLT) to produce a most expressive subspace for face representation and recognition. By PCA, the dimensionality of image space can be dramatically reduced, and noise encoded on the small eigenvectors can be removed.

However, as an optimal method for face representation, the PCA method is not optimal in terms of extracting the most discriminating features. Recently, the Bayesian algorithm [1] has been shown to be more effective for face recognition. Different from other techniques, which classify face images into *M* classes for *M* individuals, the Bayesian algorithm casts the face recognition task as a binary pattern classification problem with each of the two classes, intrapersonal variation and extrapersonal variation, modeled as a Gaussian distribution. In the probabilistic subspace, transformation variations such as expression and lighting variations can be effectively reduced. However, as shown in this paper, since the discriminating features and noise are coupled on the small eigenvectors, noise will be magnified when normalized by the small eigenvalues in the probabilistic similarity measure. In this paper, we propose an improved Bayesian algorithm in the reduced PCA space. PCA is first used to separate the noise from the transformation variations and discriminating features. Then, the Bayesian analysis is applied to the PCA subspace in order to remove the effect of transformation variations from the final feature vectors.

2. A SHORT REVIEW OF THE BAYESAIN ALGORITHM

The Bayesian algorithm classifies the face intensity difference Δ as intrapersonal variation (Ω_I) for the same individual and extrapersonal variation (Ω_E) for different individuals [1]. The MAP similarity between two images is defined as the intrapersonal a posterior probability

$$S(I_1, I_2) = P(\Omega_I \mid \Delta) = \frac{P(\Delta \mid \Omega_I)P(\Omega_I)}{P(\Delta \mid \Omega_I)P(\Omega_I) + P(\Delta \mid \Omega_E)P(\Omega_E)}. \quad (1)$$

Because of the high dimensionality, $P(\Delta \mid \Omega_I)$ and $P(\Delta \mid \Omega_E)$ cannot be estimated directly from the training set. So subspace estimate is used. To estimate $P(\Delta \mid \Omega_I)$, PCA on the set $\{\Delta \mid \Delta \in \Omega_I\}$ decomposes the image difference space into principle subspace F, called intrapersonal eigenspace with K eigenvectors and its orthogonal complementary space $\bar{F}$. The likehood can be estimated as the product of two independent marginal Gaussian densities in F and $\bar{F}$,

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

In Eq. (2), $d_F(\Delta)$ is a Mahalanobis distance in F, referred as "distance-in-feature-space" (DIFS),

$$d_F(\Delta) = \sum_{i=1}^{K} \frac{y_i^2}{\lambda_i}, \quad (3)$$

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

$$\rho = \frac{1}{N-K} \sum_{i=K+1}^{N} \lambda_i . \quad (4)$$

$P(\Delta | \Omega_E)$ can be estimated in a similar way. The principle subspace computed from the set $\{\Delta | \Delta \in \Omega_E\}$ is called extrapersonal eigenspace.

An alternative maximum likehood (ML) measure is defined as

$$S'(\Delta) = P(\Delta | \Omega_I) . \quad (5)$$

It is equivalent to evaluating a distance measure in the intrapersonal subspace,

$$D_I = d_F(\Delta) + \varepsilon^2(\Delta) / \rho . \quad (6)$$

ML measure has been shown to be simpler but almost as effective as the MAP measure in Eq. (1).

3. IMPROVED BAYESIAN ALGORITHM

3.1 Face difference model

We model the difference Δ between two face images by three components: intrinsic difference (*I*) that discriminates different individuals; transformation difference (*T*) caused by such transformations as lighting or expression changes; and random noise (*N*). *T* and *N* are two components deteriorating the recognition performance. Normally *N* is of small energy. Under a large transformation, *T* could potentially be greater than *I*. A successful subspace method should be able to reduce the effect of *T* and *N* as much as possible without sacrificing much of *I*. By analyzing the distribution of *I*, *T* and *N* in the PCA and Bayesian analysis, we can show how *T* and *N* can be effectively removed.

