Unified Subspace Analysis for Face Recognition

Xiaogang Wang and Xiaoou Tang
Department of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
{xgwang1, xtang}@ie.cuhk.edu.hk

Abstract

- PCA, LDA and Bayesian analysis are three of the most representative subspace based face recognition approaches. We show that they can be unified under the same framework. Starting from the framework, a unified subspace analysis is developed using PCA, Bayes, and LDA as three steps. It achieves better performance than the standard subspace methods.

Notation

Face data vector length: N

Training face images: $X = [\vec{x}_1, \ldots, \vec{x}_M]$

Training sample number: M

Face classes: $\{X_1, \ldots, X_L\}$

Face classes number: L

Class label: $\ell(\vec{x}_i)$

Two Kinds of Variation

Extrapersonal variation Ω_E

Intrapersonal variation Ω_I

Face Difference Model

- The difference Δ between two face images can be decomposed into three components.

$\tilde{I}$: Intrinsic difference discriminating face identity

$\tilde{T}$: Transformation difference arising from all kinds of transformations, such as lighting, expression, changes etc.

$\tilde{N}$: Noise

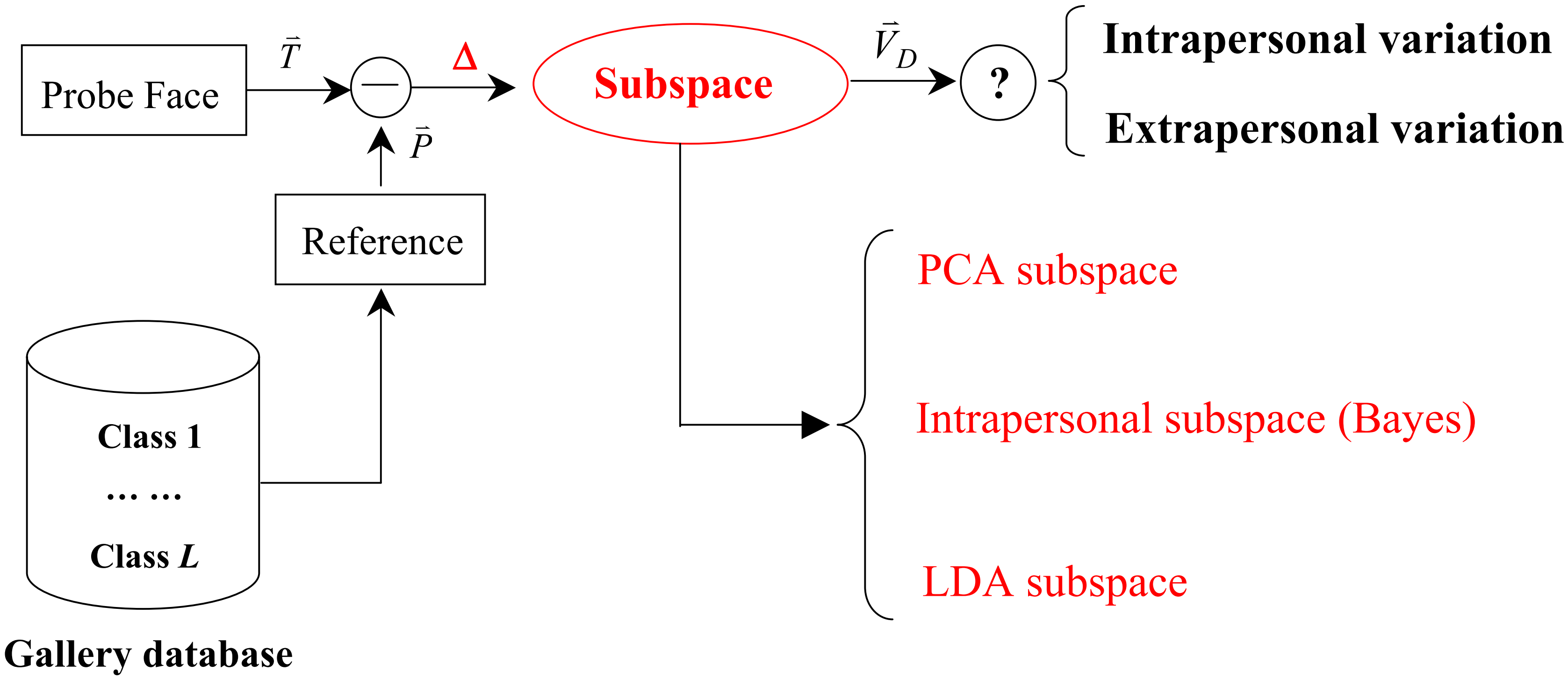
Deteriorating recognition

$$\Delta = \Omega_I + \Omega_E = \tilde{I} + \tilde{T} + \tilde{N}$$

Intrapersonal variation: $\Omega_I = \tilde{T} + \tilde{N}$

Extrapersonal variation: $\Omega_E = \tilde{I} + \tilde{T} + \tilde{N}$

Diagram of the Unified Framework for Subspace Based Face Recognition



Principal Component Analysis (PCA)

