A Unified Framework for Subspace Face Recognition

Xiaogang Wang, *Student Member*, *IEEE*, and
Xiaoou Tang, *Senior Member*, *IEEE*

Abstract—PCA, LDA, and Bayesian analysis are the three most representative subspace face recognition approaches. In this paper, we show that they can be unified under the same framework. We first model face difference with three components: intrinsic difference, transformation difference, and noise. A unified framework is then constructed by using this face difference model and a detailed subspace analysis on the three components. We explain the inherent relationship among different subspace methods and their unique contributions to the extraction of discriminating information from the face difference. Based on the framework, a unified subspace analysis method is developed using PCA, Bayes, and LDA as three steps. A 3D parameter space is constructed using the three subspace dimensions as axes. Searching through this parameter space, we achieve better recognition performance than standard subspace methods.

Index Terms—Face recognition, subspace analysis, PCA, LDA, Bayesian analysis, eigenface.

1 Introduction

Subspace analysis methods, such as PCA, LDA, and Bayes, have been extensively studied for face recognition in recent years. The eigenface method (PCA) [13] uses the Karhunen-Loeve Transform to produce the most expressive subspace for face representation and recognition. LDA or Fisher Face [1], [12], [19] is an example of the most discriminating subspace methods. It seeks a set of features best separating face classes. Another important subspace method is the Bayesian algorithm using probabilistic subspace [8], [9]. Different from other subspace methods, which classify the probe face image into L classes for L individuals, the Bayesian algorithm casts the face recognition task into a binary pattern classification problem with each of the two classes, intrapersonal and extrapersonal variations, modeled as a Gaussian distribution.

Many other subspace methods are modifications or extensions of the above three methods. Independent Component Analysis (ICA) [18], nonlinear PCA (NLPCA) [4], and Kernel PCA (KPCA) [16] are all the generalizations of PCA to address higher order statistical dependencies. Kernel-based Fisher Discriminant Analysis (KFDA) [7] extracts nonlinear discriminating features. To improve the generalization ability of LDA on different data sets, several modifications are proposed, including the Enhanced FLD model [5], LDA mixture model [3], and direct LDA [17]. In addition to directly processing image appearance, subspace methods can also be applied to other features, such as Gabor features [6], [15].

In this work, we develop a unified framework to study the three major subspace face recognition methods: PCA, LDA, and Bayes. PCA is an evaluation benchmark for face recognition. Both LDA and Bayes have achieved superior performance in the FERET test [11]. A unified framework on the three methods will greatly help to understand the family of subspace methods for further improvement.

- *The authors are with the Department of Information Engineering, The Chinese University of Hong Kong, 826B, HSH, Shatin, Hong Kong. E-mail: {xgwang1, xtang}@ie.cuhk.edu.hk.*

Manuscript received 30 Dec. 2002; revised 13 Aug. 2003; accepted 26 Jan. 2004.
Recommended for acceptance by R. Basri.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 118060.

We start our investigation by defining a face difference model with three components: intrinsic difference, transformation difference, and noise. A unified framework is then constructed by using the face difference model and a detailed subspace analysis on the three components. Using this framework, we explain the inherent relationship among different subspace methods and their unique contributions to the extraction of discriminating information from the face difference. Starting from the framework, a unified subspace analysis method is proposed using PCA, Bayes, and LDA as three steps. It is shown that the subspace dimension of each method can affect the recognition performance significantly. It is a trade off on how much noise and transformation difference is excluded, and how much intrinsic difference is retained. This eventually leads to the construction of a 3D parameter space that uses the three subspace dimensions as axes. Searching through this parameter space, we achieve better recognition performance than the standard subspace methods, which are all limited to local areas of the parameter space.

2 Review of the PCA, LDA, and Bayesian Algorithms

In appearance-based approaches, a 2D face image is viewed as a vector in the image space. A set of face image samples $\{\vec{x}_i\}$ can be represented as an N by M matrix $X = [\vec{x}_1, \ldots, \vec{x}_M]$, where N is the number of pixels in the images and M is the number of samples. Each face image $\vec{x}_i$ belongs to one of the L face classes $\{X_1, \ldots, X_L\}$, with $\ell(\vec{x}_i)$ as the class label of $\vec{x}_i$. When a probe $\vec{T}$ is the input, the face recognition task is to find its class label in the database. In this section, we briefly review the three subspace methods.

