Face Recognition and Feature Subspaces

Computer Vision

CS 543 / ECE 549

University of Illinois

Derek Hoiem

Presented by Ali Farhadi

Some slides from Lana Lazebnik, Silvio Savarese, Fei-Fei Li

Object recognition

Last Class

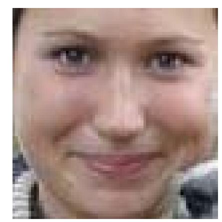
- Object instance recognition: focus on localization of miscellaneous objects

This class

- Face recognition: focus on distinguishing one face from another
- Feature subspaces: PCA and FLD
- Look at results from recent vendor test
- Look at interesting findings about human face recognition

Face detection and recognition

Detection

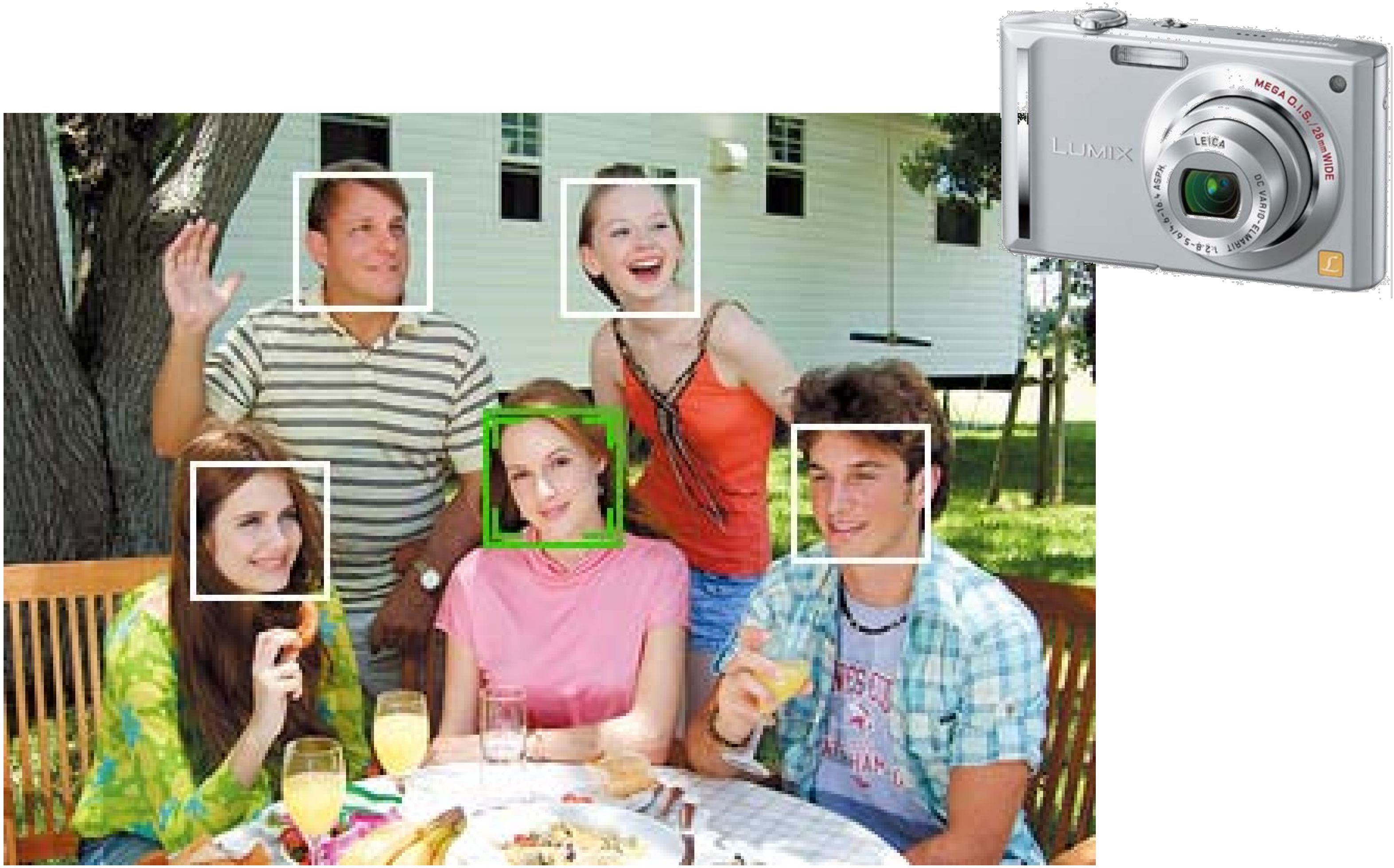


Recognition

"Sally"

