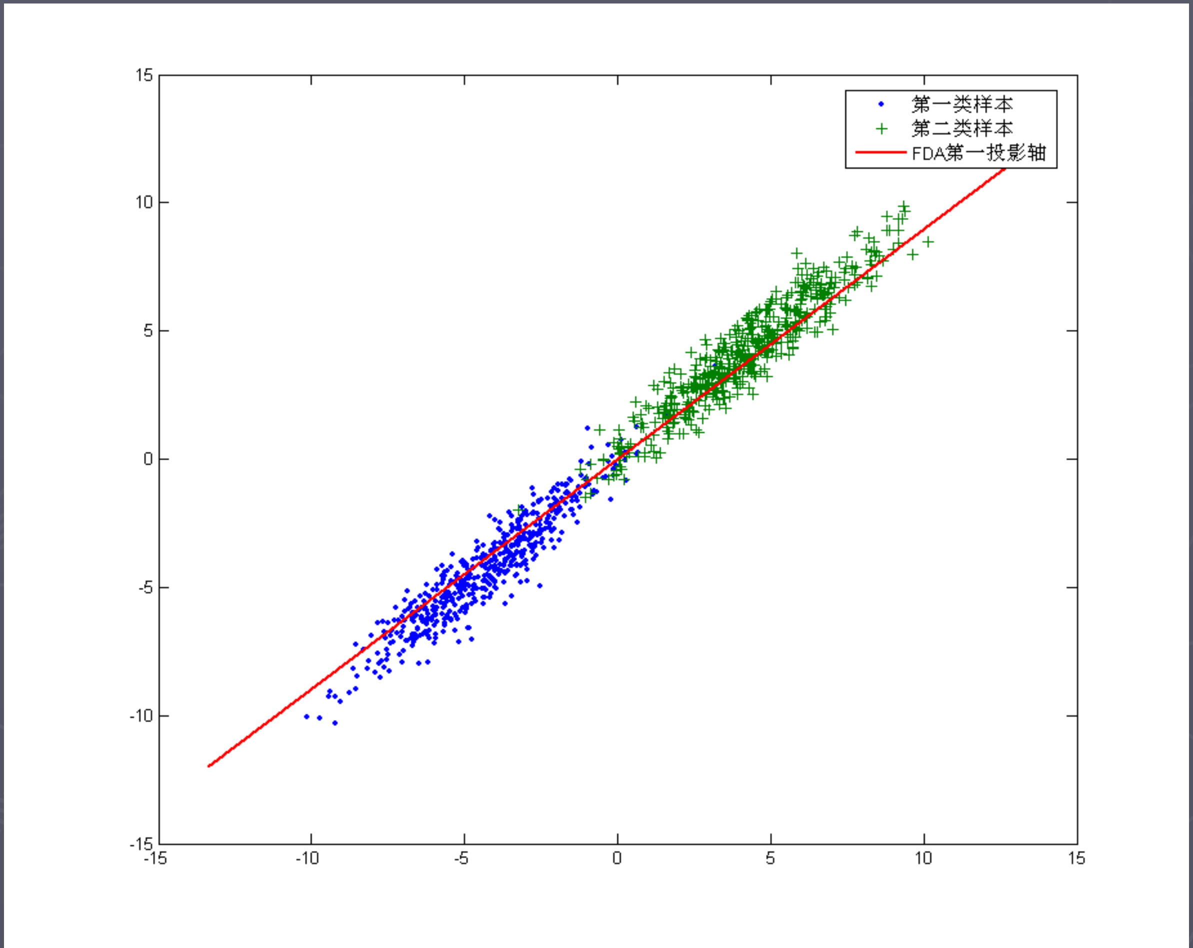
Impact on Ordination History

- by 1970 PCA was the ordination method of choice for community data
- simulation studies by Swan (1970) & Austin & Noy-Meir (1971) demonstrated the horseshoe effect and showed that the linear assumption of PCA was not compatible with the nonlinear structure of community data
- stimulated the quest for more appropriate ordination methods.