2.1 PCA

In PCA, a set of eigenfaces are computed from the eigenvectors of sample covariance matrix C,

$$C = \sum_{i=1}^{M} \left(\vec{x}_i - \vec{m}\right)\left(\vec{x}_i - \vec{m}\right)^T, \tag{1}$$

where $\vec{m}$ is the mean face of the sample set. The eigenspace $U = \left[\vec{u}_1, \ldots \vec{u}_K\right]$ is spanned by the K eigenfaces with the largest eigenvalues. For recognition, the prototype $\vec{P}$ for each face class and the probe $\vec{T}$ are projected onto U to get the weight vectors $\vec{w}_p = U^T(\vec{P} - \vec{m})$ and $\vec{w}_T = U^T(\vec{T} - \vec{m})$. The face class is found to minimize the distance

$$\varepsilon = ||\vec{w}_T - \vec{w}_p|| = ||U^T\left(\vec{T} - \vec{P}\right)||. \tag{2}$$

2.2 LDA

LDA tries to find the subspace that best discriminates different face classes by maximizing the between-class scatter matrix S_b, while minimizing the within-class scatter matrix S_w in the projective subspace. S_w and S_b are defined as

$$S_w = \sum_{i=1}^{L} \sum_{\vec{x}_k \in X_i} \left(\vec{x}_k - \vec{m}_i\right)\left(\vec{x}_k - \vec{m}_i\right)^T, \tag{3}$$

$$S_b = \sum_{i=1}^{L} n_i\left(\vec{m}_i - \vec{m}\right)\left(\vec{m}_i - \vec{m}\right)^T, \tag{4}$$

where $\vec{m}_i$ is the mean face for the individual class X_i and n_i is the number of samples in class X_i.

LDA subspace is spanned by a set of vectors W, satisfying

$$W = \arg\max \left| \frac{W^T S_b W}{W^T S_w W} \right|. \quad (5)$$

W can therefore be constructed by the eigenvectors of $S_w^{-1} S_b$. Computing the eigenvectors of $S_w^{-1} S_b$ is equivalent to simultaneous diagonalization of S_w and S_b [2]. First, compute the eigenvector matrix Φ and eigenvalue matrix Θ of S_w. Then, project the class centers onto $\Phi\Theta^{-1/2}$, thus S_b is transformed to $K_b = \Theta^{-1/2}\Phi^T S_b \Phi \Theta^{-1/2}$. After computing the eigenvector matrix Ψ and eigenvalue matrix Λ of K_b, the projection vectors of LDA can be defined as $W = \Phi\Theta^{-1/2}\Psi$. For recognition, the linear discriminant function is computed as

$$d\left(\vec{T}\right) = W^T\left(\vec{T} - \vec{P}\right). \quad (6)$$

The face class is chosen to minimize $||d||$. To avoid degeneration of S_w, most LDA methods first reduce the data dimensionality by PCA, then apply discriminant analysis in the reduced PCA subspace.

2.3 Bayesian 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. The MAP similarity between two images is defined as the intrapersonal a posterior probability

$$S(I_1, I_2) = P(\Omega_I|\Delta) = \frac{P(\Delta|\Omega_I)P(\Omega_I)}{P(\Delta|\Omega_I)P(\Omega_I) + P(\Delta|\Omega_E)P(\Omega_E)}. \quad (7)$$

To estimate $P(\Delta|\Omega_I)$, PCA on the intrapersonal difference set $\{\Delta|\Delta \in \Omega_I\}$ decomposes the image difference space into intrapersonal principal subspace F, spanned by the di largest intrapersonal eigenvectors, and its orthogonal complementary subspace $\bar{F}$, with the dimension N-di. $P(\Delta|\Omega_I)$ can be estimated as the product of two independent marginal Gaussian densities in F and $\bar{F}$,

$$\begin{aligned} P(\Delta|\Omega_I) &= \left[\frac{\exp\left(-\frac{1}{2}d_F(\Delta)\right)}{(2\pi)^{di/2}\prod_{i=1}^{di}\lambda_i^{1/2}}\right]\left[\frac{\exp(-\varepsilon^2(\Delta)/2\rho)}{(2\pi\rho)^{(N-di)/2}}\right] \\ &= \frac{\exp\left[-\frac{1}{2}(d_F(\Delta) + \varepsilon^2(\Delta)/\rho)\right]}{\left[(2\pi)^{di/2}\prod_{i=1}^{di}\lambda_i^{1/2}\right]\left[(2\pi\rho)^{(N-di)/2}\right]}. \end{aligned} \quad (8)$$

Here, $d_F(\Delta) = \sum_{i=1}^{di}\frac{y_i^2}{\lambda_i}$ is a Mahalanobis distance in F, referred as "distance-in-feature-space" (DIFS). y_i is the principal component of Δ projecting to the ith intrapersonal eigenvector, and λ_i is the corresponding eigenvalue. $\varepsilon^2(\Delta)$ is defined as "distance-from-feature-space" (DFFS), equivalent to PCA residual error in $\bar{F}$. ρ is the average eigenvalue in $\bar{F}$. $P(\Delta|\Omega_E)$ can be estimated in a similar way in the extrapersonal subspace computed from $\{\Delta|\Delta \in \Omega_E\}$.