- PCA subspace W is computed from the eigen-vectors of covariance matrix of training set $\{X_i\}$

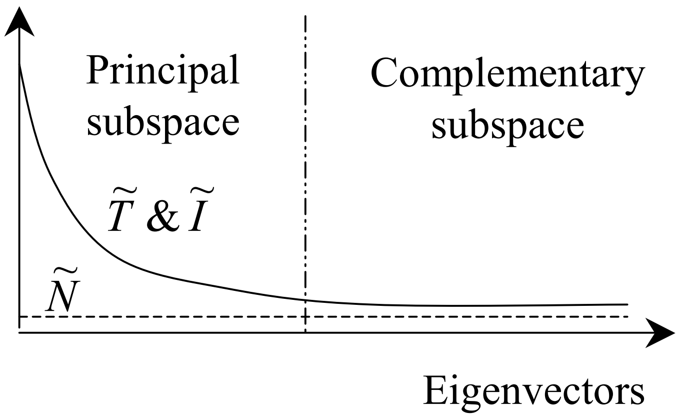
$$C = \sum_{i=1}^{M} (\vec{x}_i - \vec{m})(\vec{x}_i - \vec{m})^T \qquad \vec{m} = \frac{1}{M}\sum_{i=1}^{M} \vec{x}_i$$

- Theorem 1: The PCA subspace characterizes the difference between any two face images $\{(\vec{x}_i - \vec{x}_j)\}$, which may belong to the same individual or different individuals

$$C = \sum_{i=1}^{M} (\vec{x}_i - \vec{m})(\vec{x}_i - \vec{m})^T = \frac{1}{2M}\sum_{i=1}^{M}\sum_{j=1}^{M} (\vec{x}_i - \vec{x}_j)(\vec{x}_i - \vec{x}_j)^T$$

Principal Component Analysis (PCA)

- PCA subspace is not ideal for face recognition.
  - In PCA subspace, both $\tilde{I}$ and $\tilde{T}$ as structured signals, concentrating on the small number of principal eigenvectors. By selecting the principal components, most of the noise encoded on the large number of trailing eigenvectors is removed. But $\tilde{I}$ and $\tilde{T}$ are still coupled.



PCA subspace

PCA subspace directly computed on the set $\{\Delta\}$, which contains both intrapersonal difference and extrapersonal difference.

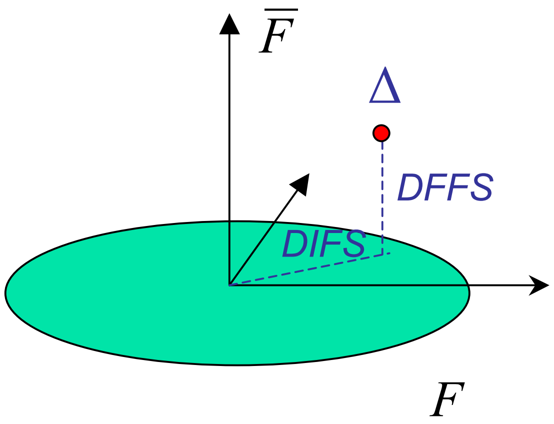
$$\Delta = \tilde{I} + \tilde{T} + \tilde{N}$$

Bayesian Face Recognition

- The similarity between two face images is based on the intrapersonal likehood $P(\Delta | \Omega_I)$
  - Apply PCA on the intrapersonal difference set $\{\Delta \mid \Delta \in \Omega_I\}$. The image space is decomposed to principal intrapersonal subspace F and its complementary subspace $\overline{F}$.