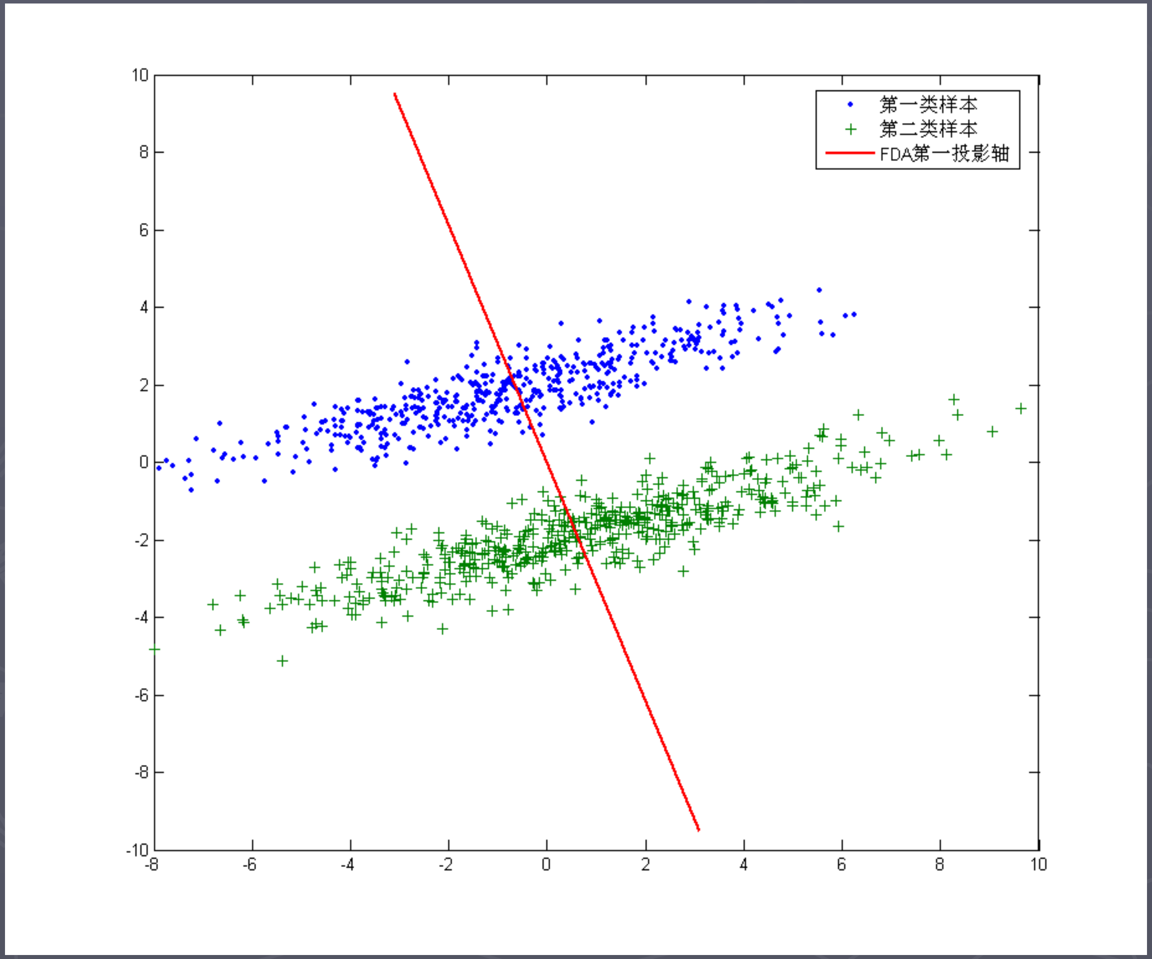
Linear discriminant analysis(LDA,FDA)

- Difference between PCA AND LDA
-
- PCA is a unsupervised method and LDA is a supervised method.
- LDA seeks the axis the projections on which of samples have the maximum between-class distance and within-class distance.