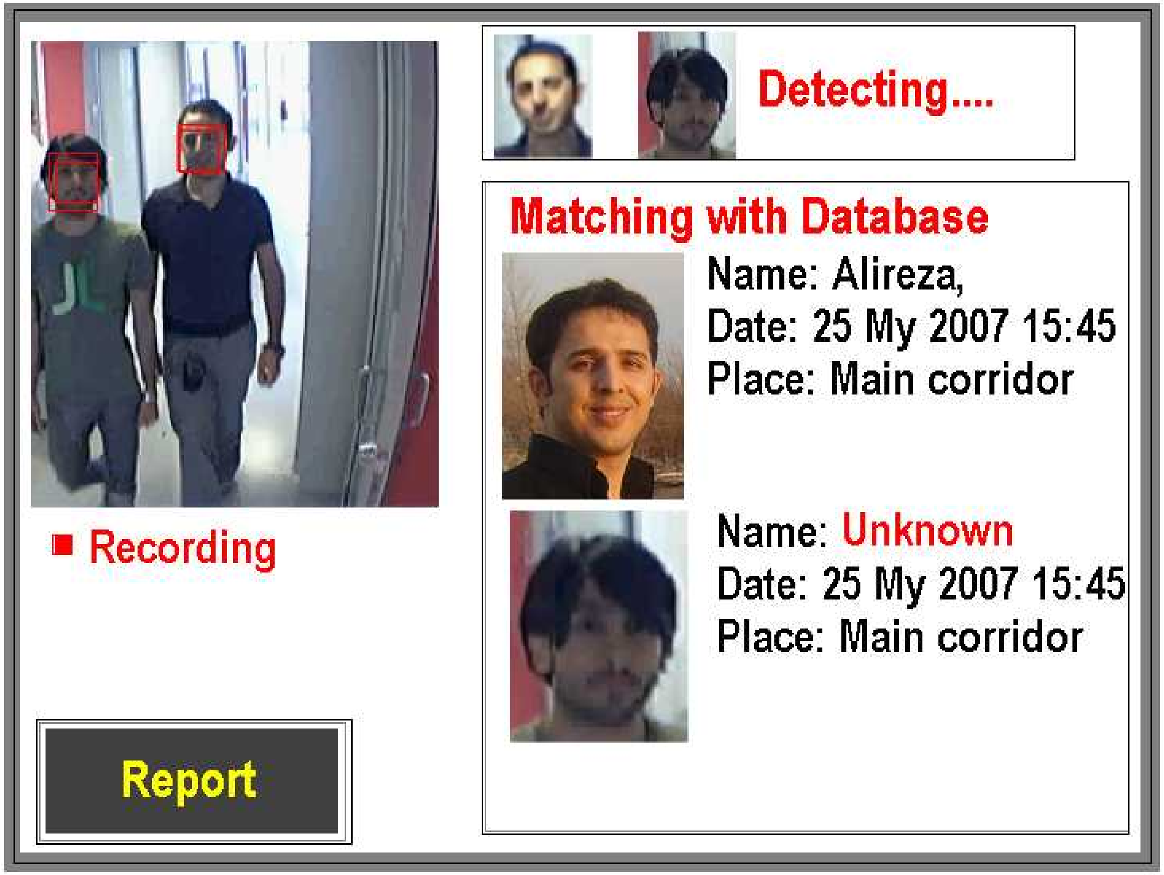
Applications of Face Recognition

- Digital photography



Applications of Face Recognition

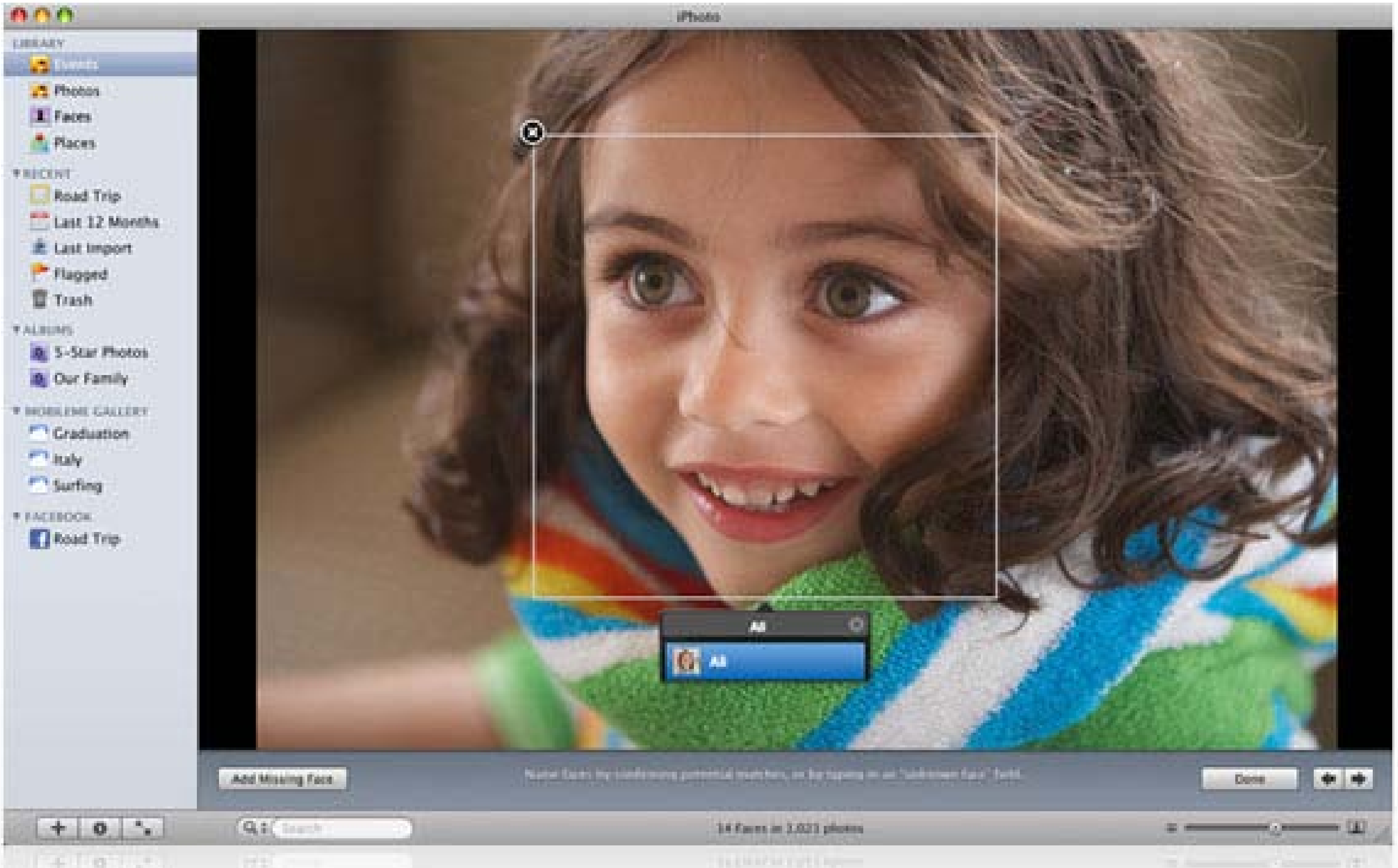
- Digital photography
- Surveillance



Applications of Face Recognition

- Digital photography
- Surveillance
- Album organization

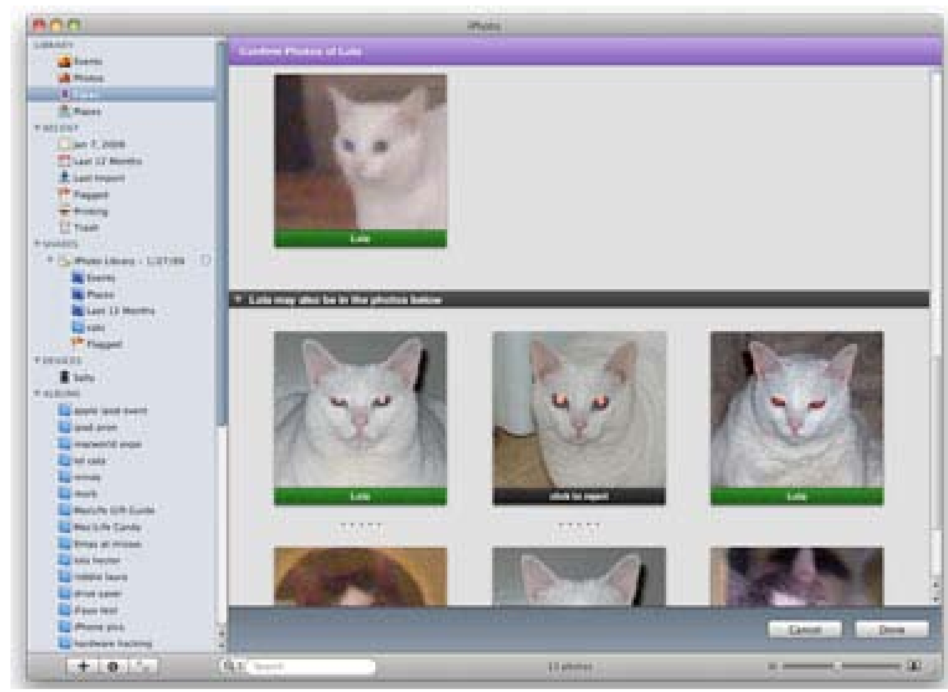
Consumer application: iPhoto 2009



http://www.apple.com/ilife/iphoto/

Consumer application: iPhoto 2009

- Can be trained to recognize pets!

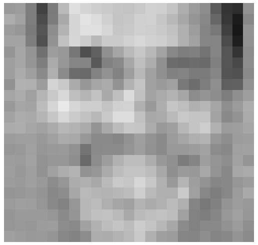
http://www.maclife.com/article/news/iphotos_faces_recognizes_cats

Consumer application: iPhoto 2009

- [Things iPhoto thinks are faces]()

Starting idea of "eigenfaces"

1. Treat pixels as a vector



$\rightarrow \mathbf{x}$

2. Recognize face by nearest neighbor

$\rightarrow \mathbf{y}_1 \ldots \mathbf{y}_n$

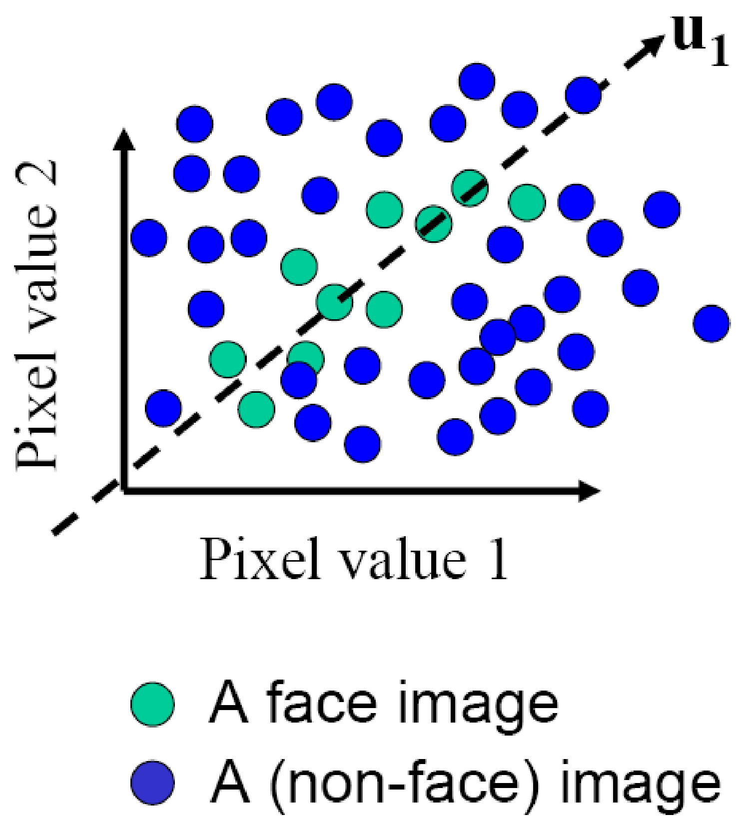
$$k = \underset{k}{\operatorname{argmin}} \left\| \mathbf{y}_k^T - \mathbf{x} \right\|$$

The space of all face images

- When viewed as vectors of pixel values, face images are extremely high-dimensional
  - 100x100 image = 10,000 dimensions
  - Slow and lots of storage
- But very few 10,000-dimensional vectors are valid face images
- We want to effectively model the subspace of face images

The space of all face images

- Eigenface idea: construct a low-dimensional linear subspace that best explains the variation in the set of face images



Principal Component Analysis (PCA)

- Given: N data points $\mathbf{x}_1, \dots, \mathbf{x}_N$ in R^d

- We want to find a new set of features that are linear combinations of original ones:

$$u(\mathbf{x}_i) = \mathbf{u}^T(\mathbf{x}_i - \boldsymbol{\mu})$$

(**μ**: mean of data points)

- Choose unit vector **u** in R^d that captures the most data variance

Forsyth & Ponce, Sec. 22.3.1, 22.3.2

Principal Component Analysis

- Direction that maximizes the variance of the projected data:

(show on board)

Maximize $$\frac{1}{N}\sum_{i=1}^{N} \underbrace{\mathbf{u}^{\mathrm{T}}(\mathbf{x}_i - \mu)}_{\text{Projection of data point}}(\mathbf{u}^{\mathrm{T}}(\mathbf{x}_i - \mu))^{\mathrm{T}}$$ subject to ||**u**||=1

$$= \quad \mathbf{u}^{\mathrm{T}} \underbrace{\left[1/N\sum_{i=1}^{N} (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^{\mathrm{T}}\right]}_{\text{Covariance matrix of data}} \mathbf{u}$$

$$= \quad \mathbf{u}^{\mathrm{T}} \Sigma \mathbf{u}$$

The direction that maximizes the variance is the eigenvector associated with the largest eigenvalue of Σ

Implementation issue

- Covariance matrix is huge (N^2 for N pixels)
- But typically # examples << N
- Simple trick
  - $\mathbf{X}$ is matrix of normalized training data
  - Solve for eigenvectors $\mathbf{u}$ of $\mathbf{X}\mathbf{X}^T$ instead of $\mathbf{X}^T\mathbf{X}$
  - Then $\mathbf{X}^T\mathbf{u}$ is eigenvector of covariance $\mathbf{X}^T\mathbf{X}$
  - May need to normalize (to get unit length vector)

Eigenfaces (PCA on face images)

1. Compute covariance matrix of face images

2. Compute the principal components ("eigenfaces")
   - K eigenvectors with largest eigenvalues

3. Represent all face images in the dataset as linear combinations of eigenfaces
   - Perform nearest neighbor on these coefficients

