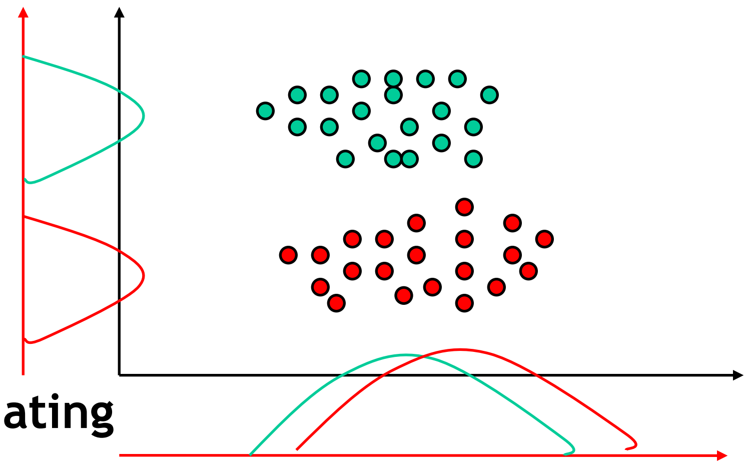
Topics of This Lecture

- **Fisher's Linear Discriminant Analysis (FLD/LDA)**
  - Derivation
  - Fisherfaces for recognition

Restrictions of PCA

- PCA minimizes projection error

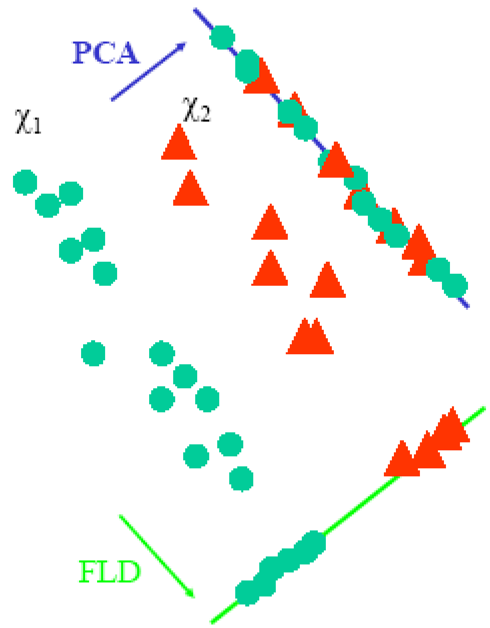


Best discriminating projection

PCA projection

- PCA is „unsupervised“, no information on classes is used
- Discriminating information might be lost

Fischer’s Linear Discriminant Analysis (FLD)



- **FLD is an enhancement to PCA**
  - **Constructs a discriminant subspace that minimizes the scatter between images of the same class and maximizes the scatter between different class images**
  - **Also sometimes called LDA…**

Mean Images

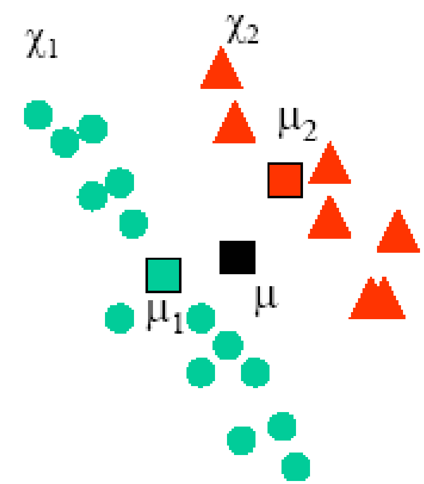
- Let X_1, X_2,…, X_k be the classes in the database and let each class X_i, i = 1,2,…,k have N_i images x_j, j=1,2,…,k.