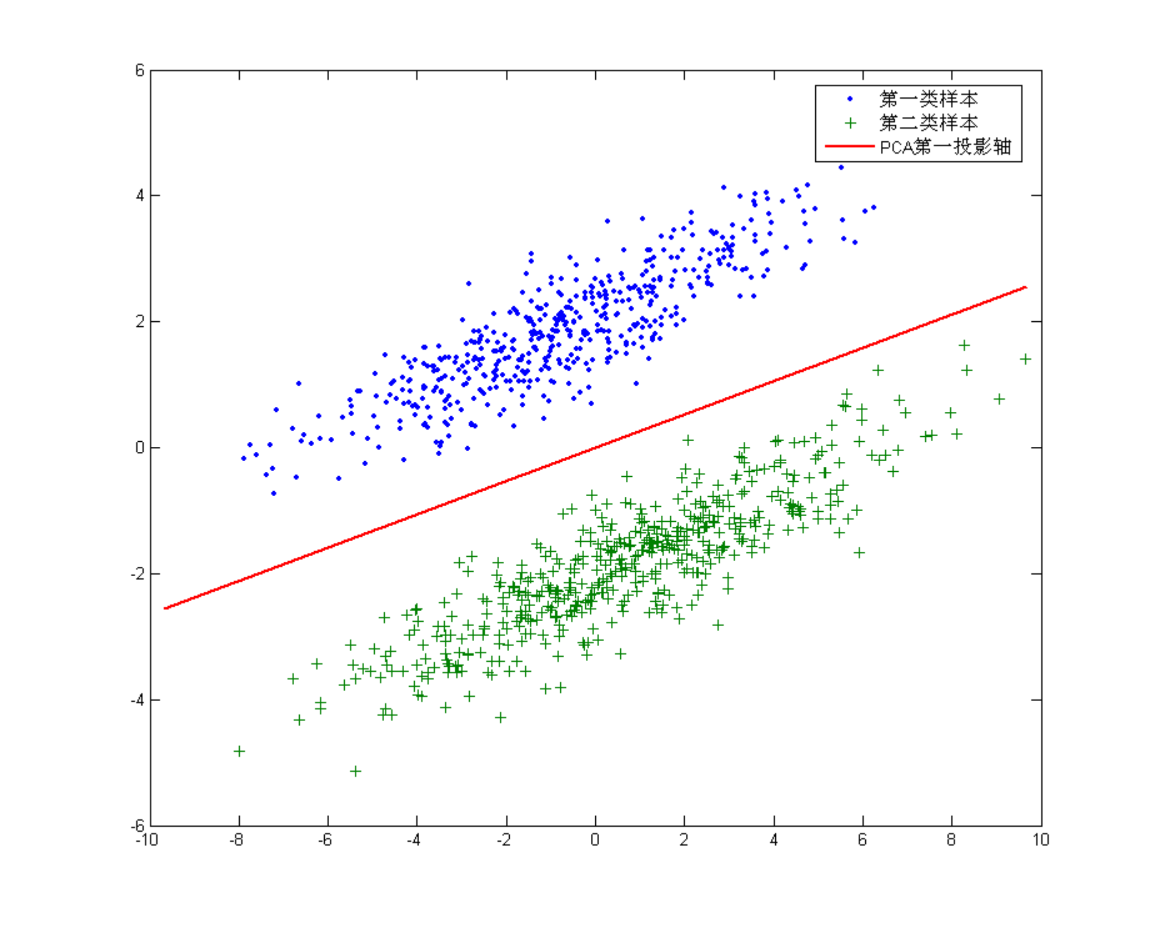
Linear discriminant analysis(LDA,FDA)



Projection axis of LDA

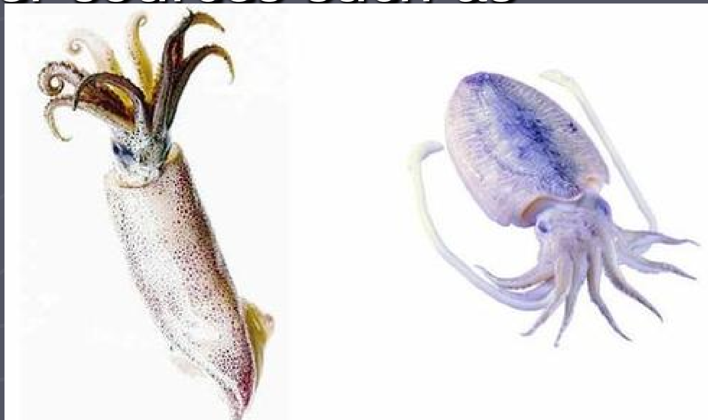


However, PCA performs badly in this case



Linear Discriminant Analysis (LDA)

- What is the goal of LDA?
  - Perform dimensionality reduction **"while preserving as much of the class discriminatory information as possible"**.
  - Seeks to find directions along which the classes are best separated.
  - Takes into consideration the scatter <u>*within-classes*</u> but also the scatter <u>*between-classes*</u>.
  - More capable of distinguishing image variation due to <u>identity</u> from variation due to other sources such as <u>illumination</u> and <u>expression</u>.