M. Turk and A. Pentland, Face Recognition using Eigenfaces, CVPR 1991

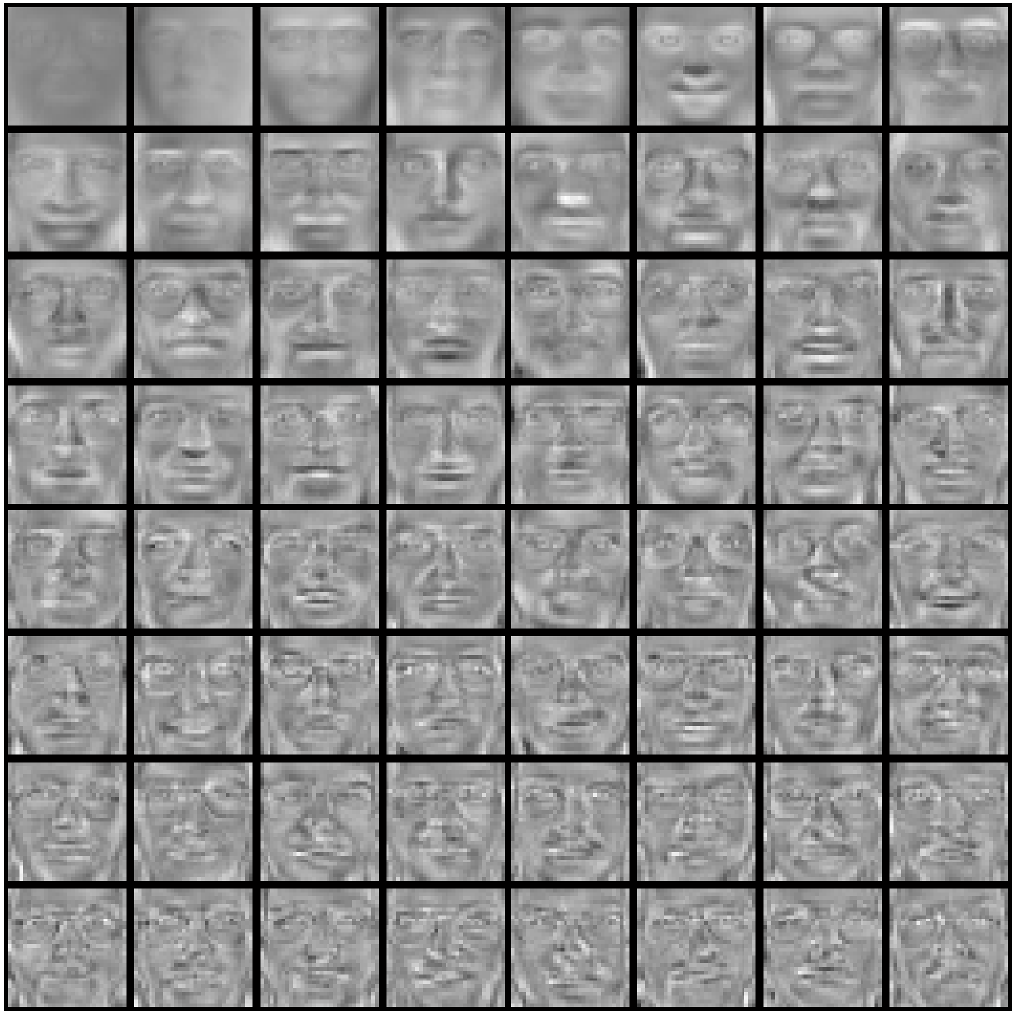
Eigenfaces example

- Training images
- $\mathbf{x}_1,\ldots,\mathbf{x}_N$

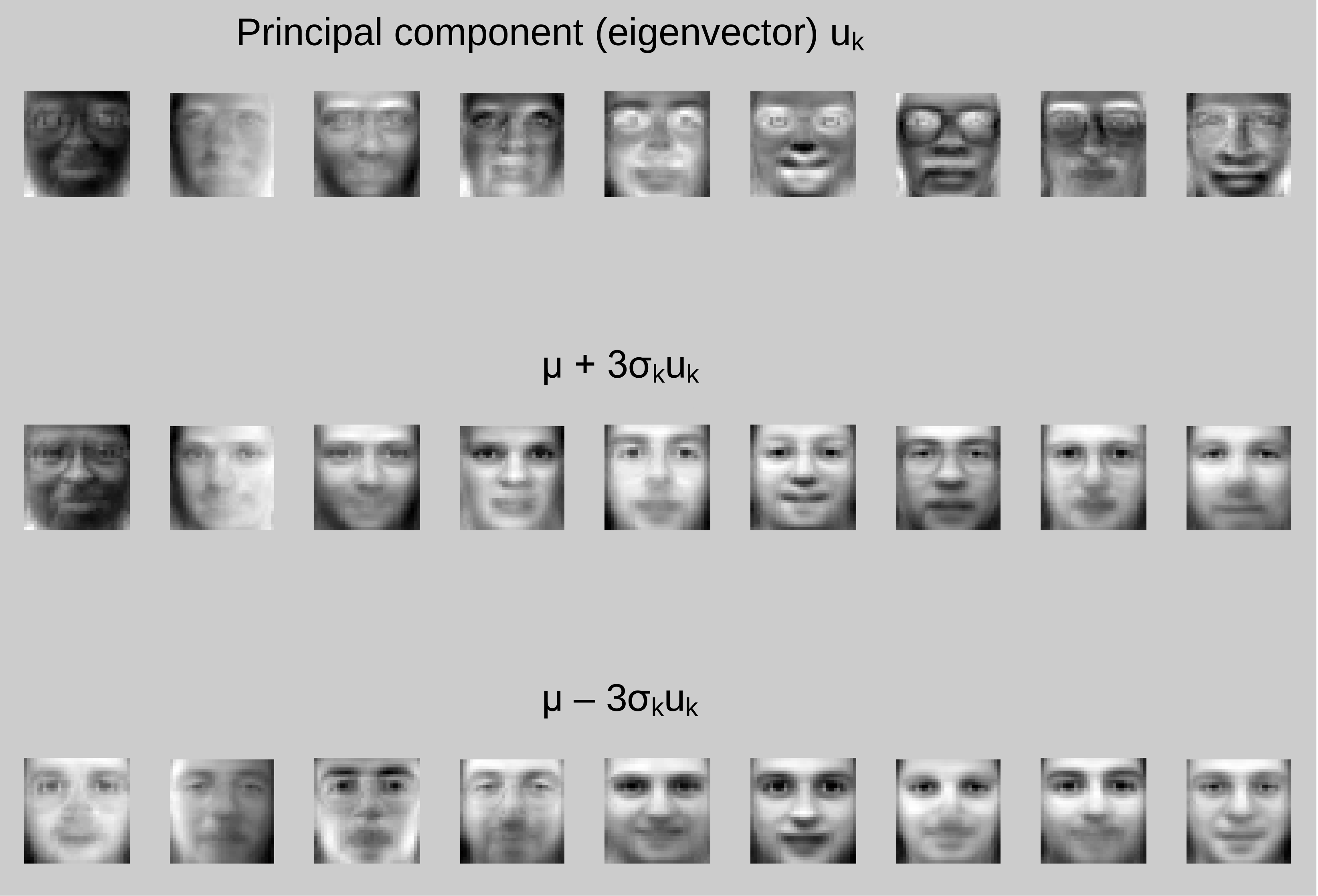
Eigenfaces example

Top eigenvectors: $u_1, \ldots u_k$

Mean: μ



Visualization of eigenfaces



Representation and reconstruction

- Face **x** in "face space" coordinates:

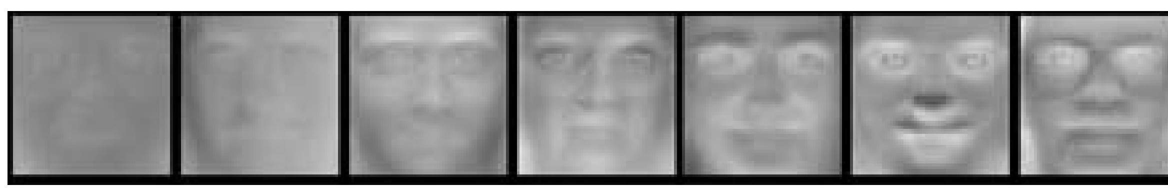
$$\mathbf{x} \rightarrow [\mathbf{u}_1^{\mathrm{T}}(\mathbf{x} - \mu), \ldots, \mathbf{u}_k^{\mathrm{T}}(\mathbf{x} - \mu)]$$
$$= \quad w_1, \ldots, w_k$$

Representation and reconstruction

- Face **x** in “face space” coordinates:

$$\mathbf{x} \rightarrow [\mathbf{u}_1^{\mathrm{T}}(\mathbf{x} - \mu), \ldots, \mathbf{u}_k^{\mathrm{T}}(\mathbf{x} - \mu)]$$
$$= \quad w_1, \ldots, w_k$$

- Reconstruction:

 =  + 

$$\hat{x} = \mu + w_1u_1+w_2u_2+w_3u_3+w_4u_4+ \ldots$$

Reconstruction

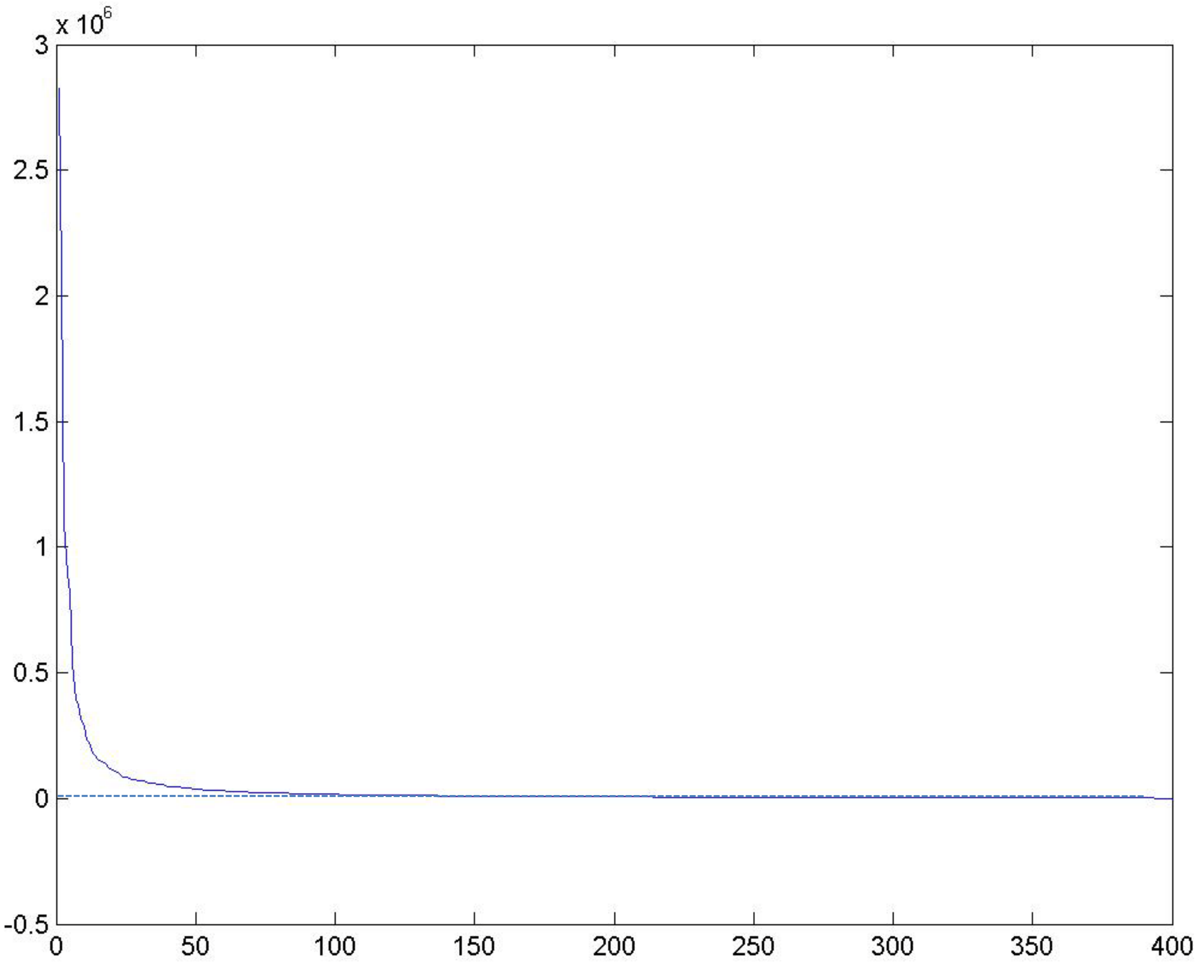
P = 4

P = 200

P = 400

After computing eigenfaces using 400 face images from ORL face database

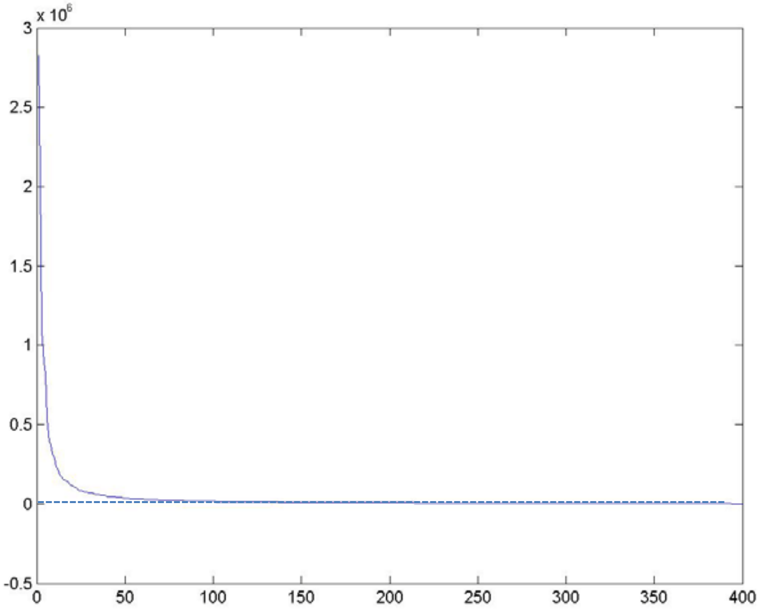
Eigenvalues (variance along eigenvectors)



Note

Preserving variance (minimizing MSE) does not necessarily lead to qualitatively good reconstruction.

P = 200



Recognition with eigenfaces

Process labeled training images

- Find mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
- Find k principal components (eigenvectors of $\boldsymbol{\Sigma}$) $\mathbf{u}_1,\ldots\mathbf{u}_k$
- Project each training image $\mathbf{x}_i$ onto subspace spanned by principal components:
  $(w_{i1},\ldots,w_{ik}) = (\mathbf{u}_1^T(\mathbf{x}_i - \boldsymbol{\mu}), \ldots , \mathbf{u}_k^T(\mathbf{x}_i - \boldsymbol{\mu}))$

Given novel image $\mathbf{x}$

- Project onto subspace:
  $(w_1,\ldots,w_k) = (\mathbf{u}_1^T(\mathbf{x} - \boldsymbol{\mu}), \ldots , \mathbf{u}_k^T(\mathbf{x} - \boldsymbol{\mu}))$
- Optional: check reconstruction error $\mathbf{x} - \hat{\mathbf{x}}$ to determine whether image is really a face
- Classify as closest training face in k-dimensional subspace

M. Turk and A. Pentland, [Face Recognition using Eigenfaces], CVPR 1991

PCA

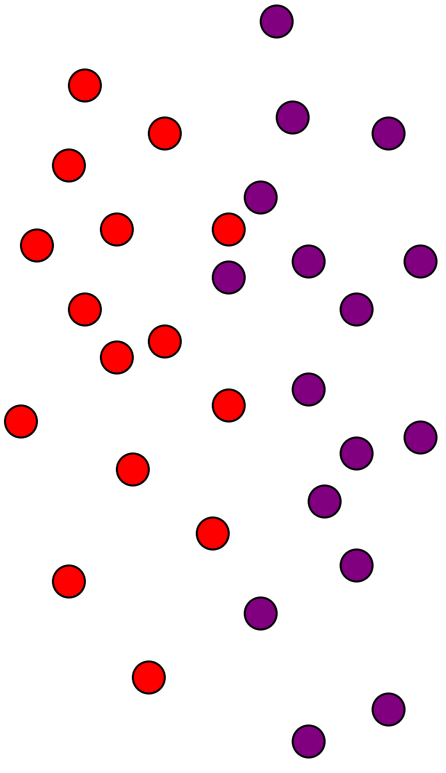
- General dimensionality reduction technique
- Preserves most of variance with a much more compact representation
  - Lower storage requirements (eigenvectors + a few numbers per face)
  - Faster matching
- What are the problems for face recognition?

Limitations

Global appearance method: not robust to misalignment, background variation

Limitations

- The direction of maximum variance is not always good for classification



A more discriminative subspace: FLD

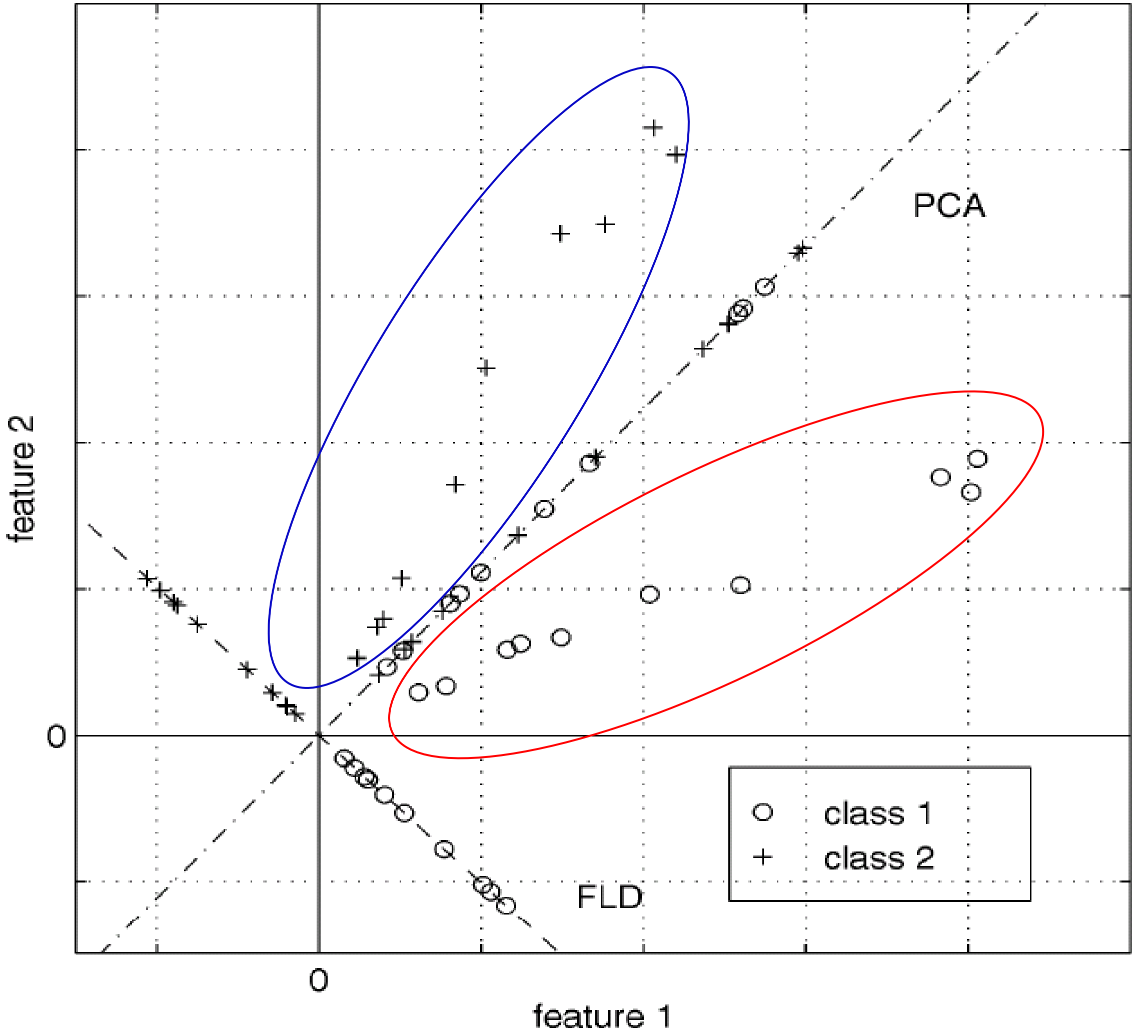
- Fisher Linear Discriminants → “Fisher Faces”