- We compute the mean image μ_i of each class X_i as:

$$\mu_i = \frac{1}{k}\sum_{j=1}^{N_i} x_j$$

- Now, the mean image μ of all the classes in the database can be calculated as:

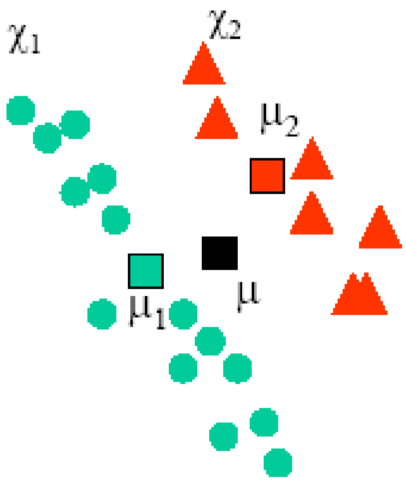
$$\mu = \frac{1}{c}\sum_{i=1}^{k} \mu_i$$



Scatter Matrices

- We calculate the within-class scatter matrix as:

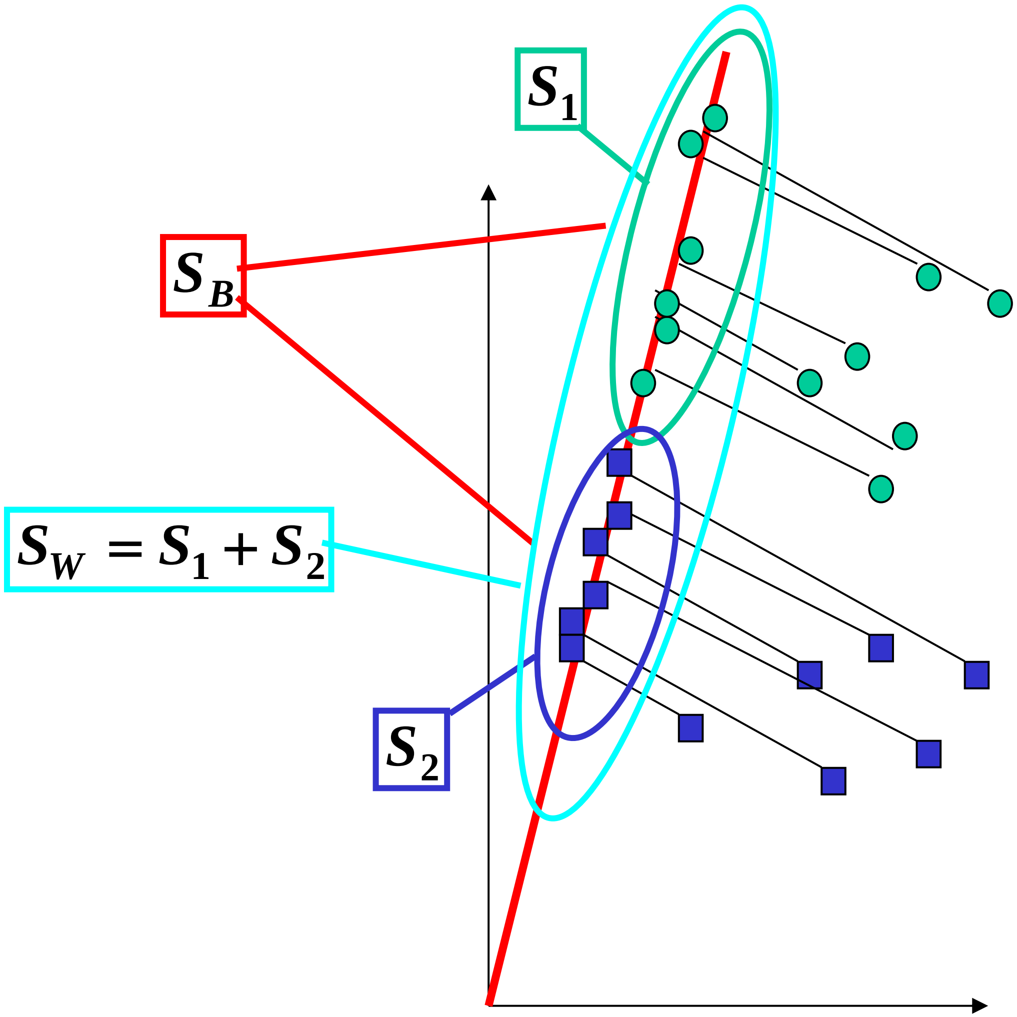
$$S_W = \sum_{i=1}^{k} \sum_{x_j \in X_i} (x_j - \mu_i)(x_j - \mu_i)^T$$



- We calculate the between-class scatter matrix as:

