Linear Discriminant Analysis (LDA)

• Reading Assignments

S. Gong et al., *Dynamic Vision: From Images to Face Recognition*, Imperial College Press, 2001 (pp. 173-175 and Appendix C: Mathematical Details, hard copy).

A. Webb *Statistical Pattern Recognition*, Arnold, 1999 (pp. 112-116, hard copy).

R. Duda et al., *Pattern Classification*, John Wiley, 2001 (pp. 117-124, hard copy).

• Case Studies

D. Swets and J. Weng, "Using Discriminant Eigenfeatuers for Image Retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligenve*, vol. 18, no. 8, pp. 831-836, 1996 (on-line).

A. Martinez and A. Kak, "PCA versus LDA", *IEEE Transactions on Pattern Analy sis and Machine Intelligenve*, vol. 23, no. 2, pp. 228-233, 2001, (on-line)

P. Belhumeur et al., "Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Transactions on Pattern Analysis and Machine Intelligenve*, vol. 19, no. 7, pp. 711-720, 1997 (on-line)

Linear Discriminant Analysis (LDA)

• Multiple classes and PCA

- Suppose there are C classes in the training data.

- PCA is based on the sample covariance which characterizes the scatter of the entire data set, *irrespective of class-membership*.

- The projection axes chosen by PCA might not provide good discrimination power.

• What is the goal of LDA?

- The objective of LDA is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible.

- It seeks to find directions along which the classes are best separated.

Figures/notesP2.ps

- It does so by taking into consideration the scatter *within-classes* but also the scatter *between-classes*.

- It is also more capable of distinguishing image variation due to identity from variation due to other sources such as illumination and expression.

• Methodology

- Suppose there are C classes
- Let μ_i be the mean vector of class i, $i = 1, 2, \ldots, C$
- Let M_i be the number of samples within class i, $i = 1, 2, \ldots, C$,
- Let $M = \sum_{i=0}^{C} M_i$ be the total number of samples. and

<u>Within-class scatter matrix:</u>

$$S_w = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (y_j - \mu_i)(y_j - \mu_i)^T$$

<u>Between-class scatter matrix:</u>

$$S_b = \sum_{i=1}^{C} (\mu_i - \mu)(\mu_i - \mu)^T$$

$$\mu = 1/C \sum_{i=1}^{C} \mu_i \quad \text{(mean of entire data set)}$$

- LDA computes a transformation that maximizes the between-class scatter while minimizing the within-class scatter:

$$\text{maximize } \frac{det(S_b)}{det(S_w)}$$

- Such a tranformation should retain class separability while reducing the variation due to sources other than identity (e.g., illumination).

• Linear tranformation implied by LDA

- The linear tranformation is given by a matrix U whose columns are the eigenvectors of $S_w^{-1} S_b$ (called *Fisherfaces*).

$$\begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_K \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ \cdots \\ u_K^T \end{bmatrix} (x - \mu) = U^T (x - \mu)$$

- The eigenvectors are solutions of the *generalized eigenvector problem*:

$$S_B u_k = \lambda_k S_w u_k$$

- There are at most $C - 1$ non-zero generalized eigenvectors (i.e.,$K < C$)

• Does S_w^{-1} always exist?

- If S_w is non-singular, we can obtain a conventional eigenvalue problem by writing:

$$S_w^{-1} S_B u_k = \lambda_k u_k$$

- In practice, S_w is often singular since the data are image vectors with large dimensionality while the size of the data set is much smaller ($M << N$)

To alleviate this problem, we can perform two projections:

(1) PCA is first applied to the data set to reduce its dimensionality.

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_N \end{bmatrix} --> PCA --> \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_K \end{bmatrix}$$

(2) LDA is then applied to further reduce the dimensionality of $C-1$.

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_K \end{bmatrix} --> LDA --> \begin{bmatrix} z_1 \\ z_2 \\ \cdots \\ z_{C-1} \end{bmatrix}$$

Figures/Fig2.ps