DIFS: $d_F(\Delta) = \sum_{i=1}^{K} \frac{y_i^2}{\lambda_i}$ DFFS: $\varepsilon^2(\Delta) = \|\Delta\|^2 - \sum_{i=1}^{K} y_i^2$

y_i is the projection weights of Δ on the intrapersonal eigenvectors, and λ_i is the intrapersonal eigenvalue



Bayesian Face Recognition

- $P(\Delta | \Omega_I)$ is computed as

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

ρ is the average eigenvalue in the complementary subspace $\overline{F}$

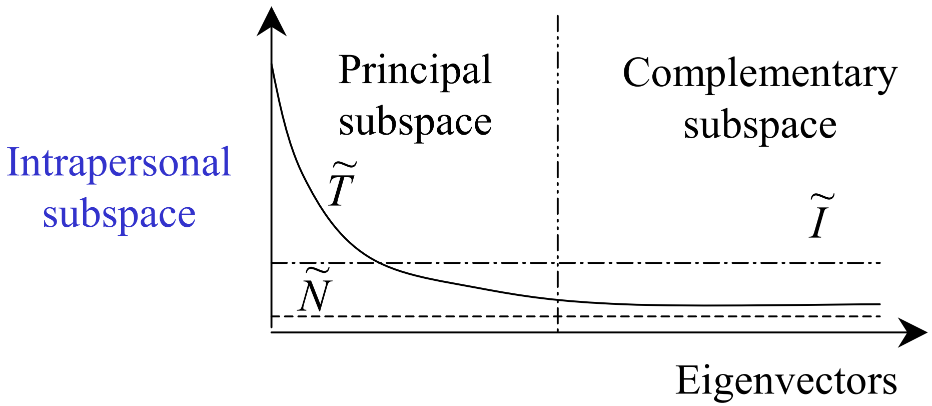
All the parameters are fixed in recognition procedure. It is equivalent to evaluating the distance

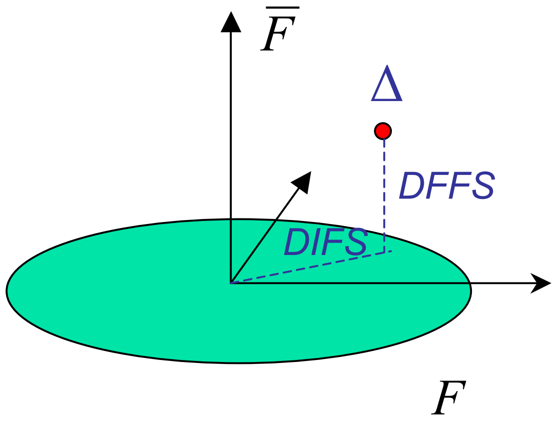
$$d(\Delta) = \sum_{i=1}^{K} \frac{y_i^2}{\lambda_i} + \varepsilon^2(\Delta)/\rho$$

Intrapersonal Subspace

- The intrapersonal subspace is computed from PCA on the intrapersonal difference set $\{\Omega_I = \tilde{T} + \tilde{N}\}$. So the axes are arranged according to the energy distribution of $\tilde{T}$.
- Most energy of the $\tilde{T}$ component will concentrate on the first few largest eigenvectors, while the $\tilde{I}$ & $\tilde{N}$ components are randomly distributed over the eigenvectors.
- The Mahalanobis distance $\sum_{i=1}^{K} y_i^2 / \lambda_i$ in the principal subspace weights the feature vectors by the inverse of eigenvalues, so it effectively reduces the $\tilde{T}$ component.
- The complementary subspace throws away most of $\tilde{T}$ the component while keep the majority of $\tilde{I}$, so $\varepsilon^2(\Delta)$ is also distinctive for recognition.