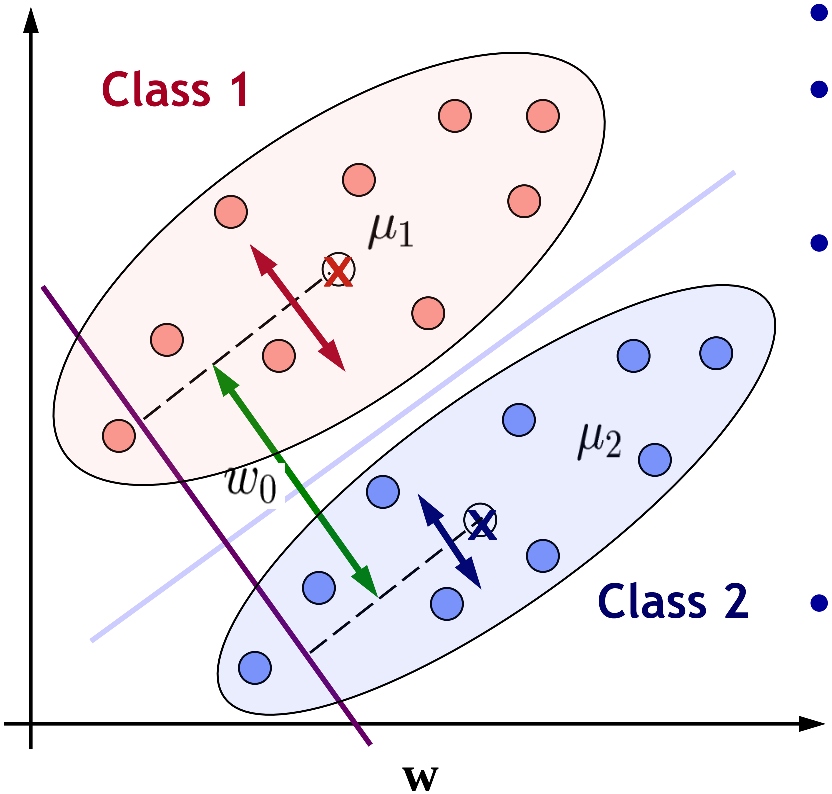
$$S_B = \sum_{i=1}^{k} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

Visualization



Good separation

Fisher's Linear Discriminant Analysis (FLD)



- Maximize distance between classes
- Minimize distance within a class
- Criterion: $J(\mathbf{w}) = \dfrac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_B\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_W\mathbf{w}}$

  $\mathbf{S}_B$ … between-class scatter matrix
  $\mathbf{S}_W$ … within-class scatter matrix

- In the two-class case, the optimal solution for $\mathbf{w}$ can be obtained as:

$$\mathbf{w} \propto \mathbf{S}_W^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$$

- Classification function:

$$y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 \underset{\text{Class 2}}{\overset{\text{Class 1}}{\gtrless}} 0$$

Multiple Discriminant Analysis

- **Generalization to K classes**

$$J(\mathbf{W}) = \frac{|\mathbf{W}^{\mathrm{T}}\mathbf{S}_B\mathbf{W}|}{|\mathbf{W}^{\mathrm{T}}\mathbf{S}_W\mathbf{W}|}$$

"Rayleigh quotient"

  - where

$$\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K] \qquad \boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n = \frac{1}{N}\sum_{k=1}^{K} N_k\boldsymbol{\mu}_k$$

$$\mathbf{S}_B = \sum_{k=1}^{K} N_k(\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^{\mathrm{T}}$$

$$\mathbf{S}_W = \sum_{k=1}^{K}\sum_{n\in\mathcal{C}_k}(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

Maximizing J(W)

- Solution from generalized eigenvalue problem

$$J(\mathbf{W}) = \frac{|\mathbf{W}^{\mathrm{T}}\mathbf{S}_B\mathbf{W}|}{|\mathbf{W}^{\mathrm{T}}\mathbf{S}_W\mathbf{W}|}$$

  - The columns of the optimal W are the eigenvectors correspon-ding to the largest eigenvectors of

$$\mathbf{S}_B\mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i$$

  - Defining $\mathbf{v} = \mathbf{S}_B^{\frac{1}{2}}\mathbf{w}$, we get

$$\mathbf{S}_B^{\frac{1}{2}}\mathbf{S}_W^{-1}\mathbf{S}_B^{\frac{1}{2}}\mathbf{v} = \lambda\mathbf{v}$$