- PCA preserves maximum variance

- FLD preserves discrimination
  - Find projection that maximizes scatter between classes and minimizes scatter within classes

Reference: [Eigenfaces vs. Fisherfaces, Belheumer et al., PAMI 1997]

Comparing with PCA



Variables

- N Sample images: $\{x_1, \cdots, x_N\}$
- c classes: $\{\chi_1, \cdots, \chi_c\}$

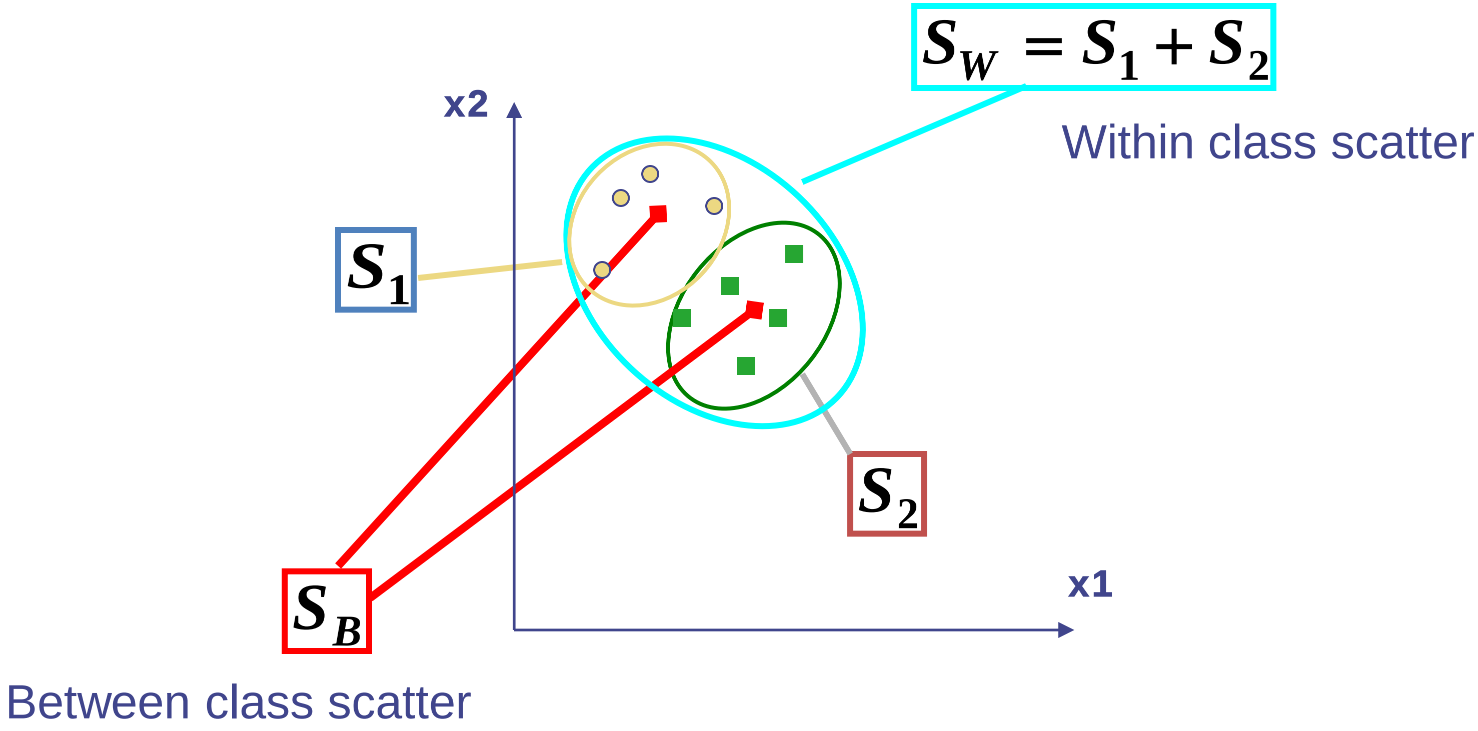
- Average of each class: $\mu_i = \frac{1}{N_i} \sum_{x_k \in \chi_i} x_k$

- Average of all data: $\mu = \frac{1}{N} \sum_{k=1}^{N} x_k$

Scatter Matrices

- Scatter of class i: $S_i = \sum_{x_k \in \chi_i} (x_k - \mu_i)(x_k - \mu_i)^T$
- Within class scatter: $S_W = \sum_{i=1}^{c} S_i$
- Between class scatter: $S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T$

Illustration



Mathematical Formulation

- After projection $y_k = W^T x_k$
  - Between class scatter $\tilde{S}_B = W^T S_B W$
  - Within class scatter $\tilde{S}_W = W^T S_W W$
- Objective

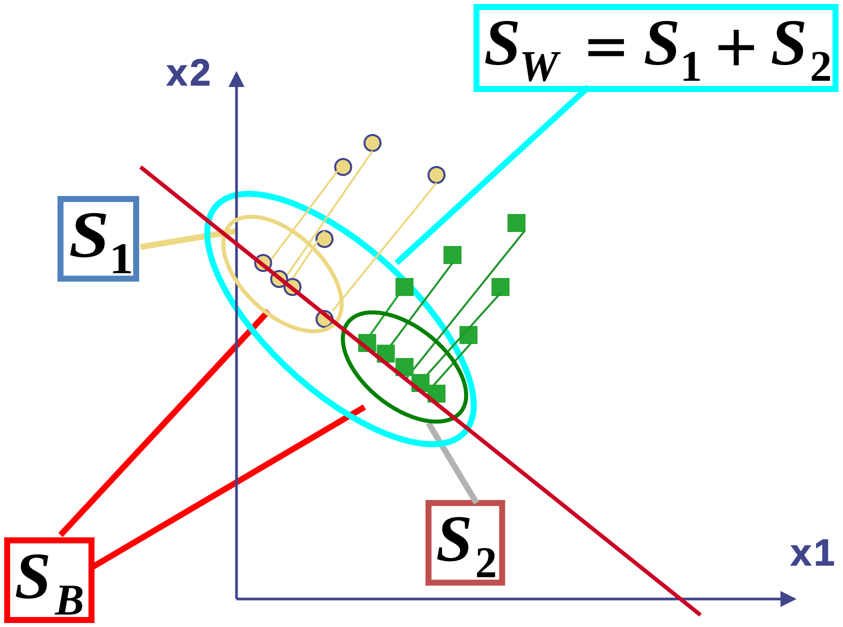
$$W_{opt} = \arg\max_{W} \frac{\left|\tilde{S}_B\right|}{\left|\tilde{S}_W\right|} = \arg\max_{W} \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|}$$

- Solution: Generalized Eigenvectors

$$S_B w_i = \lambda_i S_W w_i \qquad i = 1, \ldots, m$$

- Rank of $\boldsymbol{W_{opt}}$ is limited
  - Rank(S_B) <= |C|-1
  - Rank(S_W) <= N-C

Illustration



Recognition with FLD

- Similar to "eigenfaces"

- Compute within-class and between-class scatter matrices

$$S_i = \sum_{x_k \in \chi_i} (x_k - \mu_i)(x_k - \mu_i)^T \qquad S_W = \sum_{i=1}^{c} S_i \qquad S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

- Solve generalized eigenvector problem

$$W_{opt} = \arg\max_{\mathrm{W}} \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|} \qquad S_B w_i = \lambda_i S_W w_i \qquad i = 1, \ldots, m$$