Intrapersonal Subspace





$$d(\Delta) = \sum_{i=1}^{K} \frac{y_i^2}{\lambda_i} + \varepsilon^2(\Delta)/\rho$$

Intrapersonal subspace is computed from the eigenvectors of

$$C_I = \sum_{\ell(\vec{x}_i)=\ell(\vec{x}_j)} (\vec{x}_i - \vec{x}_j)(\vec{x}_i - \vec{x}_j)^T$$

$$\Omega_I = \tilde{T} + \tilde{N}$$

Linear Discriminant Analysis

- LDA seeks for the subspace best discriminating different classes. The projection vectors W maximize the ratio between the between-class scatter matrix S_b and within-class scatter matrix S_w
- W can be computed from the eigenvectors of $S_w^{-1}S_b$
- In face recognition, the training sample number is small (M<<N). The rank of S_w is at most M-L. So S_w, the N by N matrix may become singular.
- Usually, the dimensionality of face data is first reduced to M-C using PCA, and then apply LDA in the reduced PCA subspace.

LDA Subspace

- Theorem 2: The within-class scatter matrix is identical to the covariance C_I of intrapersonal subspace in Bayes, which characterizes the distribution of face variation for the same individuals. Using the mean face image to describe each individual class, the between class scatter matrix characterizes the variation between any two mean face images.

$$S_W = \sum_{i=1}^{L} \sum_{\vec{x}_k \in X_i} (\vec{x}_k - \vec{m}_i)(\vec{x}_k - \vec{m}_i)^T$$

$$= \frac{1}{2n} \sum_{i=1}^{L} \sum_{\vec{x}_{k_1}, \vec{x}_{k_2} \in X_i} \left(\vec{x}_{k_1} - \vec{x}_{k_2}\right)\left(\vec{x}_{k_1} - \vec{x}_{k_2}\right)^T$$

$$= C_I$$

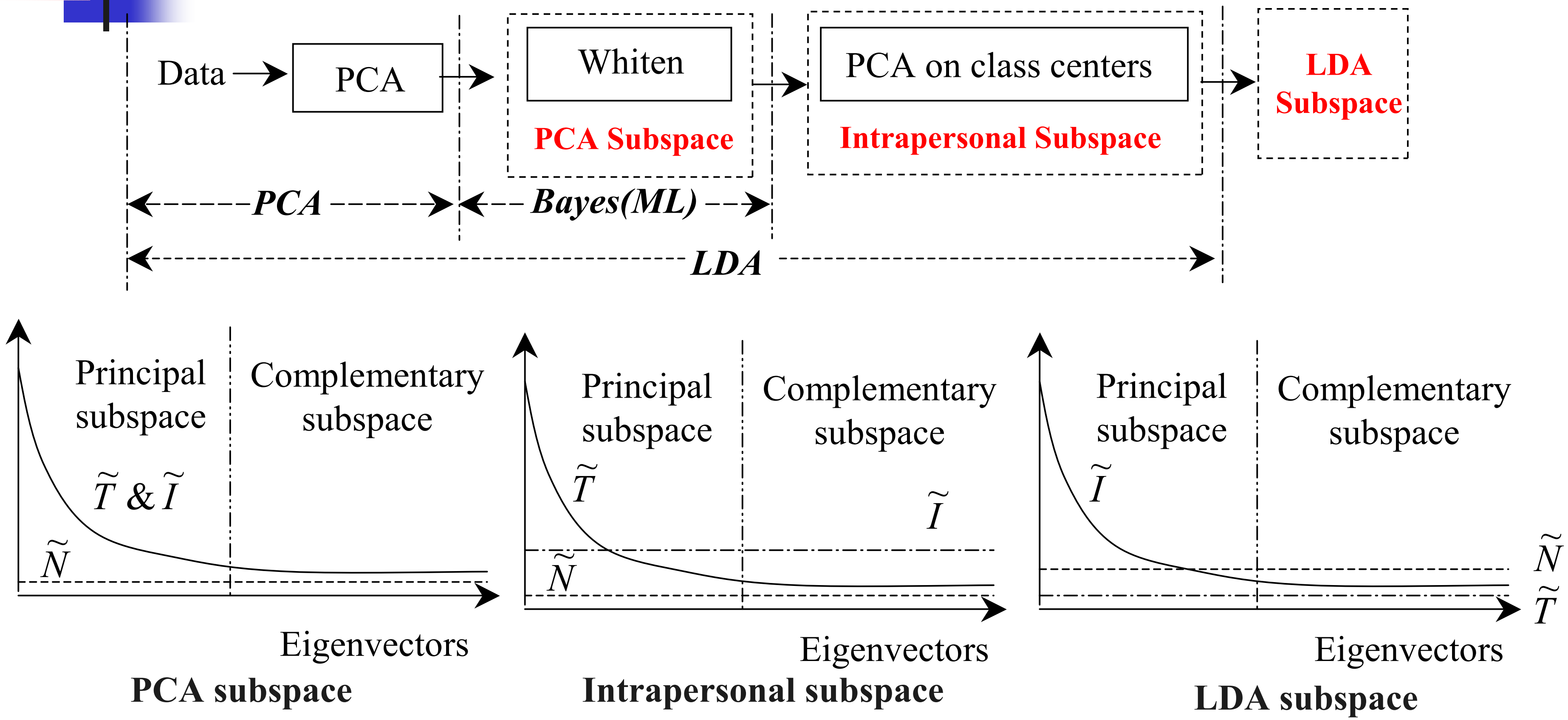
$$S_b = \sum_{i=1}^{L} (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T$$