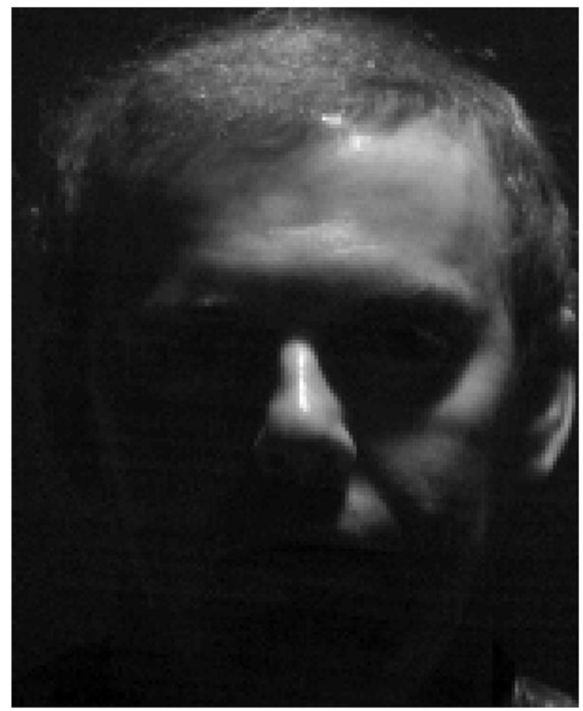
    which is a regular eigenvalue problem.

  $\Rightarrow$ Solve to get eigenvectors of $\mathbf{v}$, then from that of $\mathbf{w}$.

- For the K-class case we obtain (at most) K-1 projections.
  - (i.e. eigenvectors corresponding to non-zero eigenvalues.)

Face Recognition Difficulty: Lighting

- The same person with the same facial expression, and seen from the same viewpoint, can appear dramatically different when light sources illuminate the face from different directions.

Application: Fisherfaces

- **Idea:**
  - Using Fisher's linear discriminant to find class-specific linear projections that compensate for lighting/facial expression.
- **Singularity problem**
  - The within-class scatter is always singular for face recognition, since #training images << #pixels
  - This problem is overcome by applying PCA first

$$W_{opt}^T = W_{fld}^T U_{pca}^T$$

where

$$U_{pca} = \arg\max_U |U^T S_T U|, \qquad S_T = S_B + S_W$$

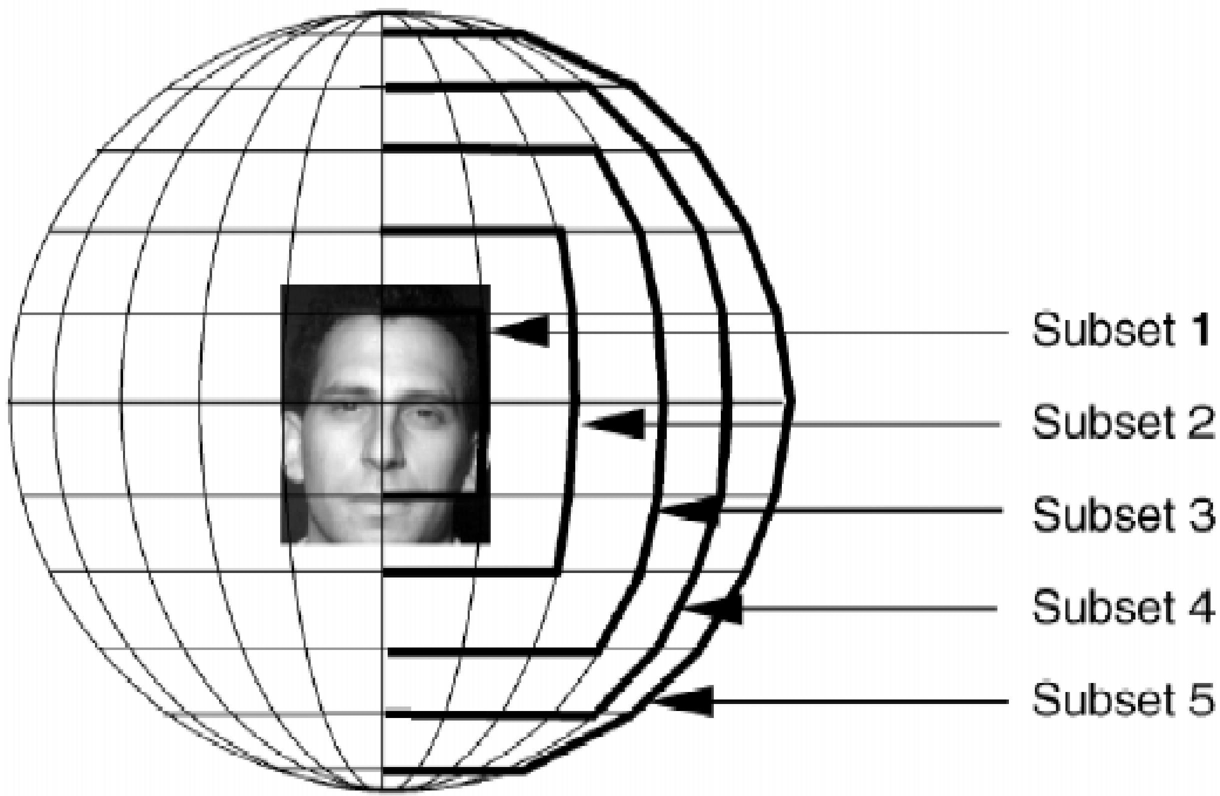
$$W_{fld} = \arg\max_W \frac{|W^T U_{pca}^T S_B U_{pca} W|}{|W^T U_{pca}^T S_W U_{pca} W|}$$