An alternative maximum likelihood (ML) measure, using the intrapersonal likelihood $S'(\Delta) = P(\Delta|\Omega_I)$ alone, is proven to be simpler but almost as effective as the MAP measure [9]. In recognition, all the parameters in (8) are constant except $d_F(\Delta)$ and $\varepsilon^2(\Delta)$. So, it is equivalent to evaluate the distance

$$D_I = d_F(\Delta) + \varepsilon^2(\Delta)/\rho. \quad (9)$$

3 A Unified Framework

The three methods reviewed in Section 2 are developed with different specific considerations. In this study, we formulate an indepth subspace analysis to construct a unified framework for the three methods. Under this framework, we study the inherent connections of the three methods in order to understand the reason behind the different performance of each method. A unified subspace analysis method is then proposed based on the theorems and insights provided by the unified framework.

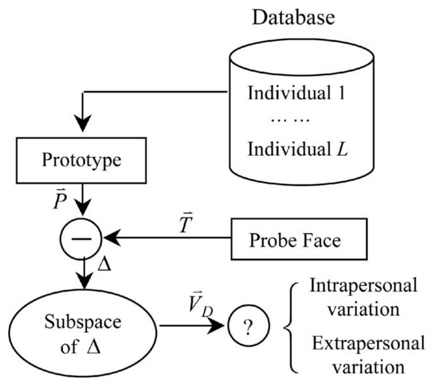

Fig. 1. Diagram of the unified framework for face recognition.

To construct the framework, let us first look at the matching criterions and focus on the difference $\Delta = \vec{T} - \vec{P}$ between the probe $\vec{T}$ and the prototype $\vec{P}$. From (2), (6), and (9), we can see that the recognition process of the three methods can be described by the same framework as shown in Fig. 1. When a probe face image $\vec{T}$ is the input, we compute the difference Δ between $\vec{T}$ and each class prototype $\vec{P}$. Then, Δ is projected into an image subspace to compute the feature vector $\vec{V}_D$. Finally, based on the feature vector and the specific distance metric, Δ is classified as intrapersonal or extrapersonal variations.

The two key components of this framework are the image difference Δ and its subspace. Especially, using a set of theorems, we will show that all the three subspaces for PCA, Bayes, and LDA can be computed from the face difference set instead of the original image set. This is the central point of the unified framework. We model Δ by three key components: intrinsic difference ($\tilde{I}$) that discriminates different face identity; transformation difference ($\tilde{T}$), arising from all kinds of transformations, such as expression and illumination changes; noise ($\tilde{N}$), which randomly distributes in the face images. The intrapersonal variation Ω_I is composed of $\tilde{T}$ and $\tilde{N}$, since it comes from the same individual. Notice that extrapersonal variation Ω_E is not equivalent to $\tilde{I}$. In Ω_E, $\tilde{I}$, $\tilde{T}$, and $\tilde{N}$ exist together, since $\tilde{T}$ and $\tilde{N}$ cannot be canceled when computing the difference of the images of two individuals.

$\tilde{T}$ and $\tilde{N}$ are the two components deteriorating recognition performance. Normally, $\tilde{N}$ is of small energy. The main difficulty for face recognition comes from $\tilde{T}$, which can change the face appearance substantially [10]. A successful face recognition algorithm should be able to reduce the energy of $\tilde{T}$ as much as possible without sacrificing much of $\tilde{I}$. To improve recognition efficiency, it is also useful to compact $\tilde{I}$ onto a small number of features. In the following, we propose a set of theorems to analyze the properties of the three subspaces for PCA, LDA, and Bayes in order to find out how they suppress the $\tilde{T}$ and $\tilde{N}$ components, and extract the $\tilde{I}$ component in their respective subspaces. Due to length limit, detailed proofs of the theorems are not shown in this paper. They can be found in [14].

3.1 PCA Subspace

Eigenfaces are computed from the ensemble covariance matrix C. Equation (1) shows that C is derived from all the training face

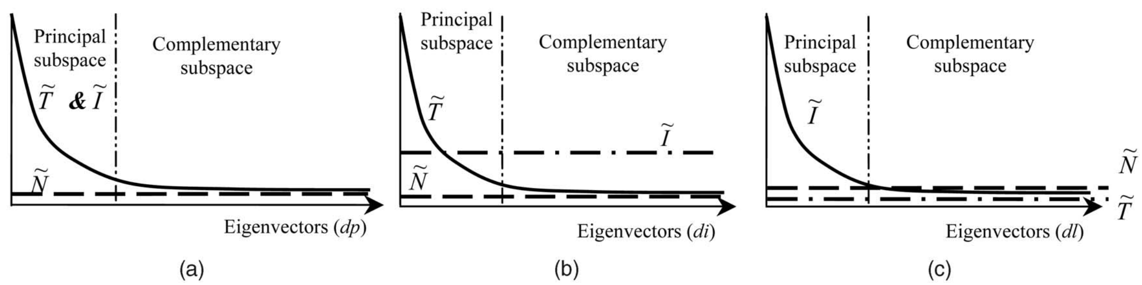

Fig. 2. Energy distribution of the three components $\tilde{I}$, $\tilde{T}$, and $\tilde{N}$ on (a) eigenvectors in the PCA subspace, (b) the intrapersonal subspace, and (c) the LDA subspace.