Major difference

- PCA: Class label is not considered
- LDA: Class label is fully exploited

- Training sample → Label

Linear Discriminant Analysis (LDA)

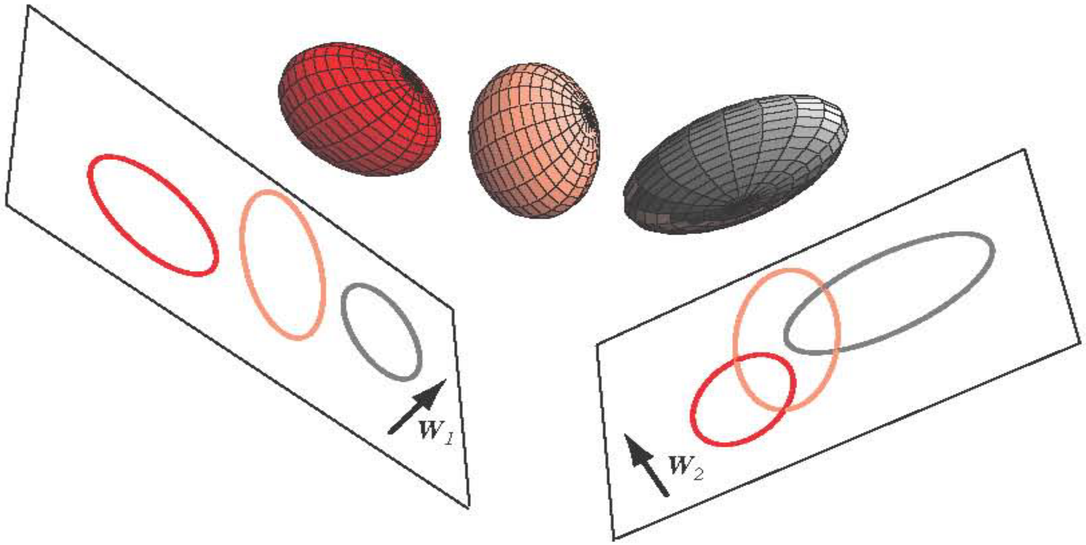


FIGURE 3.6. Three three-dimensional distributions are projected onto two-dimensional subspaces, described by a normal vectors $\mathbf{W}_1$ and $\mathbf{W}_2$. Informally, multiple discriminant methods seek the optimum such subspace, that is, the one with the greatest separation of the projected distributions for a given total within-scatter matrix, here as assoc
$\mathbf{W}_1$. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Clas*
Copyright © 2001 by John Wiley & Sons, Inc.

How does LDA differ

- Remember the label of each class and try to make every class be different from the others

Linear Discriminant Analysis (LDA)

- Notation

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

Within-class scatter matrix:

$$S_w = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (x_{ij} - \boldsymbol{\mu}_i)(x_{ij} - \boldsymbol{\mu}_i)^T$$

Between-class scatter matrix:

(each sub-matrix has rank 1 or less, i.e., outer product of two vectors)

(S_b has at most rank C-1) $$S_b = \sum_{i=1}^{C} (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$$

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

Linear Discriminant Analysis (LDA)