$$= \frac{1}{2M} \sum_{i=1}^{L} \sum_{j=1}^{L} \left(\vec{m}_i - \vec{m}_j\right)\left(\vec{m}_i - \vec{m}_j\right)^T$$

LDA Subspace

- Computing LDA subspace can be divided into three steps. PCA and Bayes can be viewed as the intermediate steps of LDA.
- PCA subspace significantly reduces the noise $\tilde{N}$ and data dimension.
- Compute the intrapersonal subspace from the within-class matrix and whiten the projection data by dividing intrapersonal eigenvalues, such that the transformation difference $\tilde{T}$ is significantly reduced.
- PCA is again applied on the whitened class centers. It further reduces the noise and concentrates the energy of intrinsic difference $\tilde{I}$ onto a small number of features.

LDA Subspace



Energy distribution of the three components $\tilde{I}$, $\tilde{T}$ and $\tilde{N}$ on eigenvectors in the PCA subspace, the intrapersonal subspace, and the LDA subspace.

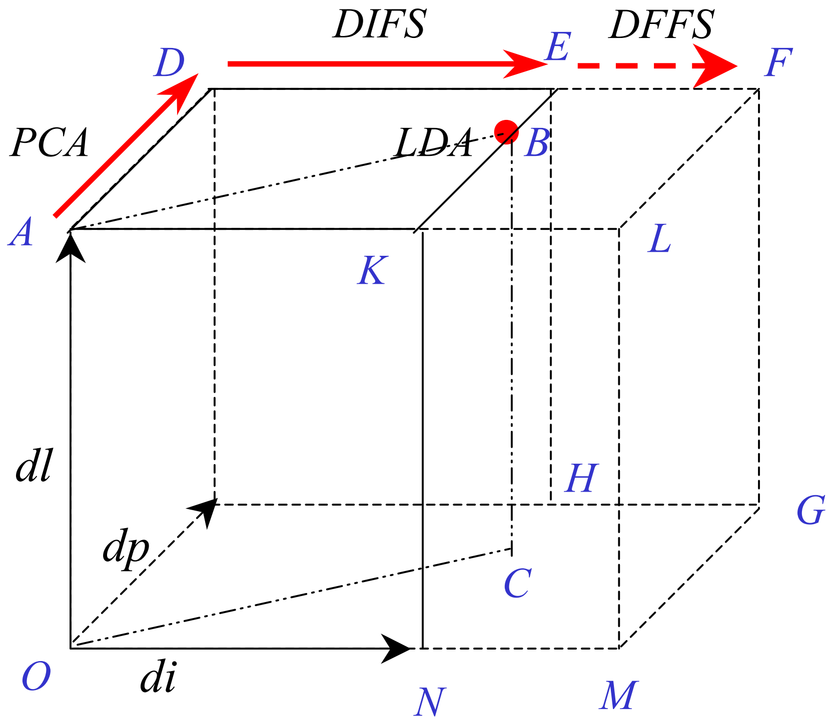
Compare Different Subspaces

Behavior of the subspaces on characterizing the face difference

Algorithm	Subspace	Decompose face image difference	
		Principal subspace	Complementary subspace
PCA	PCA subspace	$\tilde{T}+\tilde{I}$	$\tilde{N}$
Bayes	Intrapersonal subspace	$\tilde{T}$	$\tilde{I}+\tilde{N}$
LDA	LDA subspace	$\tilde{I}$	$\tilde{T}+\tilde{N}$

- The subspace dimension of each method can affect the recognition performance.
- Conventional LDA fails to attain the best performance without significant changes in each individual step. It is directly computed from the eigenvectors of $S_w^{-1}S_b$. In fact, it fixes the PCA and intrapersonal subspace as *M-L* dimension, and LDA subspace at *L-1* dimension.