3.2. Intrapersonal subspace

Intrapersonal subspace plays a critical role in the Bayesian algorithm. ML measure using intrapersonal subspace alone is almost as effective as the MAP measure. Ω_I is composed of *T* and *N*, since it comes from the same individual, and *T* is the principle component,

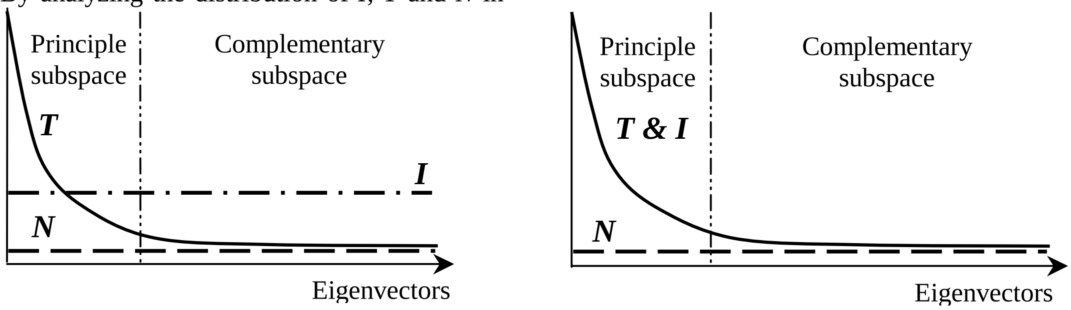
$$\Omega_I = T + N . \quad (7)$$

As shown in Figure 1, PCA on the intrapersonal variation set arranges the axes according to the energy distribution of *T*. When we project a face difference Δ (either intrapersonal or extrapersonal) onto the intrapersonal subspace, most energy of the *T* component will concentrate on the first few largest eigenvectors, while the *I* and *N* components are randomly distributed over all of the eigenvectors. This is because *I* and *N* are independent of *T*, which forms the principle vectors of the intrapersonal subspace. In Eq. (6), the Mahalanobis distance in *F* weights the feature vectors by the inverse of eigenvalues. This effectively reduces the *T* component since the principle components with large eigenvalues are significantly diminished. $\varepsilon^2(\Delta)$ is also a distinctive component for recognition, since it throws away most of the component *T* on the largest eigenvectors, while keeps the majority of *I*.

3.3. Bayesian analysis in reduced PCA space

As shown in the previous section, the Bayesian algorithm successfully separates *T* from *I*. However, *I* and *N* are still coupled on the small eigenvectors. Even though *N* is usually of small energy, when it is normalized by the small eigenvalues as shown in Eq. (3) and (6), the effect of *N* could be significantly enlarged in the probabilistic measure.

To solve this problem, we first apply PCA on the original image vectors. As shown in Figure 1, in the PCA subspace, both *T* and *I*, as structured signals embedded in



(a) Intrapersonal subspace (b) PCA subspace

Figure 1. Energy distribution of the three components *I*, *T* and *N* on eigenvectors in the intrapersonal subspace (a) and the standard PCA subspace (b).

the original face image, will concentrate on the small number of principle eigenvectors. By selecting the principle components, most noise encoded on the large number of trailing eigenvectors is removed from *T* and *I*. In the following Bayesian analysis, intrapersonal and extrapersonal subspaces are derived from the reduced PCA space to separate *T* from *I*. Since the space dimensionality has been dramatically reduced by PCA, likehood can be estimated directly from DIFS,

$$P(\Delta \mid \Omega) = \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{K}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{K/2}\prod_{i=1}^{K}\lambda_i^{1/2}} . \qquad (8)$$

Through such a two-step PCA and Bayesian analysis, we can finally separate the two interfering components *T* and *N* from the discriminating feature *I*.

4. EXPERIMENT

We test this method on images of 1195 individuals from the Feret database [3], with two images for each individual. We use images of 495 people for training, and the remaining images of 700 people for testing. So there are totally 990 face images in the training set, 700 face images in the gallery, and 700 face images for probe.