- Methodology

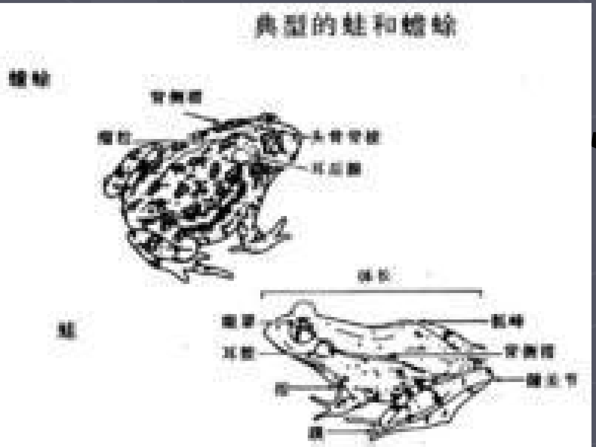
projection matrix

$$\mathbf{y} = U^T(\mathbf{x} - \mu)$$

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

$$\max \frac{|U^T S_b U|}{|U^T S_w U|} = \max \frac{|\tilde{S}_b|}{|\tilde{S}_w|}$$

products of eigenvalues !



$\tilde{S}_b, \tilde{S}_w$: scatter matrices of the projected data $\mathbf{y}$

Linear Discriminant Analysis (LDA)

- Linear transformation implied by LDA
  - The LDA solution is given by the eigenvectors of the *generalized eigenvector problem*:

$$S_B u_k = \lambda_k S_w u_k$$

  - The linear transformation is given by a matrix U whose columns are the eigenvectors of the above problem (i.e., called *Fisherfaces*).

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

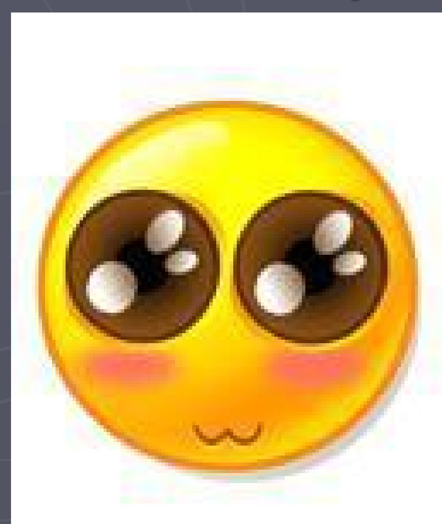
- **Important:** Since S_b has at most rank C-1, the max number of eigenvectors with non-zero eigenvalues is C-1 (i.e., max dimensionality of sub-space is C-1)

Linear Discriminant Analysis (LDA)

- 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 ($M << N$)



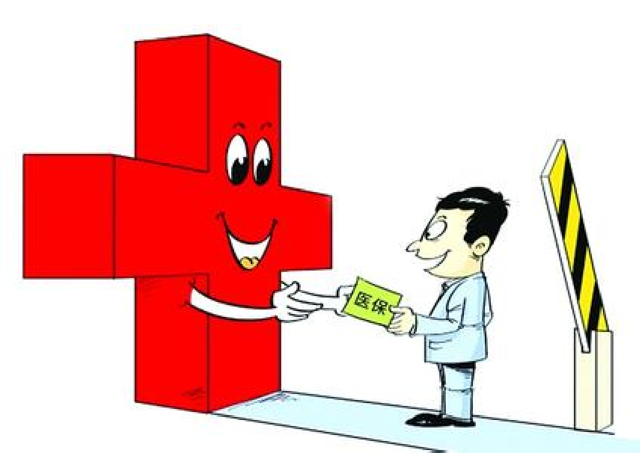
Linear Discriminant Analysis (LDA)

- Does S_w^{-1} always exist? – cont.
  - To alleviate this problem, we can use PCA first:
    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 find the most discriminative directions:

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



Linear Discriminant Analysis (LDA)

- Case Study: Using Discriminant Eigenfeatures for Image Retrieval
  - D. Swets, J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831-836, 1996.
- Content-based image retrieval
  - The application being studied here is *query-by-example* image retrieval.
  - The paper deals with the problem of *selecting a good set of image features* for content-based image retrieval.

Linear Discriminant Analysis (LDA)

- Assumptions
  - "Well-framed" images are required as input for training and query-by-example test probes.
  - Only a small variation in the size, position, and orientation of the objects in the images is allowed.