images subtracting of the mean face. We have shown that C can also be formulated as [14]

$$C = \sum_{i=1}^{M}\sum_{j=1}^{M}\left(\vec{x}_i - \vec{x}_j\right)\left(\vec{x}_i - \vec{x}_j\right)^T. \quad (10)$$

Therefore, the eigenvectors for $\{\vec{x}_i\}$ can also be computed as the eigenvectors for the face difference set $\{(\vec{x}_i - \vec{x}_j)\}$, containing all the difference between any pair of face images in the training set.

Theorem 1. *The PCA subspace characterizes the distribution of face difference between any two face images, which may belong to the same individual or different individuals.*

3.2 Intrapersonal and Extrapersonal Subspaces

In the Bayesian algorithm, the eigenvectors of intrapersonal subspace are computed from the difference set $\{(\vec{x}_i - \vec{x}_j)|\ell(\vec{x}_i) = \ell(\vec{x}_j)\}$, for which the covariance matrix is

$$C_I = \sum_{\ell(\vec{x}_i)=\ell(\vec{x}_j)}\left(\vec{x}_i - \vec{x}_j\right)\left(\vec{x}_i - \vec{x}_j\right)^T. \quad (11)$$

The eigenvectors of extrapersonal subspace are derived from the difference set $\{(\vec{x}_i - \vec{x}_j|\ell(\vec{x}_i) \neq \ell(\vec{x}_j)\}$, with covariance matrix

$$C_E = \sum_{\ell(\vec{x}_i)\neq\ell(\vec{x}_j)}\left(\vec{x}_i - \vec{x}_j\right)\left(\vec{x}_i - \vec{x}_j\right)^T. \quad (12)$$

Comparing with (10), we have $C = C_I + C_E$. Since the sample number for C_E is far greater than the sample number of C_I, the energy of C_E usually dominantes the computation of C. Therefore, the extrapersonal eigenfaces are similar to the PCA eigenfaces.

Theorem 2. *The intrapersonal subspace and extrapersonal subspace are both contained in the PCA eigenspace, with the extrapersonal eigenfaces being similar to the PCA eigenfaces.*

Experiments in [9] have shown that the ML measure using the intrapersonal subspace alone is almost as effective as the MAP measure using the two subspaces. In this paper, we prove that the extrapersonal subspace is similar to the PCA subspace and cannot contribute much to separating $\tilde{T}$ and $\tilde{I}$ since $\tilde{T}$ and $\tilde{I}$ are coupled together in Ω_E. The improvement of the Bayesian algorithm over the PCA method benefits mostly from the intrapersonal subspace. Therefore, given the significant additional computational cost of MAP for little improvement over ML, we focus our study on ML only in the framework.

3.3 LDA Subspace

The LDA subspace is derived from S_w and S_b. Similar to the previous analysis of the PCA and Bayesian approaches, we can also study the LDA subspace using image difference. For simplicity, we assume that each class has the same sample number n. Similar to the proof of Theorem 1, we have

$$S_w = \sum_{i=1}^{L}\sum_{\vec{x}_{k_1},\vec{x}_{k_2}\in X_i}\left(\vec{x}_{k_1} - \vec{x}_{k_2}\right)\left(\vec{x}_{k_1} - \vec{x}_{k_2}\right)^T = C_I \quad (13)$$

and

$$\begin{aligned} S_b &= \sum_{i=1}^{L} n\left(\vec{m}_i - \vec{m}\right)\left(\vec{m}_i - \vec{m}\right)^T \\ &= \frac{n}{2L}\sum_{i=1}^{L}\sum_{j=1}^{L}\left(\vec{m}_i - \vec{m}_j\right)\left(\vec{m}_i - \vec{m}_j\right)^T. \end{aligned} \quad (14)$$

So, we have the following theorem.

Theorem 3. *S_w is identical to C_I, the covariance matrix of the intrapersonal subspace, which characterizes the distribution of face variation for the same individual. Using the mean face image to describe each individual class, S_b, characterize the distribution of the difference between any two mean face images.*

3.4 Comparison of the Three Subspaces

Although PCA and LDA were initially developed based on statistical distribution of original face images, the above theorems show that they can also be computed from the distribution of face differences. We now can compare these subspaces using the same face difference model.

As shown in Fig. 2a, since PCA subspace characterizes the difference between any two face images, it concentrates both $\tilde{T}$ and $\tilde{I}$ as structural signals on a small number of principal eigenvectors. By selecting the principal components, most noise encoded on the large number of trailing eigenvectors is removed from $\tilde{T}$ and $\tilde{I}$. Because of the continuing existence of $\tilde{T}$, the PCA subspace is not ideal for face recognition.