Slide credit: Peter Belhumeur

[Belhumeur et.al. 1997]

Fisherfaces: Experiments

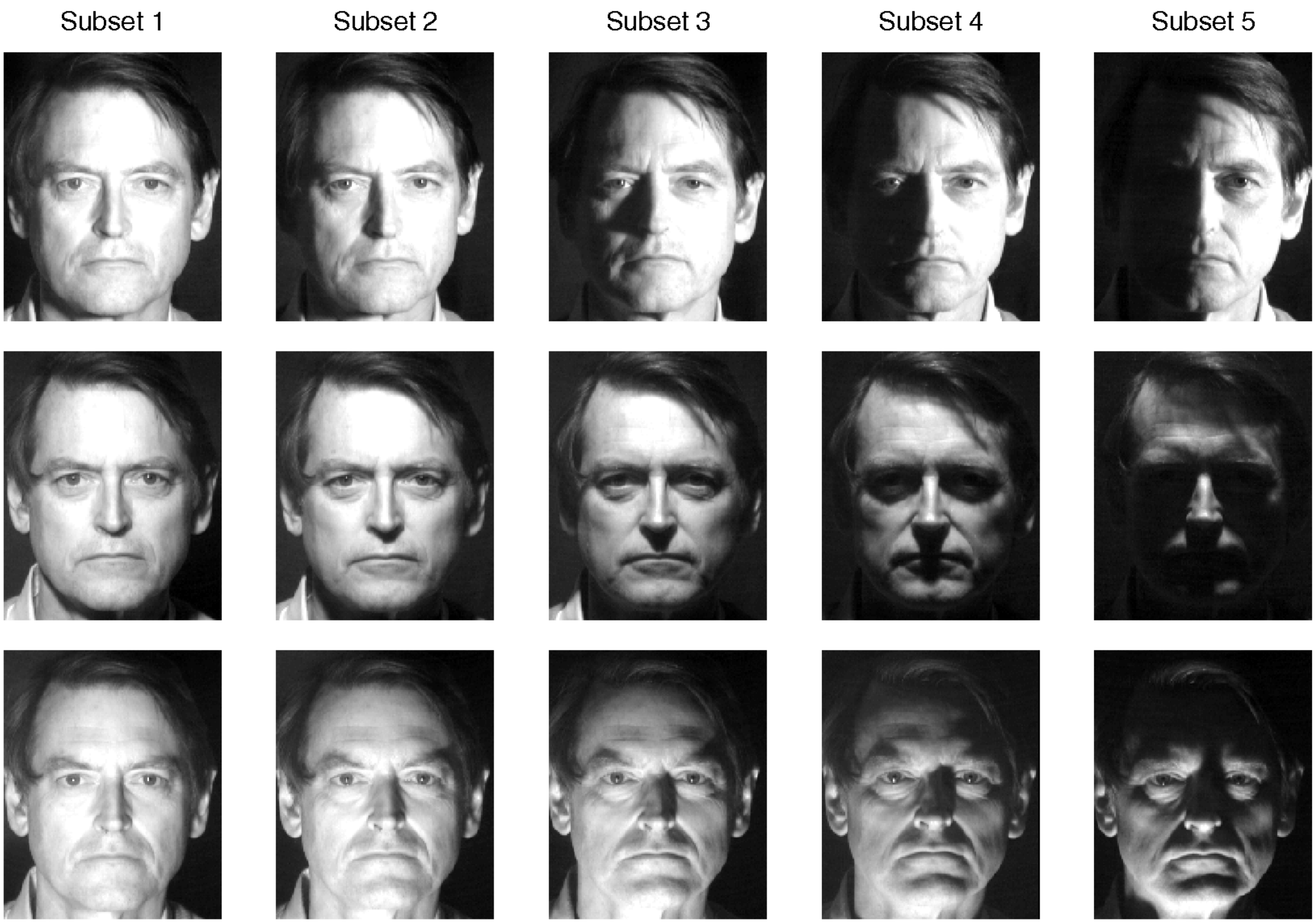
- **Variation in lighting**



[Belhumeur et.al. 1997]