- Project to FLD subspace and classify by nearest neighbor

$$\hat{x} = W_{opt}{}^T x$$

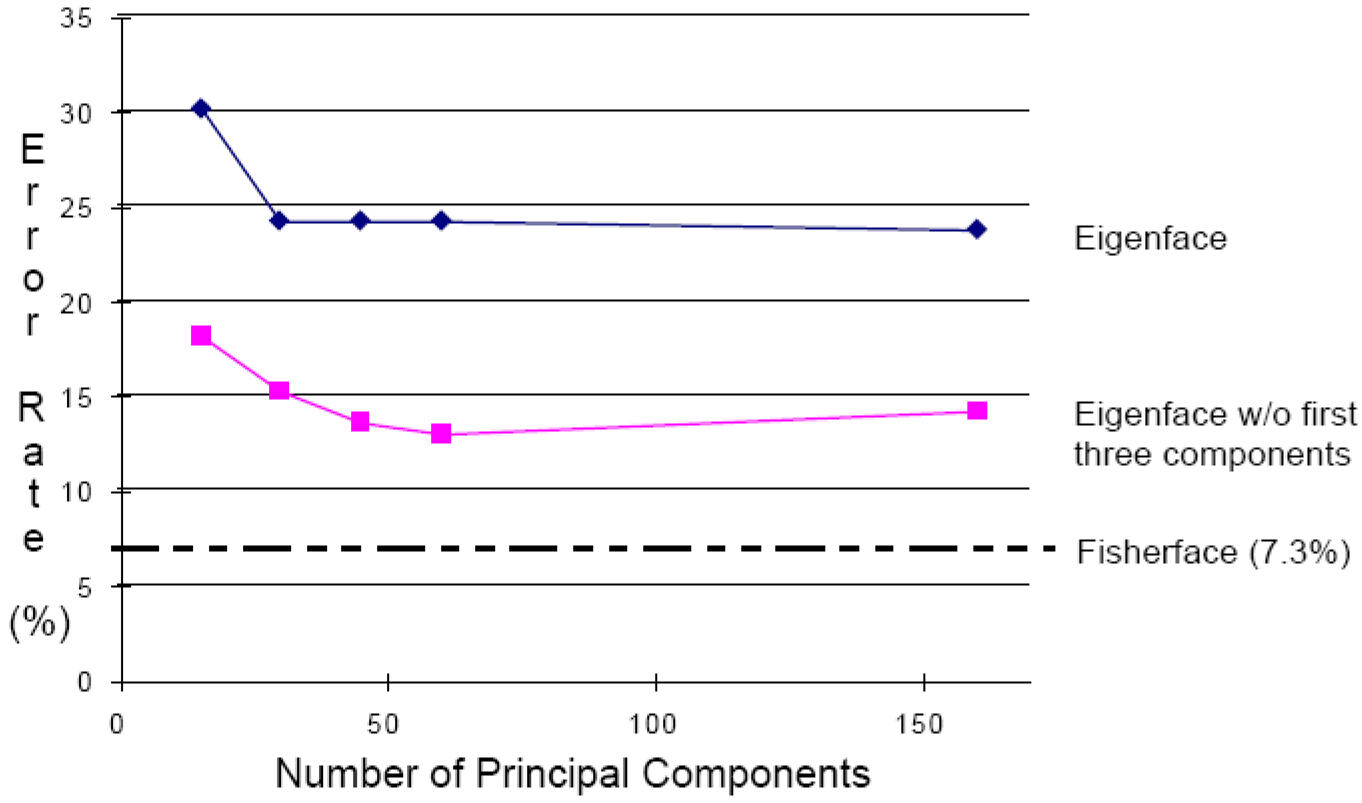
Results: Eigenface vs. Fisherface

- Input: 160 images of 16 people
- Train: 159 images
- Test: 1 image

- Variation in Facial Expression, Eyewear, and Lighting

Reference: [Eigenfaces vs. Fisherfaces, Belheumer et al., PAMI 1997]

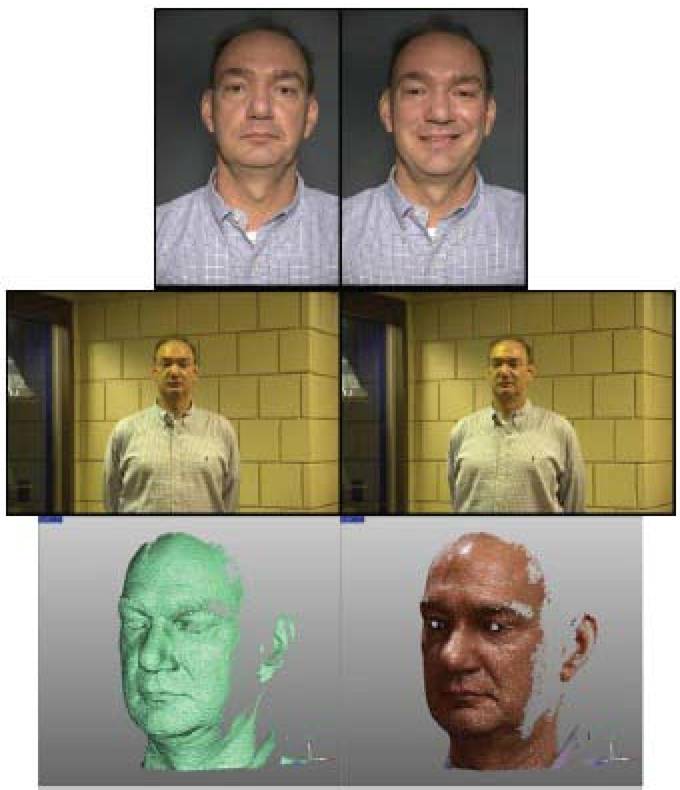
Eigenfaces vs. Fisherfaces



Reference: [Eigenfaces vs. Fisherfaces, Belheumer et al., PAMI 1997]

Large scale comparison of methods

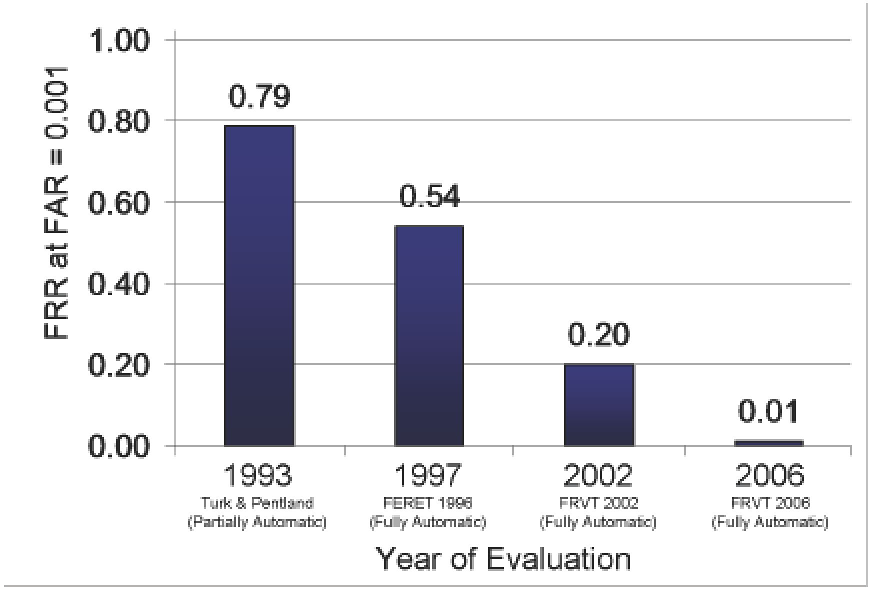
- FRVT 2006 Report
- Not much (or any) information available about methods, but gives idea of what is doable