For the Bayesian algorithm, the intrapersonal subspace plays a critical role. Since the intrapersonal variation only contains $\tilde{T}$ and $\tilde{N}$, PCA on the intrapersonal variation arranges the eigenvectors according to the energy distribution of $\tilde{T}$, as shown in Fig. 2b. When we project a face difference Δ (either intrapersonal or extrapersonal) onto the intrapersonal subspace, most energy of the $\tilde{T}$ component will concentrate on the first few largest eigenvectors, while the $\tilde{I}$ and $\tilde{N}$ components distribute over all of the eigenvectors. This is because $\tilde{I}$ and $\tilde{N}$ are somewhat independent of $\tilde{T}$, which forms the principal vectors of the intrapersonal subspace. In (9), the Mahalanobis distance in F weights the feature vectors by the inverse of eigenvalues. This effectively reduces the $\tilde{T}$ component since the principal components with large eigenvalues are significantly diminished. $\varepsilon^2(\Delta)$ is also an effective feature for recognition since it throws away most of the $\tilde{T}$ component on the largest eigenvectors, while keeping the majority of $\tilde{I}$ in $\bar{F}$.

TABLE 1
Behavior of Subspaces on Characterizing the Face Image Difference

Algorithm	Subspace	Decompose Face Image Difference	
		Principal subspace	Complementary subspace
PCA	Eigenspace	$\tilde{I}+\tilde{T}$	$\tilde{N}$
LDA	LDA subspace	$\tilde{I}$	$\tilde{T}+\tilde{N}$
Bayesian	Intrapersonal subspace	$\tilde{T}$	$\tilde{I}+\tilde{N}$
	Extrapersonal subspace	$\tilde{I}+\tilde{T}$	$\tilde{N}$

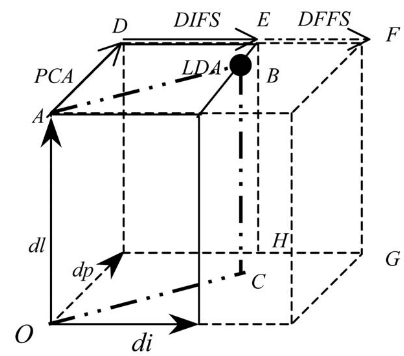

Fig. 4. 3D parameter space.

The Bayesian algorithm successfully separates $\tilde{T}$ from $\tilde{I}$. However, $\tilde{I}$ and $\tilde{N}$ are still coupled on the small eigenvectors. When normalized by the small eigenvalues, the effect of $\tilde{N}$ could be significantly enlarged in the probabilistic measure. Another drawback of the Bayesian algorithm is that the intrinsic difference $\tilde{I}$ is not compacted and, thus, spreads over F and $\bar{F}$. This leads to high computational cost. The computational cost of DFFS is equivalent to the correlation of two high dimensional image vectors.

Finally, we look at the LDA subspace. The LDA procedure can be divided into three steps. First, PCA is used to reduce the data dimensionality. Noise is significantly reduced in this step. In the second step, to whiten S_w, we first compute its eigenvector matrix Φ and eigenvalue matrix Θ. From Theorem 3, we know that Φ spans the intrapersonal subspace, therefore Θ essentially represents the energy distribution of $\tilde{T}$. The whitening process projects data onto intrapersonal subspace Φ and normalizes it by $\Theta^{-1/2}$. So, Theorem 3 effectively proves that this step reduces $\tilde{T}$ in a manner similar to the key component (ML measure) in the Bayesian analysis.

In the third step of LDA, PCA is again applied to the whitened class centers. When averaging images in each class to compute the class centers, the noise $\tilde{N}$ is further reduced. This is useful since $\tilde{N}$ may have been enlarged in the whitening process. Since both $\tilde{T}$ and $\tilde{N}$ have been reduced up to this point, the main energy in the class centers is $\tilde{I}$, which spreads over the entire eigenvector axes after the whitening as shown in Fig. 2b. PCA on the class centers therefore serves two purposes. First, it can further reduce the noise. Second, it compacts the energy of $\tilde{I}$ into a small number of principal components, as shown in Fig. 2c.

The subspace analysis results of the three methods on the face difference model are summarized in Table 1. Our analysis shows that PCA and the key component of Bayesian analysis can be viewed as the intermediate steps of LDA. In theory, the LDA subspace should be able to separate $\tilde{I}$ from $\tilde{T}$ and $\tilde{N}$. However, the original LDA has not achieved this because it uses fixed parameters in each step and, thus, fails to take full advantage of the information provided by the PCA and Bayesian analysis steps. Our subspace analysis provides a clear understanding of the unique contribution of each of the three subspaces to the processing of the face difference model. Especially, our study shows that the degree of control over the $\tilde{T}$ and $\tilde{N}$ components in the face difference by each of the three methods depends on the dimensionality of the three subspaces, PCA subspace (*dp*), intrapersonal subspace (*di*), and LDA subspace (*dl*) in Fig. 2.