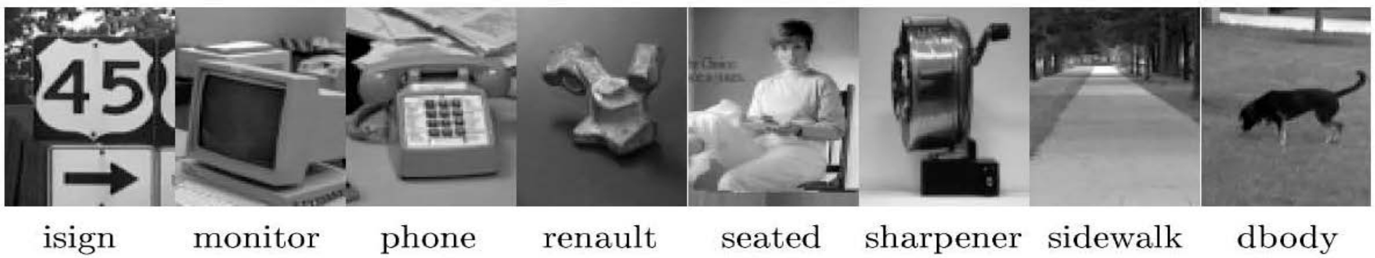


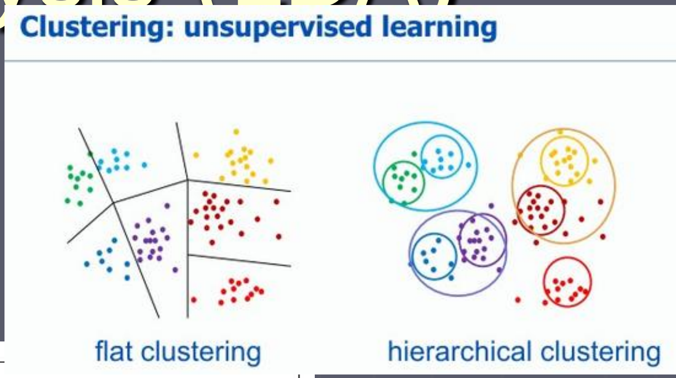
Figure 6. Representatives of training images from some of the various classes learned. Objects of interest are at the center of the fovea images. In the learning phase for this experiment, the training images were generated using manual extraction of the areas of interest.

Linear Discriminant Analysis (LDA)

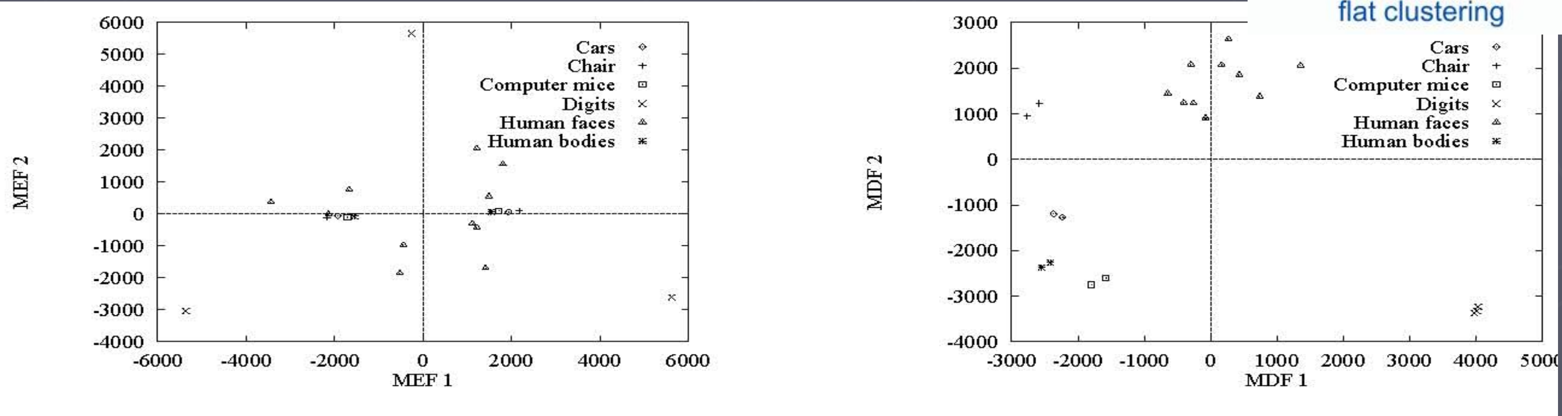
- Some terminology
  - Most Expressive Features (MEF): the features (projections) obtained using PCA

  - Most Discriminating Features (MDF): the features (projections) obtained using LDA.