RWTH AACHEN UNIVERSITY

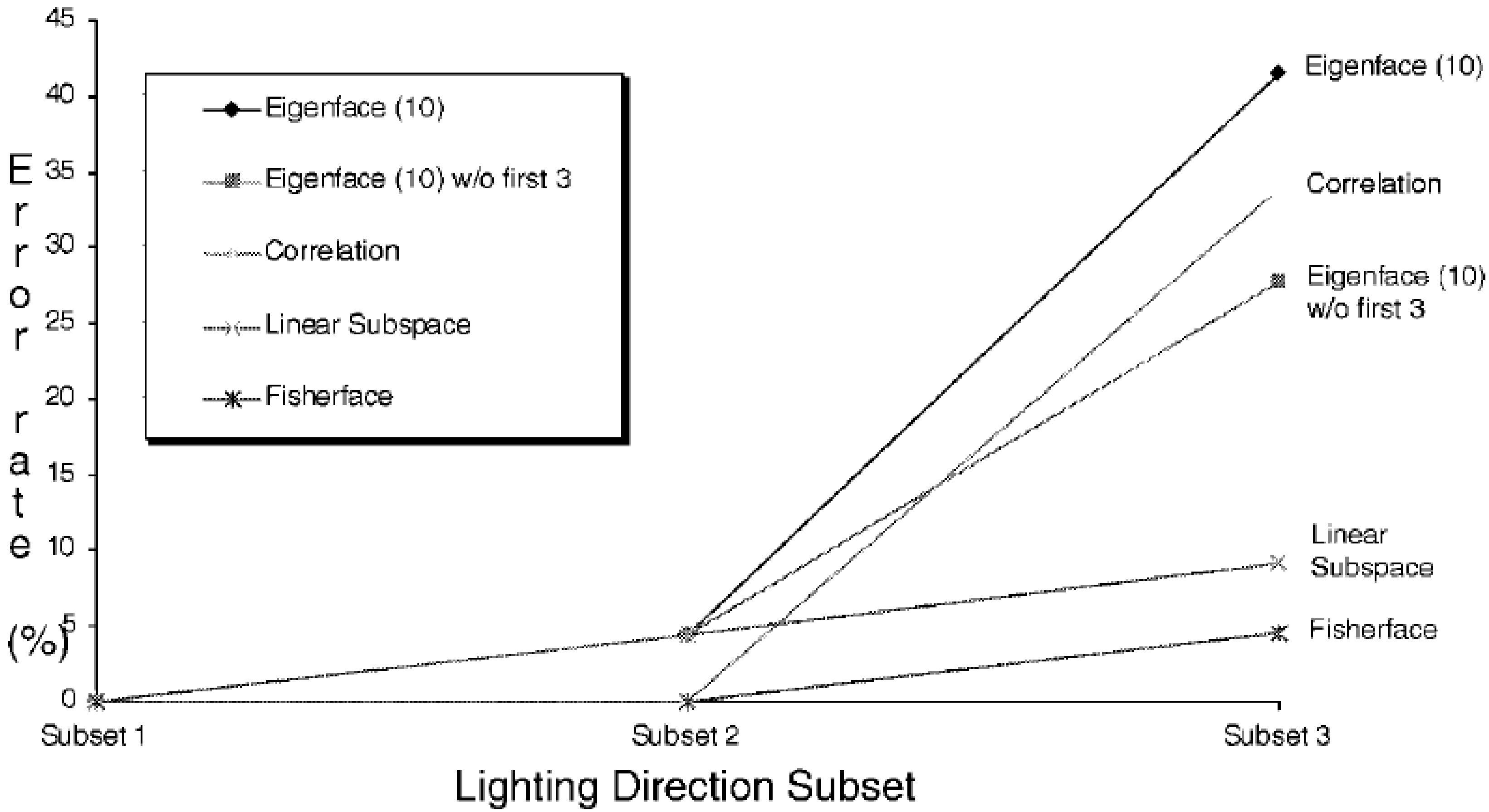
Fisherfaces: Experiments



Slide credit: Peter Belhumeur