Using the above framework, we propose a unified subspace analysis method as shown in Fig. 3. Using the three subspace dimensions as parameter axes, it provides a new 3D parameter space to improve the recognition performance, as shown in Fig. 4. It controls $\tilde{I}$, $\tilde{T}$, and $\tilde{N}$ components in the face difference by adjusting the dimensionality of the three subspaces. The choice of the three parameters greatly affects the system performance. The original PCA, LDA, and Bayes methods only occupy some local lines or areas in the 3D parameter space. PCA changes parameters in the dp direction on line AD. DIFS and DFFS of the Bayesian algorithm change on the line DEF in the di direction. Fisher Face [1] corresponds to point $B(dp = di = M - L, dl = L - 1)$ in the graph. All these methods change parameters only in the local regions. However, for the new algorithm, optimal parameters may be searched in the full 3D space. We can clearly see this advantage in the experiments.

4 Experiment

In this section, we conduct experiments on the FERET face database [11]. We select 1,195 people from the FERET database, with two face images (FA/FB) for each person. Images of 495 people are randomly selected for training, and the remaining 700 people are used for testing. So, there are a total of 990 face images in the training set, 700 face images in the gallery, and 700 face images for probe. In the preprocessing procedure, all the images are normalized by the eye for scaling, translation, and rotation, such that the eye centers are in fixed positions. A 27×41 mask template is used to remove the background and most of the hair. Histogram equalization is applied to the face images for photometric normalization.

1. Project face data to PCA subspace and adjust the PCA dimension (*dp*) to reduce most noise.
2. Apply Bayesian analysis in the reduced PCA subspace and adjust the dimension (*di*) of intrapersonal subspace to reduce the transformation difference.
3. Project class centers to the *di* intrapersonal eigenvectors, and normalize the projections by intrapersonal eigenvalues to compute the whitened class centers. Apply PCA to the whitened class centers to compute a discriminant feature vector of dimension *dl*. The face class is recognized using the *dl* discriminant features.

Fig. 3. Unified subspace analysis method using the framework.

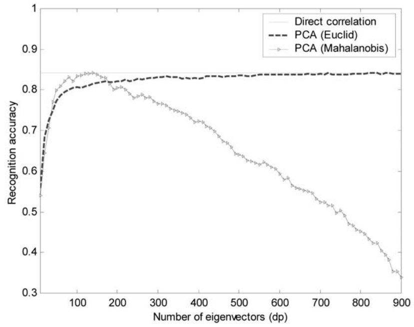

Fig. 5. PCA recognition accuracy.

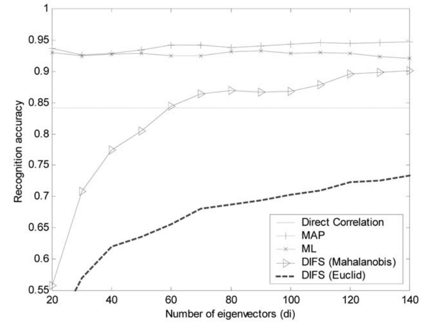

Fig. 6. Bayes recognition accuracy.

4.1 PCA Experiment

Euclid and Mahalanobis distance measures are used for the PCA recognition. We use direct correlation (84.1 percent) as a benchmark since it is essentially a direct use of face difference without subspace analysis. The overall results in Fig. 5 show that PCA is no better than direct correlation in terms of recognition accuracy. When using the Mahalanobis distance measure, it reaches peak accuracy (84.3 percent) with around 150 eigenvectors, and then drops with further increase of dimensionality. The high dimensional components with small eigenvalues are significantly magnified in whitening. Since these dimensions tend to contain more noise than structural signal, they will deteriorate the recognition results.

4.2 Bayesian Experiment

Experimental results for the Bayesian algorithm are reported in Fig. 6. It has achieved around 93 percent accuracy with 10 percent improvement over direct correlation. The eigenvectors of the intrapersonal subspace are arranged by the energy of $\tilde{T}$. When only a small number of eigenvectors are selected, the principal subspace does not have enough information on $\tilde{I}$, so the accuracy of DIFS, a main step of Bayes, is low (below 60 percent for 20 eigenvectors). However, the lost information can be compensated from DFFS in $\bar{F}$. When we use the Euclid instead of Mahalanobis distance measure for DIFS, the accuracy drops greatly and becomes even worse than PCA. This shows the effectiveness of whitening in the Bayesian analysis. The accuracy of ML using DIFS + DFFS is high since it combines the two components together. The MAP measure using both intrapersonal and extrapersonal subspaces is around 94 percent, slightly better than ML.