Linear Discriminant Analysis (LDA)



- Clustering effect



(a) MEF space (b) MDF space

- Methodology

1) Generate the features for each image in the training set.
2) Given an query image, compute its features using the same procedure.
3) Find the ***k*** **closest neighbors** for retrieval (e.g., using Euclidean distance).

Linear Discriminant Analysis (LDA)

- Experiments and results
  - Face images
    - A set of face images was used with 2 expressions, 3 lighting conditions.
    - Testing was performed using a disjoint set of images:
      - One image, randomly chosen, from each individual.

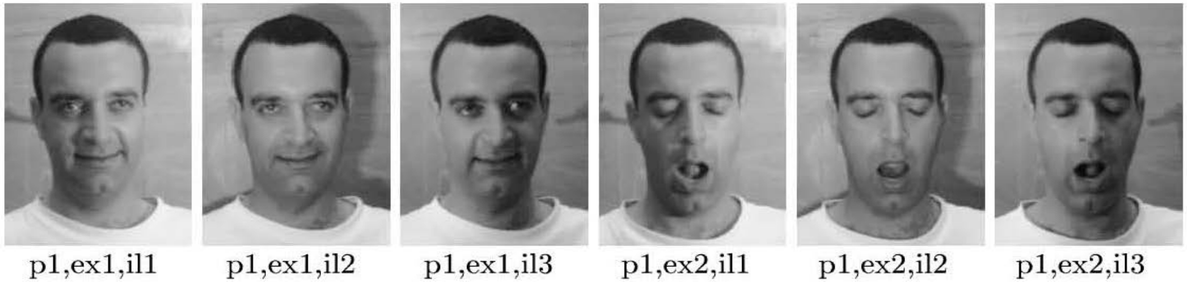


Figure 5. A sample of the Weizmann Institute face data. The frontal images for an individual contain two expressions; for each expression, a set of three images with differing lighting conditions forms the set of images available for an individual.

Linear Discriminant Analysis (LDA)

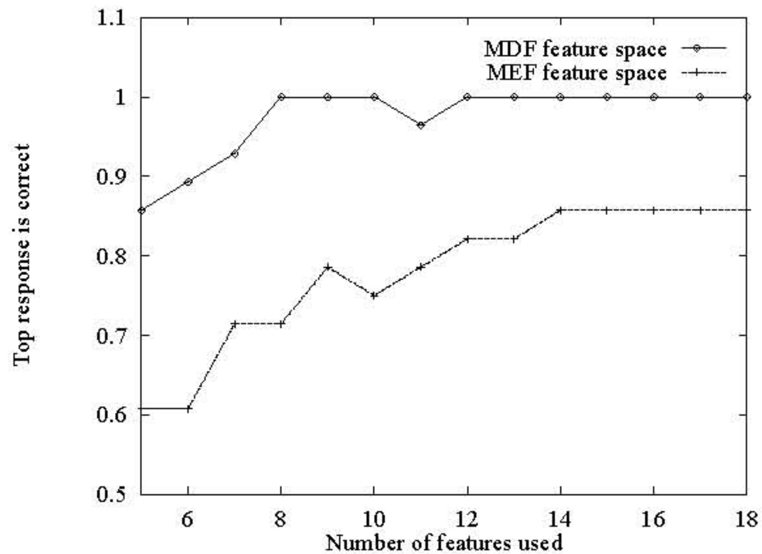


Figure 4. The performance of the ~~system for different numbers of MEF and MDF features, respectively. The number of features from the subspace used was varied to show how the MDF subspace~~ outperforms the MEF subspace. 95% of the variance for the MDF subspace was attained when 15 features were used; 95% of variance for the MEF subspace did not occur until 37 features were used. Using 95% of the MEF variance resulted in an 89% recognition rate, and that rate was not improved using more features.