(1) *dp-di* accuracy surface

Results of Bayesian analysis using ML measure in the reduced PCA space is reported in Table 1. $P(\Delta \mid \Omega_I)$ is directly estimated using the DIFS. In Table 1, the vertical direction is the dimensionality of the reduced PCA space (*dp*) and the horizontal direction is the number of intrapersonal eigenvectors selected for DIFS (*di*).

We can see that for each PCA subspace, the recognition accuracy of the DIFS changes with the number

Table 1. Recognition results of Bayesian analysis in the reduced PCA space.

PCA		DIFS (*di*)									
dp	Euclid	10	20	50	100	150	200	250	300	400	490
50	0.773	0.277	0.609	0.937	N/A	N/A	N/A	N/A	N/A	N/A	N/A
100	0.807	0.271	0.581	0.854	0.954	N/A	N/A	N/A	N/A	N/A	N/A
150	0.817	0.276	0.573	0.814	0.909	0.960	N/A	N/A	N/A	N/A	N/A
200	0.821	0.276	0.580	0.813	0.893	0.923	0.953	N/A	N/A	N/A	N/A
250	0.827	0.269	0.571	0.806	0.877	0.910	0.949	0.944	N/A	N/A	N/A
300	0.831	0.271	0.567	0.806	0.879	0.937	0.937	0.944	0.930	N/A	N/A
400	0.829	0.267	0.566	0.803	0.871	0.910	0.923	0.929	0.943	0.916	N/A
500	0.836	0.266	0.563	0.804	0.871	0.907	0.916	0.927	0.931	0.930	0.670
600	0.839	0.266	0.561	0.803	0.869	0.907	0.919	0.926	0.923	0.937	0.897
700	0.840	0.267	0.560	0.803	0.869	0.907	0.920	0.926	0.931	0.927	0.911
900	0.840	0.266	0.560	0.804	0.869	0.907	0.917	0.926	0.930	0.926	0.909
On raw data		0.267	0.559	0.804	0.869	0.907	0.919	0.930	0.930	0.926	0.906

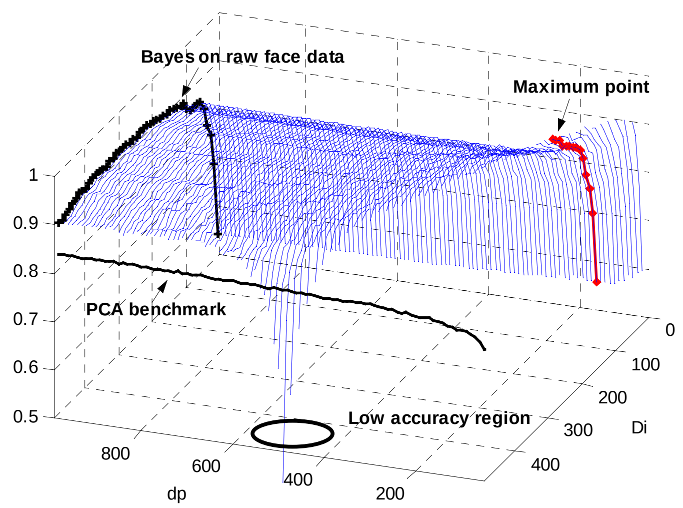


Figure 2. Accuracy curves for Bayesian analysis in the reduced PCA space.

of intrapersonal eigenvectors, thus forms an accuracy curve. The accuracy curves for different PCA subspace then form a *dp*-*di* accuracy surface. We plot the *dp*-*di* accuracy surface in Fig. 2. There are two benchmark curves in the 3D space of Fig. 2. One is the accuracy curve of the traditional PCA method as reported in the second column in Table 1. Comparison with this curve clearly shows the improvement of the Bayesian analysis. The second curve is the DIFS accuracy curve of the standard Bayesian algorithm based on raw image vectors. It is reported in the bottom row of Table 1. We will compare it with the accuracy curves in different PCA spaces.