Unified Subspace Analysis



3D parameter space

dp: PCA subspace dimension

di: Intrapersonal subspace dimension

dl: LDA subspace dimension

Unified Subspace Analysis

1. Project the face data to PCA subspace and adjust the PCA dimension (*dp*) to reduce the noise.
2. Apply Bayesian analysis in the PCA subspace and adjust the dimension (*di*) of intrapersonal subspace. The PCA subspace and intrapersonal subspace may be computed from an enlarged training set containing the extra samples not in the classes to be recognized.
3. Compute the class centers of the L individuals in the gallery, and project them to the intrapersonal subspace, whitened by the intrapersonal eigenvalues.
4. Apply PCA on the whitened *L* class centers to compute the discriminant feature vector of dimension (*dl)*

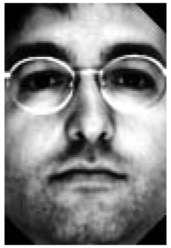
Unified Subspace Analysis

- Advantages
  - It provides a new 3D parameter space to improve the recognition performance. The optimal parameters can be found in the full 3D space, while original PCA, LDA and Bayes only occupy some local areas in this 3D parameter space
  - It adopts different training data at different training steps according to the special requirement of each step. For the intrapersonal subspace estimation (step2), we use a enlarged training set that contains individuals both inside and outside the gallery to effectively estimate $\tilde{T}$. Then for the discriminant analysis step (step4), we only use the individuals in the gallery, so that the features extracted are specifically tuned for the individuals in the gallery.

Experiments

- Data set from FERET face database
  - There are two face images (FA/FB) for each individual
  - 990 face images of 495 people for training
  - Another 700 people for testing
    - 700 face images in gallery as reference
    - 700 face images for probe