[Belhumeur et.al. 1997]

Fisherfaces: Experimental Results



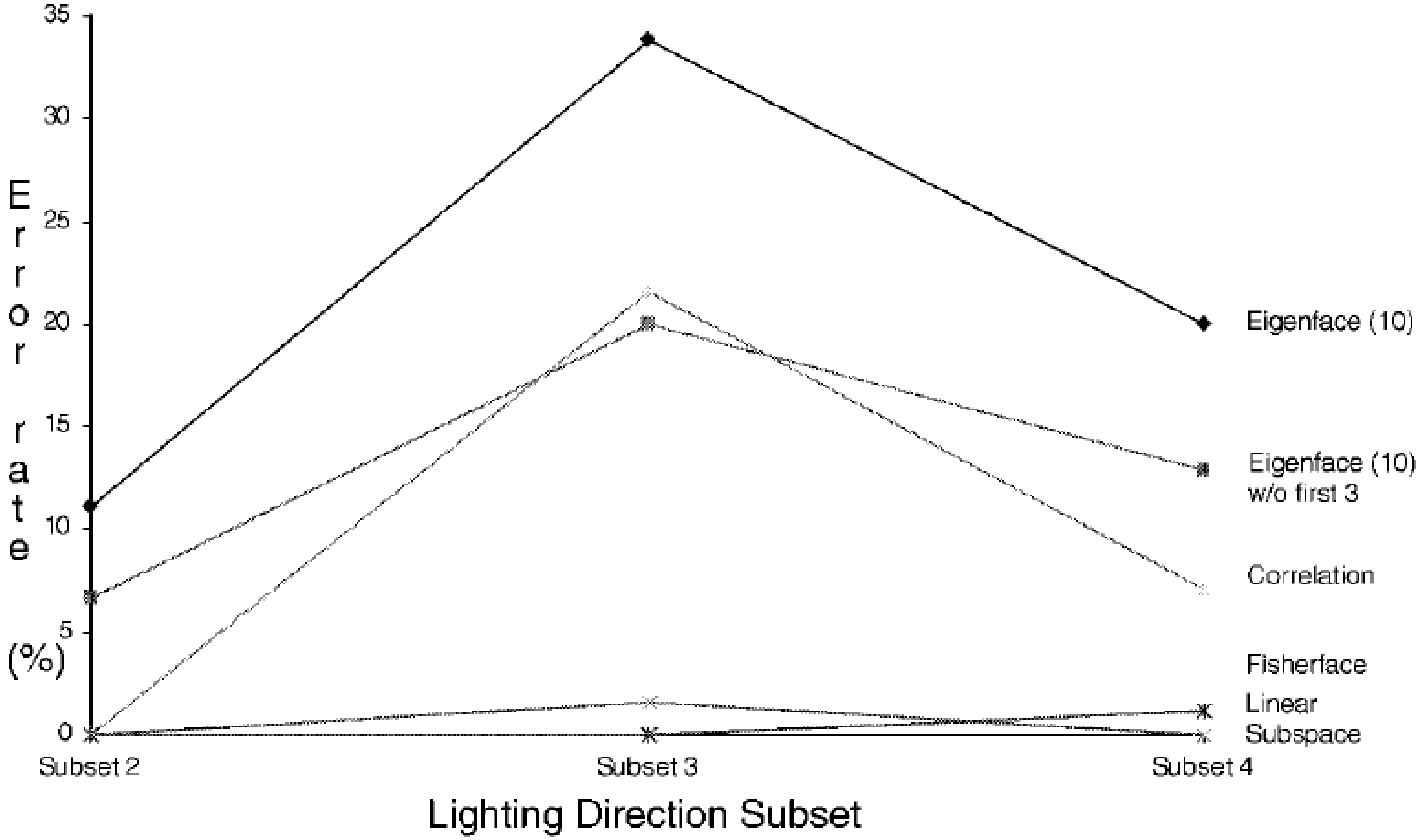
[Belhumeur et.al. 1997]