4.3 Bayesian Analysis in Reduced PCA Space

After comparing the PCA and Bayesian methods individually, we now investigate how these two subspace dimensions interact with each other. We first apply PCA on the face data to reduce the dimensionality and remove the noise. Then, the Bayesian analysis using DIFS in the intrapersonal subspace is implemented in the reduced PCA subspace. This corresponds to the dp-di plane in the 3D space in Fig. 4. Results are reported in Table 2. Since there are 990 face images and 495 classes in the training set, the rank of S_w is bounded by 495. The maximum value for *di* is $\min\{d_p, 495\}$.

The shape of the dp-di accuracy table clearly reflects the effect of noise $\tilde{N}$. When *dp* is small, there is little noise in the PCA subspace. So, the recognition accuracy monotonically increases with di as more discriminating information $\tilde{I}$ is added. However, as *dp* increases, noise begins to appear in the PCA subspace and starts to affect the accuracy. The accuracy begins to decrease after reaching a peak point before *di* reaches the full dimensionality. The decrease in accuracy for large *di* is because noise distributed on the small eigenvectors is magnified by the inverse of the small eigenvalues. Interestingly, for this training set, the parameters proposed in Fisherface [1] lead to very low accuracy around $(dp = 495, di = 495)$. This shows the importance of parameter selection for a given training set.

We plot the highest accuracy of each accuracy row of different *dp* in Fig. 7. The maximum point with 96 percent 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.

TABLE 2
Recognition Accuracy of Bayesian Analysis in the Reduced PCA Subspace

	Euclid	*dp*	DIFS (*di*)									
			10	20	50	100	150	200	250	300	400	490
PCA	0.773	50	0.277	0.609	0.937	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	0.807	100	0.271	0.581	0.854	0.954	N/A	N/A	N/A	N/A	N/A	N/A
	0.817	150	0.276	0.573	0.814	0.909	0.960	N/A	N/A	N/A	N/A	N/A
	0.821	200	0.276	0.580	0.813	0.893	0.923	0.953	N/A	N/A	N/A	N/A
	0.831	300	0.271	0.567	0.806	0.879	0.937	0.937	0.944	0.930	N/A	N/A
	0.836	500	0.266	0.563	0.804	0.871	0.907	0.916	0.927	0.931	0.930	0.670
	0.840	700	0.267	0.560	0.803	0.869	0.907	0.920	0.926	0.931	0.927	0.911
	0.840	900	0.266	0.560	0.804	0.869	0.907	0.917	0.926	0.930	0.926	0.909
Bayes (DIFS in ML)on raw data			0.267	0.559	0.804	0.869	0.907	0.919	0.930	0.930	0.926	0.906

The vertical direction is the dimension of the PCA subspace (dp) *and the horizontal direction is the dimension of the intrapersonal subspace* (di).

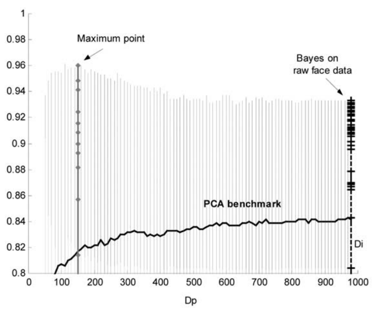

Fig. 7. Highest accuracy of Bayesian analysis in each PCA subspace.

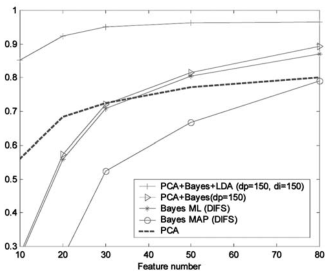

Fig. 9. Recognition accuracies using small feature number for each step of the framework.

4.4 Extract Discriminant Features from Intrapersonal Subspace

We now investigate the effect of the third dimension dl in the 3D parameter space. For ease of comparison, we choose four representative points on the dp-di surface, and report the accuracy along the dimension of dl as shown in Fig. 8. The curves first increase to a maximum point and then drop with further increase of dl. For traditional LDA, the dl dimension is usually chosen as $L - 1$, which corresponds to the last point of the curve with $di = 495$. The result is clearly much lower than the highest accuracy in Fig. 8. As discussed in Section 3, this dimension mainly serves to compact $\tilde{I}$ and remove more noise $\tilde{N}$ so dl should be reasonably small instead of being fixed by L. The best results on the curve plots are indeed better than only using the first two dimensions. Since the computational cost for face recognition is proportional to the feature number used, we compare the recognition accuracies using small feature number for each step of the framework, as shown in Fig. 9. For Bayes, DIFS measure is used for comparison, since DIFS + DFFS measure is costly to compute even for a small feature number. The results clearly demonstrate the improvement on recognition efficiency by the addition of dl dimension.

As shown by these experiments, although we have not explored the entire 3D parameter space, better results are already found comparing to the standard subspace methods. A careful investigation of the entire parameter space should lead to further improvement.