The shape of *dp*-*di* accuracy surface clearly reflects the effect of noise. When *dp* is small, there is little noise in the PCA subspace. So the recognition accuracy monotonically increases with *di* as more discriminating information *I* is added, and finally reaches the highest point at the full dimensionality of the intrapersonal subspace. However, as *dp* increases, noise begins to appear in the PCA subspace and causes a change in the accuracy curve shape. The curve starts to decrease after reaching a peak point before *di* reaches the full dimensionality. The decrease of accuracy at the end of the curve is because noise distributed on the small eigenvectors is magnified by the inverse of the small eigenvalues as shown in Eq. (3).

This effect of noise is especially severe when both *dp* and *di* are around 495, i.e. the largest possible *di*. In this region, the accuracy becomes as low as 67%. Because of the large *dp*, noise has become a fairly significant problem. When *di* becomes the same size of *dp*, all the energy in the PCA subspace, including noise, are selected for the Bayesian analysis. Noise concentrated on the last few very small eigenvectors will be drastically magnified because of the very small eigenvalues.

We plot the highest accuracy of each accuracy curve of different *dp* in Figure 3. The maximum point with 96% accuracy could be found at (*dp*=150, *di*=150). In this PCA subspace, noise has been removed and all of the eigenvectors can be used for Bayesian recognition.

(2) Recognition performance comparison

In this section we compare the improved Bayesian algorithm in the reduced PCA space with the standard Bayesian algorithm. Accumulative accuracies using both ML and MAP measures are reported in Figure 4. In the improved algorithm, face data is first projected onto the PCA subspace with a dimension 150. ML and MAP measures are computed using the same dimensionality 150. The performance of traditional PCA method is also reported. The results show that the Bayesian method is clearly better than the PCA method. By adding PCA in front of a Bayesian analysis, the performance can be further improved.

4. CONCLUSION

In this paper, we propose an improved Bayesian algorithm in the reduced PCA space. Using this method, both transformation variations and noise are separated from the discriminating features. The new method is much less sensitive to noise than the standard Bayes. Experiments on the Feret database demonstrate the improvement of the new approach.

ACKNOWLEDGMENTS

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong SAR (Project no. CUHK 4190/01E).

REFERENCES

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

[2] M. Turk and A. Pentland, "Eigenfaces for Recognition", *J. of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.

[3] P. J. Phillips, H. Moon, and S. A. Rozvi, “The Feret Evaluation Methodolody for Face Recognition Algorithms”, *IEEE Trans. PAMI*, Vol. 22, No. 10, pp. 1090-1104, Oct. 2000.

[4] W. Zhao, R. Chellappa, A. Rosenfeld and P. Phillips. “Face Recognition: A Literature Survey”, *Technical Report*, 2000.

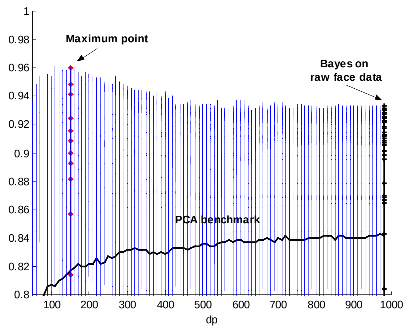


Figure 3. Highest accuracy of Bayes analysis in each PCA space.

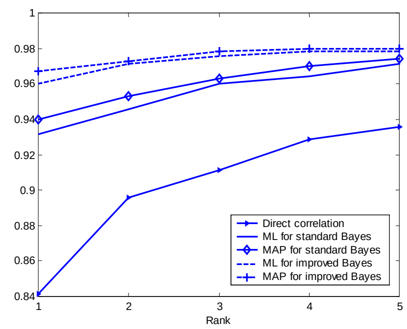


Figure 4. Accumulative accuracy of standard Bayes, improved Bayes and direct correlation.