Linear Discriminant Analysis (LDA)

- Examples of correct search probes

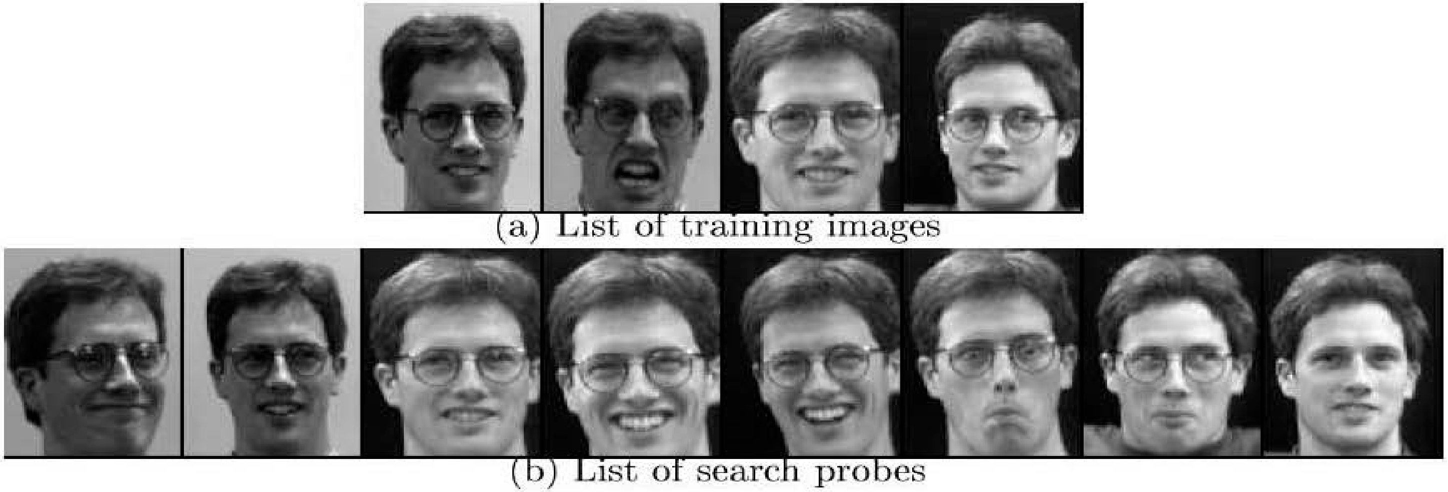


Figure 8. Example of how well within-class variation is handled. The system correctly retrieved images from the class defined by the training samples for each of the search probes.

Linear Discriminant Analysis (LDA)

- Example of a failed search probe

(a) Search probe (b) Training images

Figure 7. Example of a failed search probe. The retrieval failed to select the appropriate class due to a lack of 3D rotation in the set of training images.

Linear Discriminant Analysis (LDA)

► Case Study: PCA versus LDA

- A. Martinez, A. Kak, "PCA versus LDA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, 2001.

► Is LDA always better than PCA?

- There has been a tendency in the computer vision community to prefer LDA over PCA.
- This is mainly because LDA deals directly with discrimination between classes while PCA does not pay attention to the underlying class structure.
- Main results of this study:

(1) When the training set is small, PCA can outperform LDA.

(2) When the number of samples is large and representative for each class, LDA often outperforms PCA.

Linear Discriminant Analysis (LDA)

- Is LDA always better than PCA? – cont.

Fig. 3. Images of one subject in the AR face database. The images (a)-(m) were taken during one session and the images (n)-(z) at a different session.

Linear Discriminant Analysis (LDA)

- Is LDA always better than PCA? – cont.

LDA is not always better when training set is small

Linear Discriminant Analysis (LDA)

- Is LDA always better than PCA? – cont.

LDA often outperforms PCA when training set is large

Advances

- **Jian Yang**, David Zhang, Xu Yong, and Jing-yu Yang, Two-dimensional Discriminant Transform for Face Recognition, *Pattern Recognition*, 2005, 38(7), 1125-1129.
- Y. Xu, D. Zhang, Represent and fuse bimodal biometric images at the feature level: complex-matrix-based fusion scheme, Optical Engineering， 49(3), 037002, 2010
- Y. Xu, Quaternion-Based Discriminant Analysis Method for Color Face Recognition, PLoS ONE, 7(8): e43493， 2012
- http://www.yongxu.org/lunwen.html