5 Conclusion

Starting from the face difference model that decomposes a face difference into three major components, intrinsic difference $\tilde{I}$, transformation difference $\tilde{T}$, and noise $\tilde{N}$, we unify the three major subspace face recognition methods, PCA, Bayes, and LDA under the same framework. Using this framework, we study how each of the three methods contributes to the extraction of discriminating information $\tilde{I}$ in the face difference. This eventually leads to the construction of a 3D parameter space that uses the three subspace dimensions as axes. Searching through this parameter space, we achieve better recognition performance than the standard subspace methods. Searching through the whole 3D parameter space may be time consuming. A possible strategy is suggested by the steps of our experiments. First, find high accuracy points on the PCA (Mahalanobis) curve to narrow down dp, then use the found dp to compute the di curve to find the best di value and then choose dl according to the accuracy curve in LDA subspace. Thus, by choosing dp, di, and dl sequentially, we only need to test $\mathrm{o}(\max(dp) + \max(di) + \max(dl))$ points instead of $\mathrm{o}(\max(dp) \times \max(di) \times \max(dl))$ points. In this framework, the intrapersonal variation is modeled as Gaussian distribution. However, for significant $\tilde{T}$, such as large changes in pose, this assumption may break down. One way to solve this problem is to first "normalize" the large lighting and pose changes. Another way is to model the intrapersonal difference by more complex distributions instead of a single Gaussian distribution.

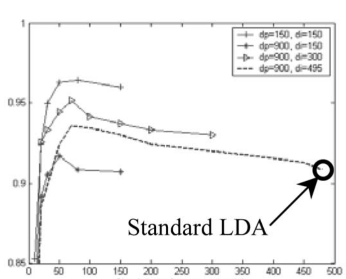

Fig. 8. Accuracies using different number of discriminant features extracted from intrapersonal subspace.

Acknowledgments

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, CUHK 4224/03E, and N_CUHK 409/03).

References

[1] P.N. Belhumeur, J. Hespanda, and D. Kriegeman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.

[2] K. Fukunnaga, *Introduction to Statistical Pattern Recognition.* Academic Press, second ed., 1991.

[3] H. Kim, D. Kim, and S.Y. Bang, "Face Recognition Using LDA Mixture Model," *Proc. Int'l Conf. Pattern Recognition,* pp. 486-489, 2002.

[4] M.A. Kramer, "Nonlinear Principle Components Analysis Using Autoassociative Neural Networks," *Am. Instit. Chemical Eng. J.,* vol. 32, no. 2, p. 1010, 1991.

[5] C. Liu and H. Wechsler, "Enhanced Fisher Linear Discriminant Models for Face Recognition," *Proc. Int'l Conf. Pattern Recognition,* vol. 2, pp. 1368-1372, 1998.

[6] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Analysis for Face Recognition," *IEEE Trans. Image Processing,* vol. 11, no. 4, pp. 467-476, Apr. 2002.

[7] Q. Liu, R. Huang, H. Lu, and S. Ma, "Kernel-Based Optimized Feature Vectors Selection and Discriminant Analysis for Face Recognition," *Proc. Int'l Conf. Pattern Recognition,* pp. 362-365, 2002.

[8] B. Moghaddam, "Principle Manifolds and Probabilistic Subspace for Visual Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 6, pp. 780-788, June 2002.

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

[10] Y. Moses, Y. Adini, and S. Ullman, "Face Recognition: the Problem of Compensating for Changes in Illumination Direction," *Proc. European Conf. Computer Vision,* pp. 286-295, 1994.

[11] P.J. Phillips, H. Moon, and S.A. Rozvi, "The FERET Evaluation Methodolody for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[12] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 8, pp. 831-836, Aug. 1996.

[13] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.

[14] X. Wang and X. Tang, "Unified Subspace Analysis for Face Recognition," *Proc. Int'l Conf. Computer Vision,* pp. 679-686, 2003.

[15] X. Wang and X. Tang, "Bayesian Face Recognition Using Gabor Features," *Proc. ACM SIGGM 2003 Multimedia Biometrics Methods and Applications Workshop (WBMA),* Nov. 2003.

[16] M. Yang, N. Ahuja, and D. Kriegman, "Face Recognition Using Kernel Eigenfaces," *Proc. Int'l Conf. Image Processing,* vol. 1, pp. 37-40, 2000.

[17] H. Yu and J. Yang, "A Direct LDA Algorithm for High-Dimensional Data-with Application to Face Recognition," *Pattern Recognition,* vol. 34, pp. 2067-2070, 2001.

[18] P.C. Yunen and J.H. Lai, "Face Representation Using Independent Component Analysis," *Pattern Recognition,* vol. 35, pp. 1247-1257, 2002.

[19] W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical Performance Analysis of Linear Discriminant Classifiers," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 164-169, 1998.