Fisherfaces: Experiments

- **Variation in facial expression, eye wear, lighting**

[Belhumeur et.al. 1997]

Fisherfaces: Experimental Results

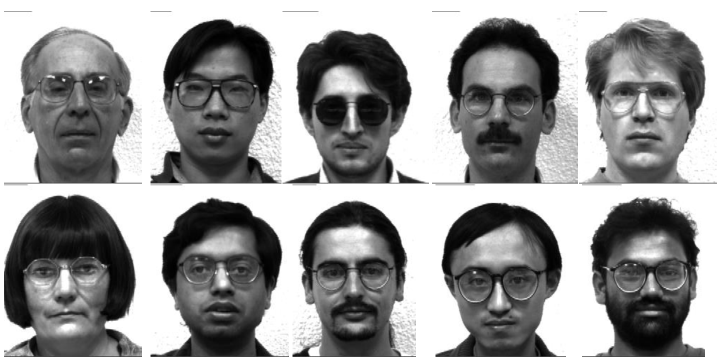


[Belhumeur et.al. 1997]

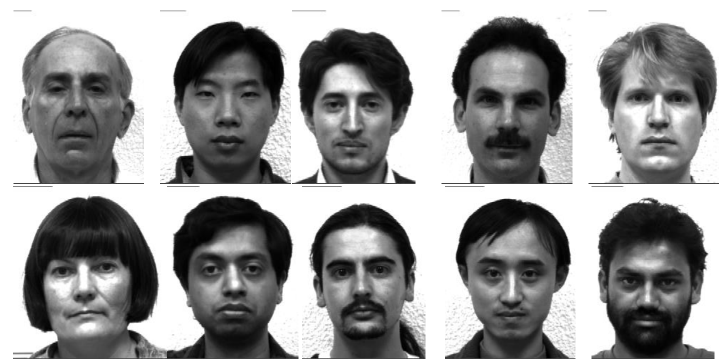
Example Application: Fisherfaces

- Visual discrimination task
  - Training data:

C_1: Subjects with glasses

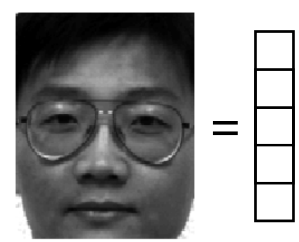


C_2: Subjects without glasses



  - Test:

– glasses?



Take each image as a vector of pixel values and apply FLD…

Image source: Yale Face Database

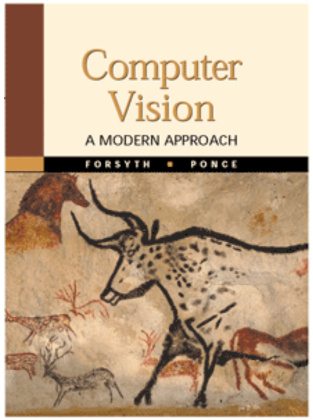
Fisherfaces: Interpretability

- **Example Fisherface for recognition “Glasses/NoGlasses“**

[Belhumeur et.al. 1997]

References and Further Reading

- Background information on PCA/FLD can be found in Chapter 22.3 of
  - D. Forsyth, J. Ponce, *Computer Vision - A Modern Approach*. Prentice Hall, 2003



- Important Paper (available on webpage)
  - P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Trans. PAMI, Vol. 19(7), 1997.