FVRT Challenge

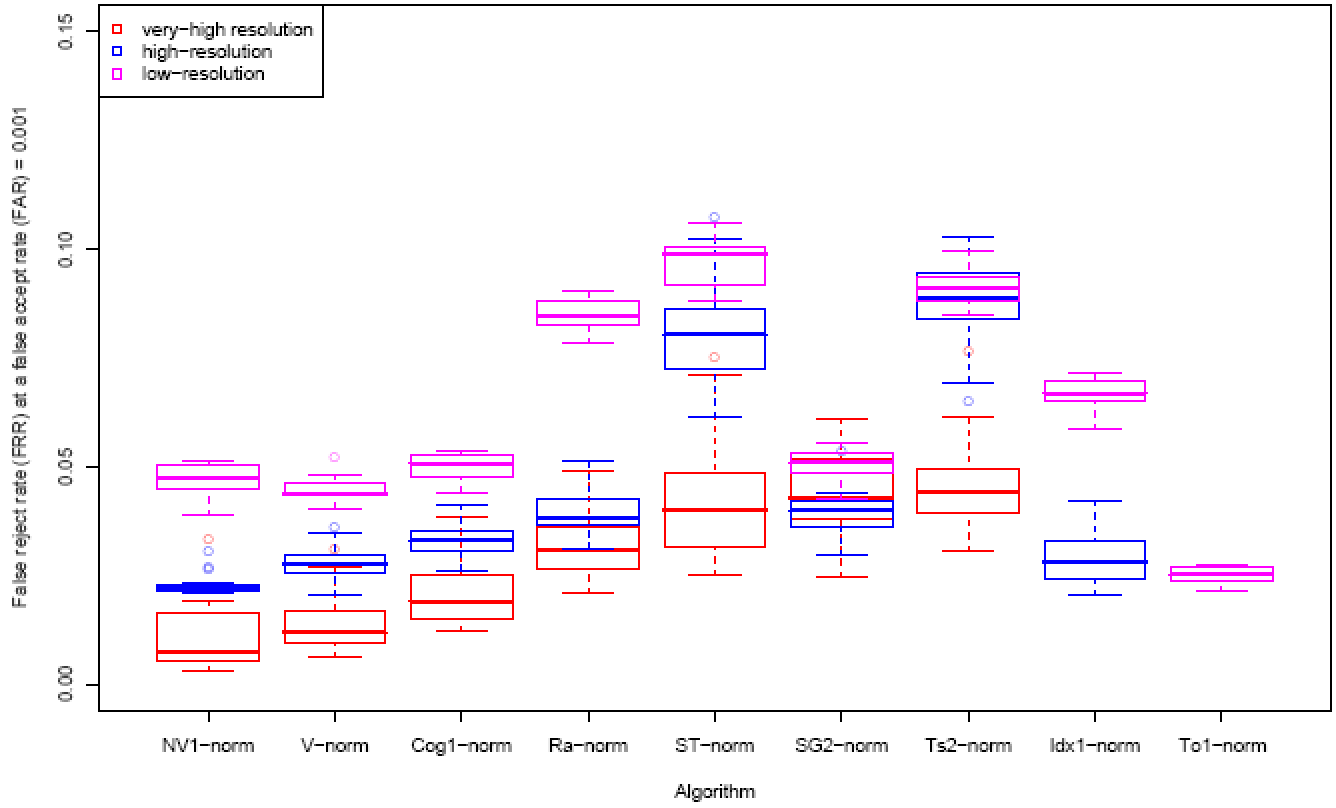
- Frontal faces
  - FVRT2006 evaluation

False Rejection Rate at False Acceptance Rate = 0.001



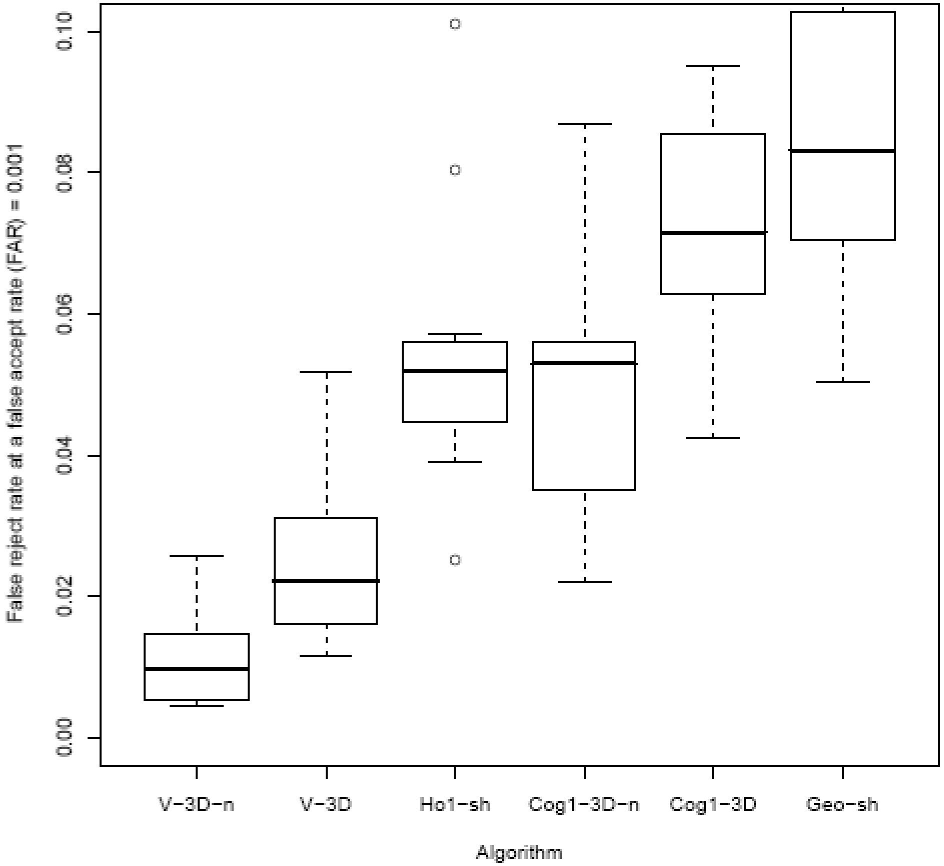
FVRT Challenge

- Frontal faces
  - FVRT2006 evaluation: controlled illumination



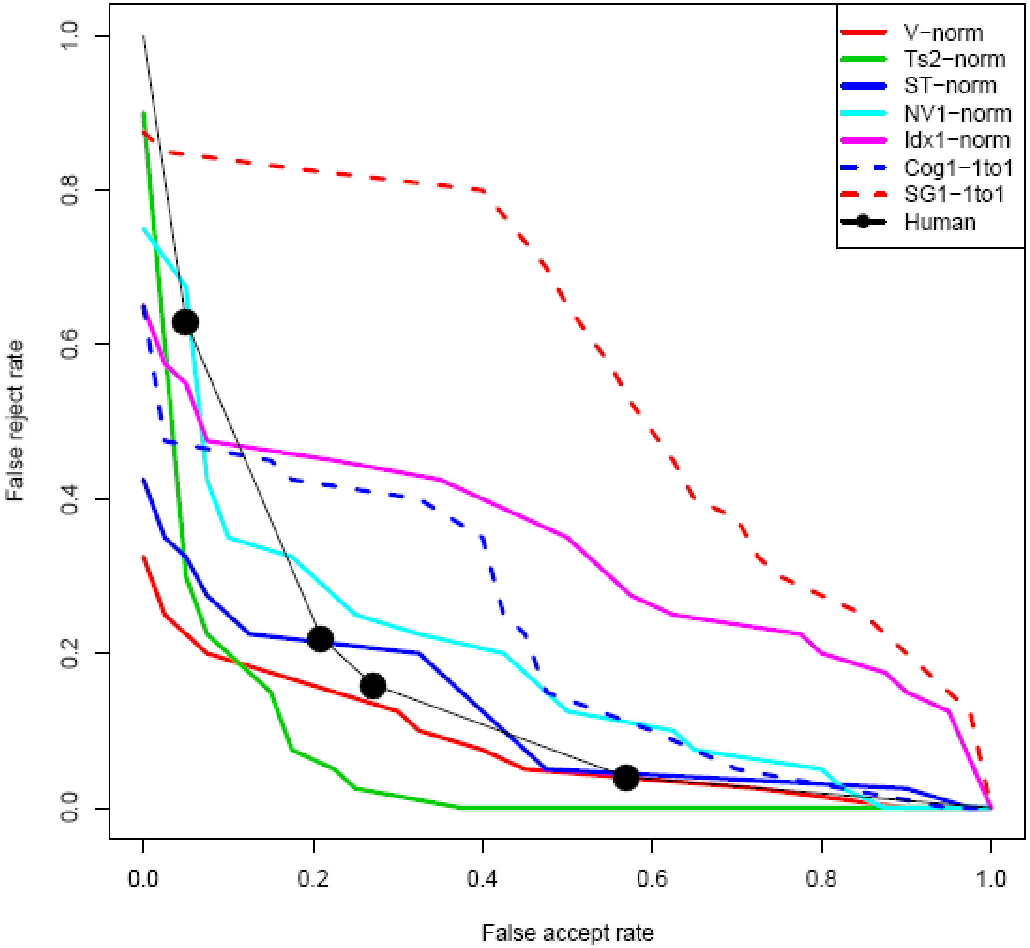
FVRT Challenge

- Frontal faces
  - FVRT2006 evaluation: uncontrolled illumination



FVRT Challenge

- Frontal faces
  - FVRT2006 evaluation: computers win!



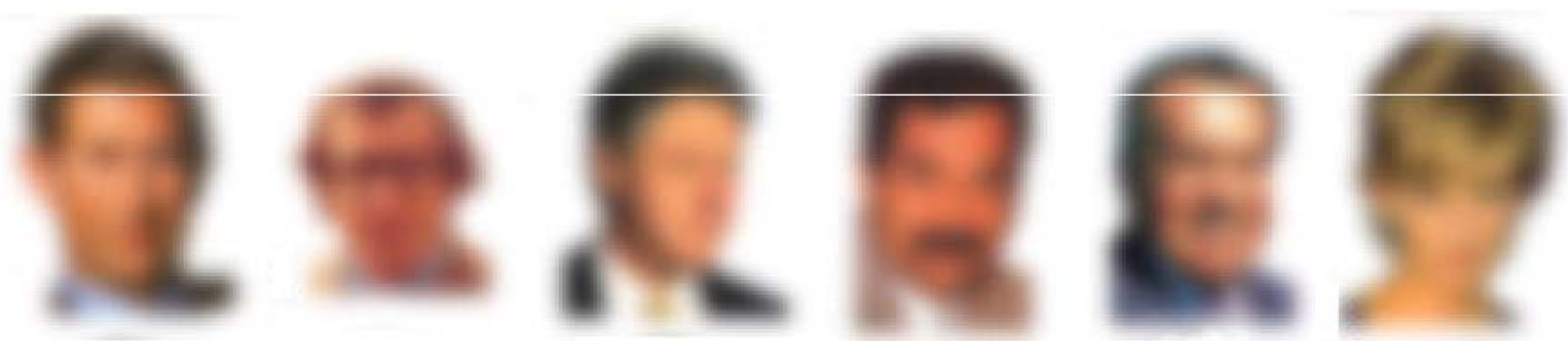
Face recognition by humans

Face recognition by humans: 20 results (2005)

Slides by Jianchao Yang

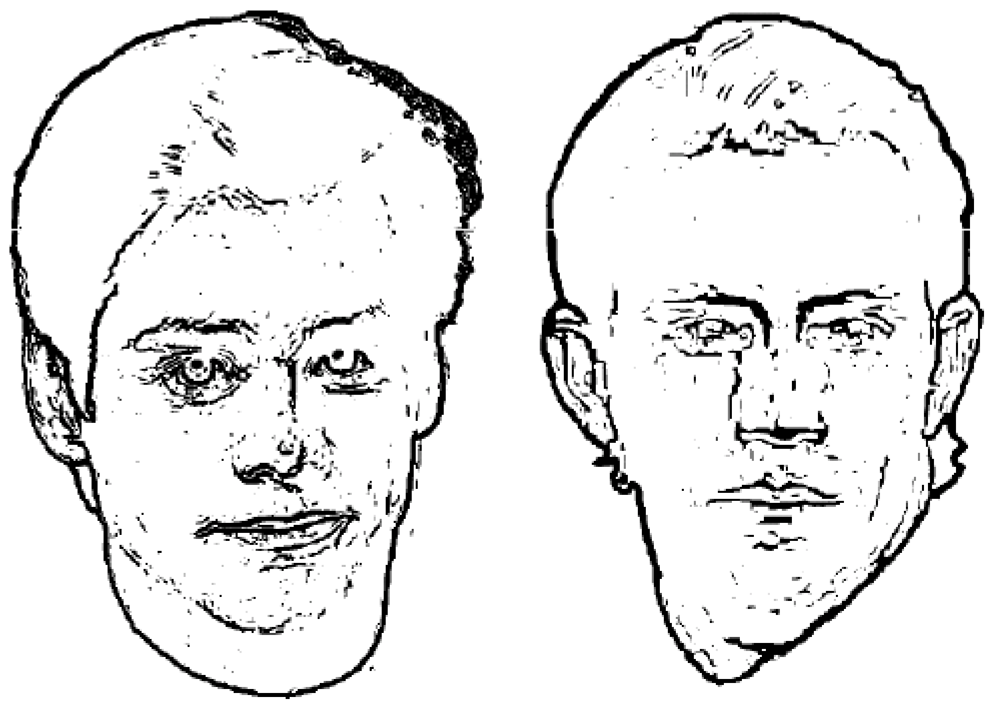
Result 1

- Humans can recognize faces in extremely low resolution images.