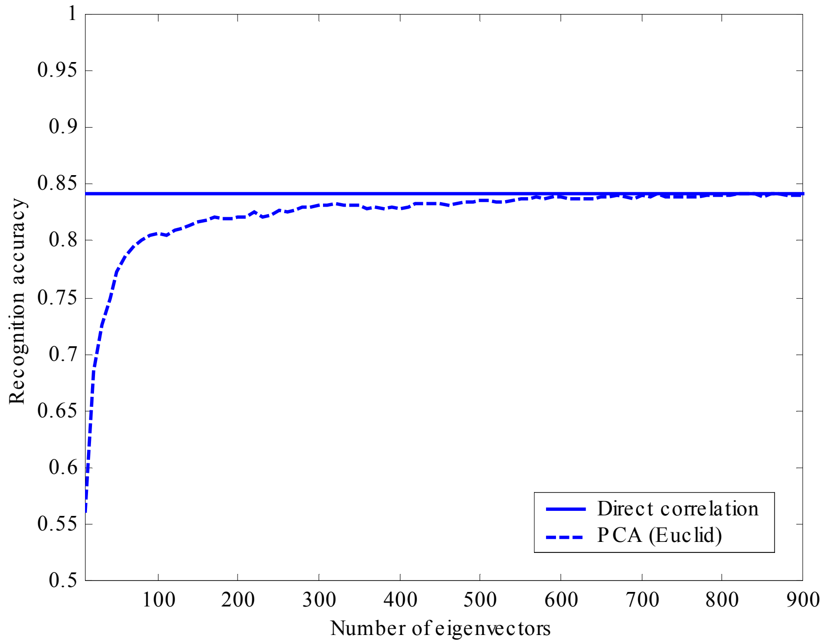
Examples of FA/FB pair



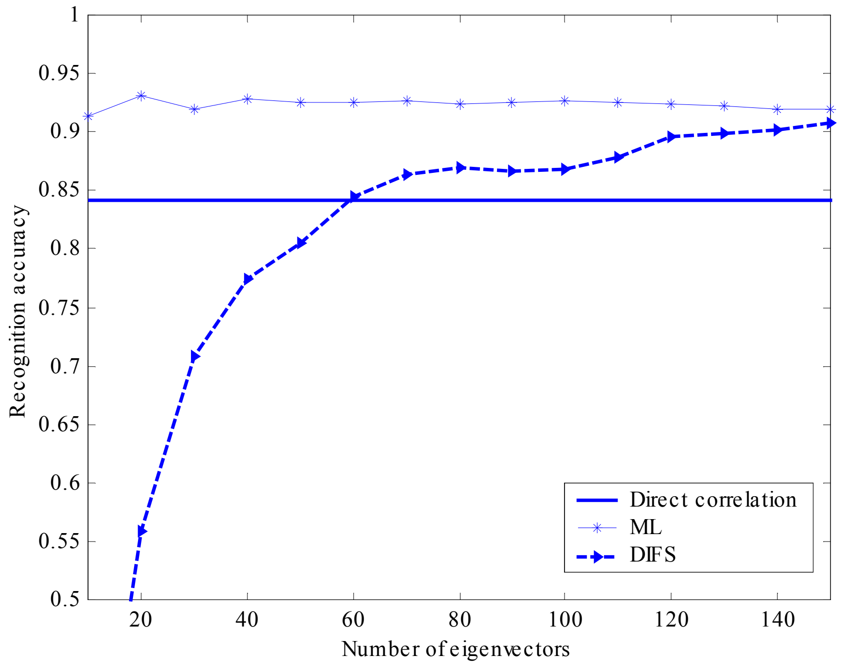
Normalized face image

Experiments

- PCA



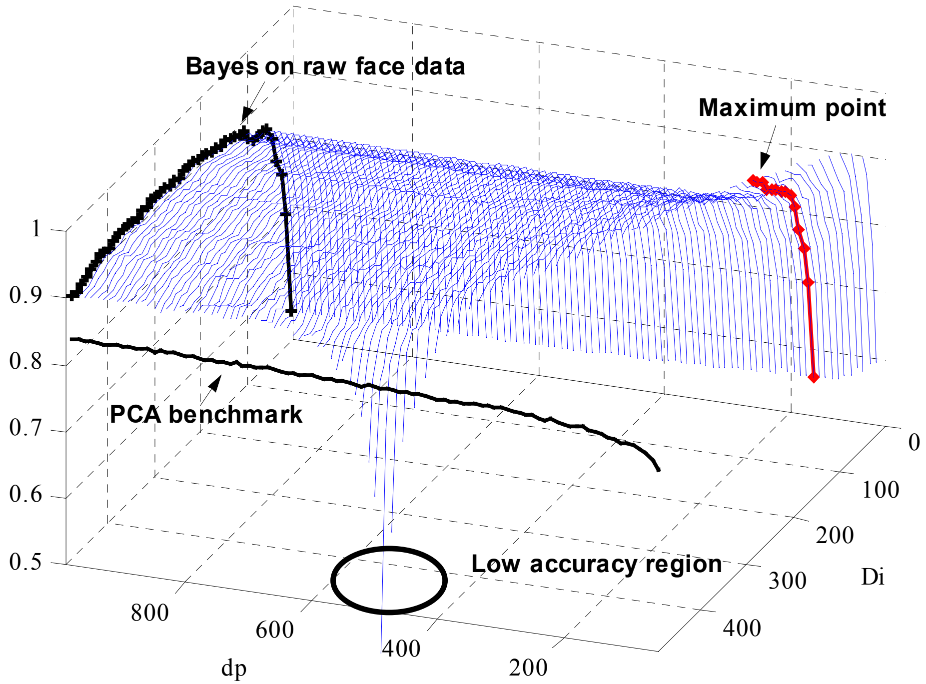
- Bayes



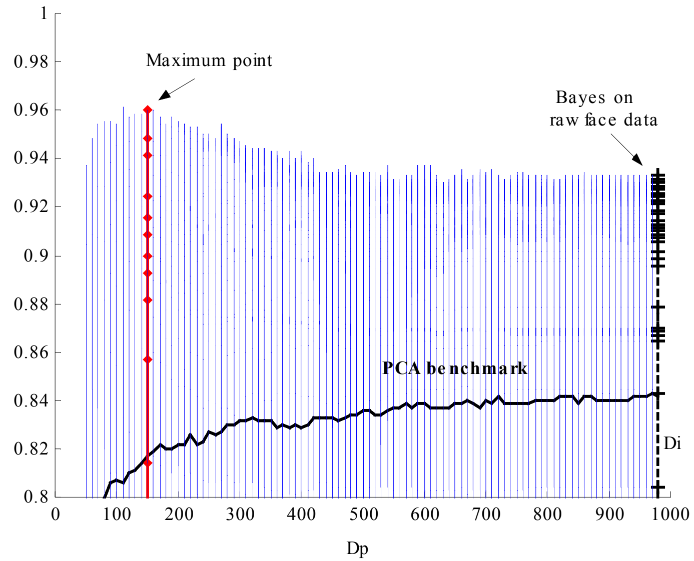
DIFS: $d(\Delta) = \sum_{i=1}^{K} \frac{y_i^2}{\lambda_i}$ DFFS: $\varepsilon^2(\Delta) = \|\Delta\|^2 - \sum_{i=1}^{K} y_i^2$ ML: DIFS+DFFS

Experiments

- Bayesian analysis in the reduced PCA space



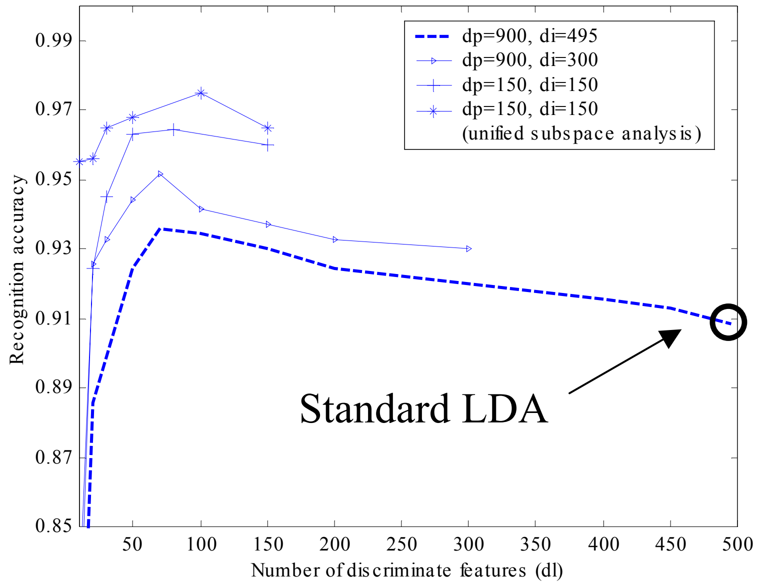
Accuracy curves for Bayesian analysis in PCA subspace



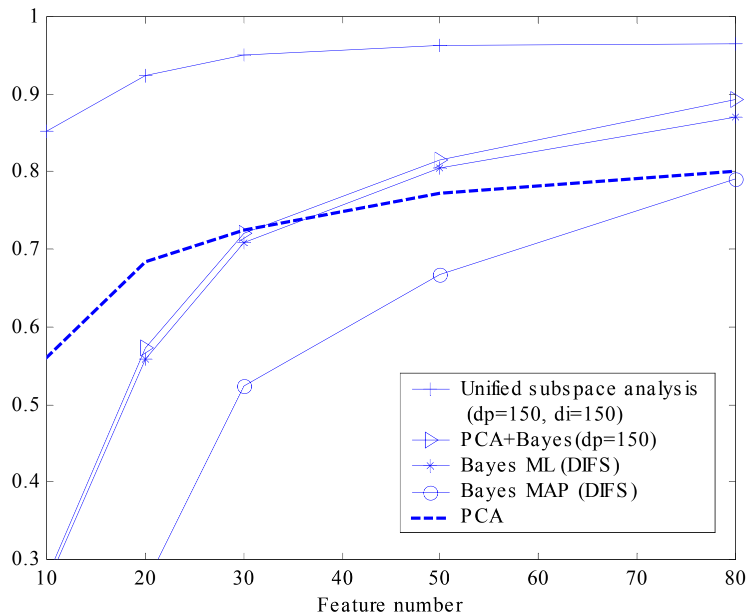
Highest accuracy of Bayes analysis in each PCA subspace

Experiments

- Extract discriminant features from intrapersonal subspace



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



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