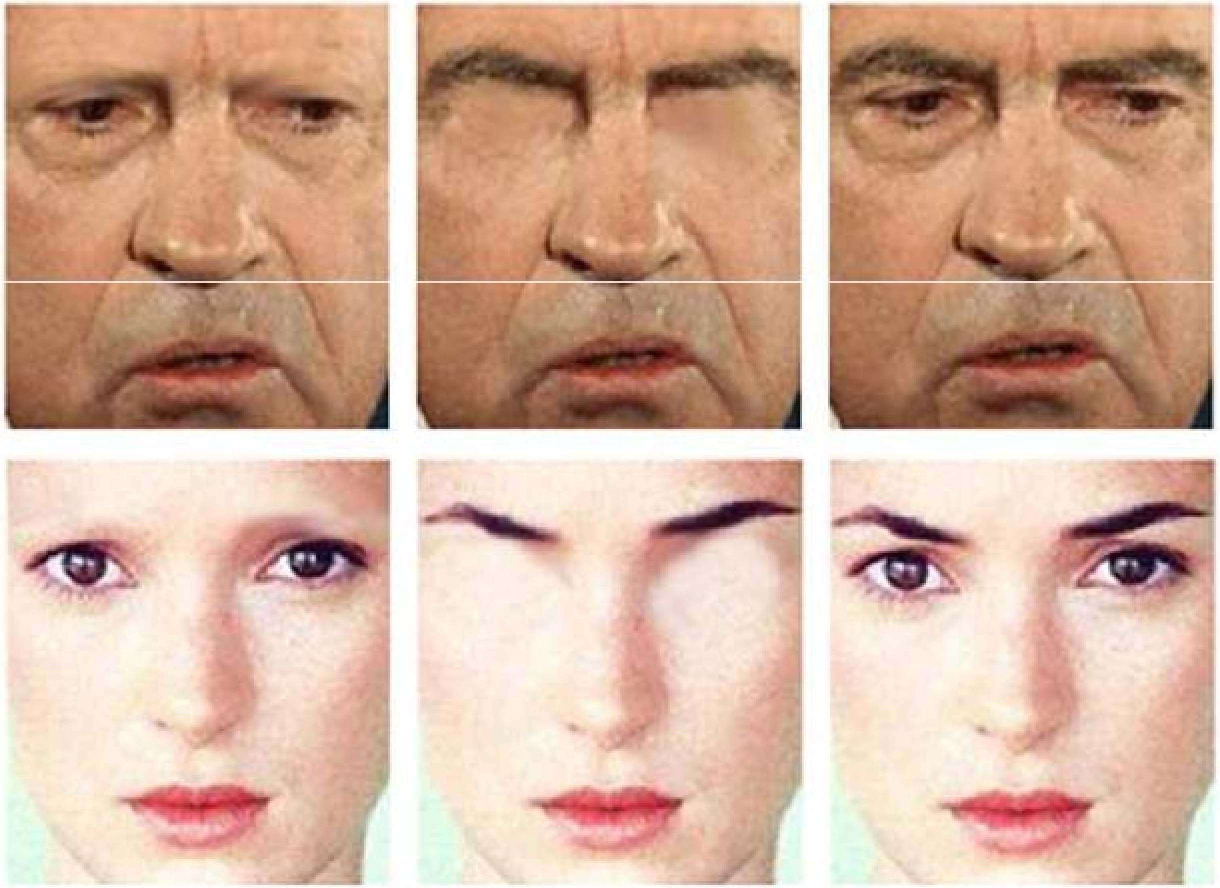
Result 3

- High-frequency information by itself does not lead to good face recognition performance



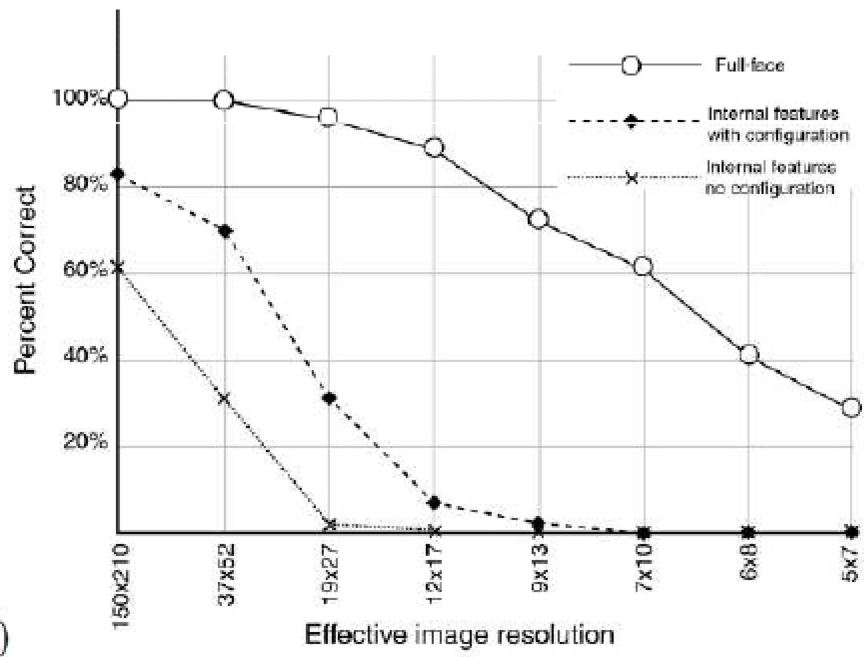
Result 5

- Eyebrows are among the most important for recognition



Result 6

- Both internal and external facial cues are important and they exhibit non-linear interactions

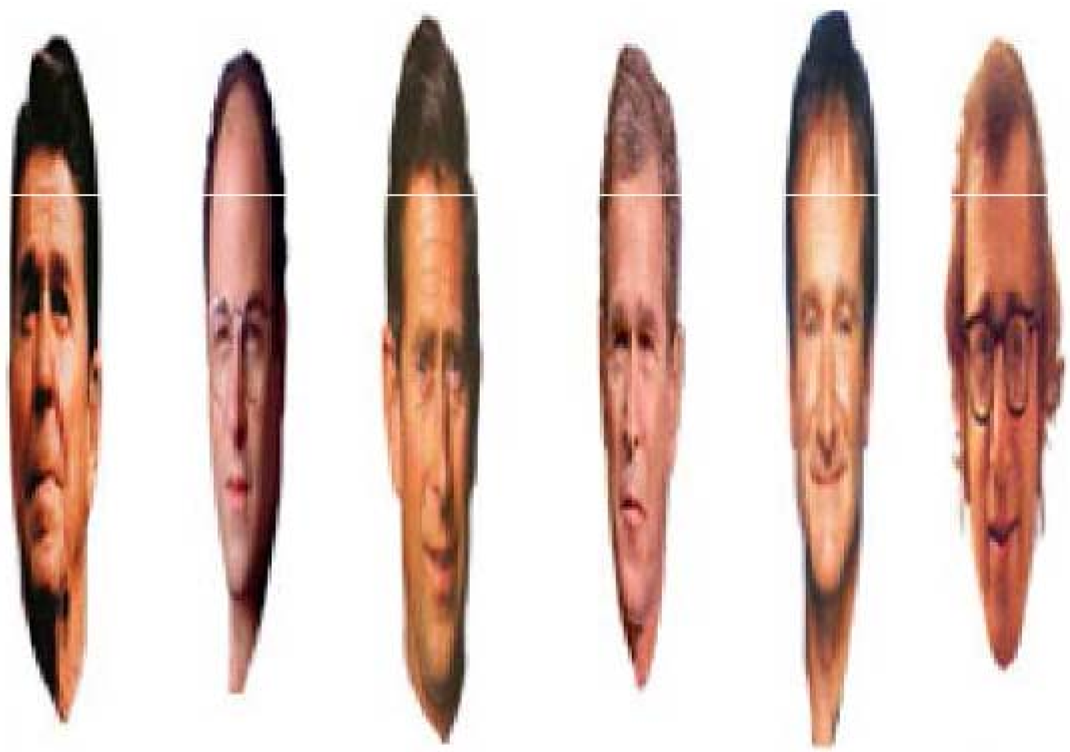


(a)

(b)

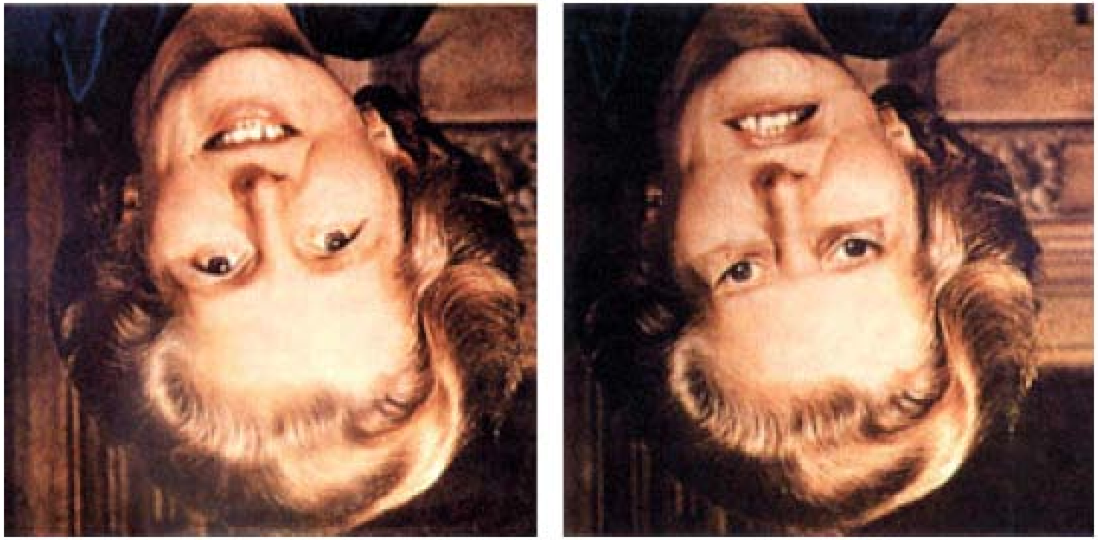
Result 7

- The important configural relations appear to be independent across the width and height dimensions



Result 8

- Vertical inversion dramatically reduces recognition performance

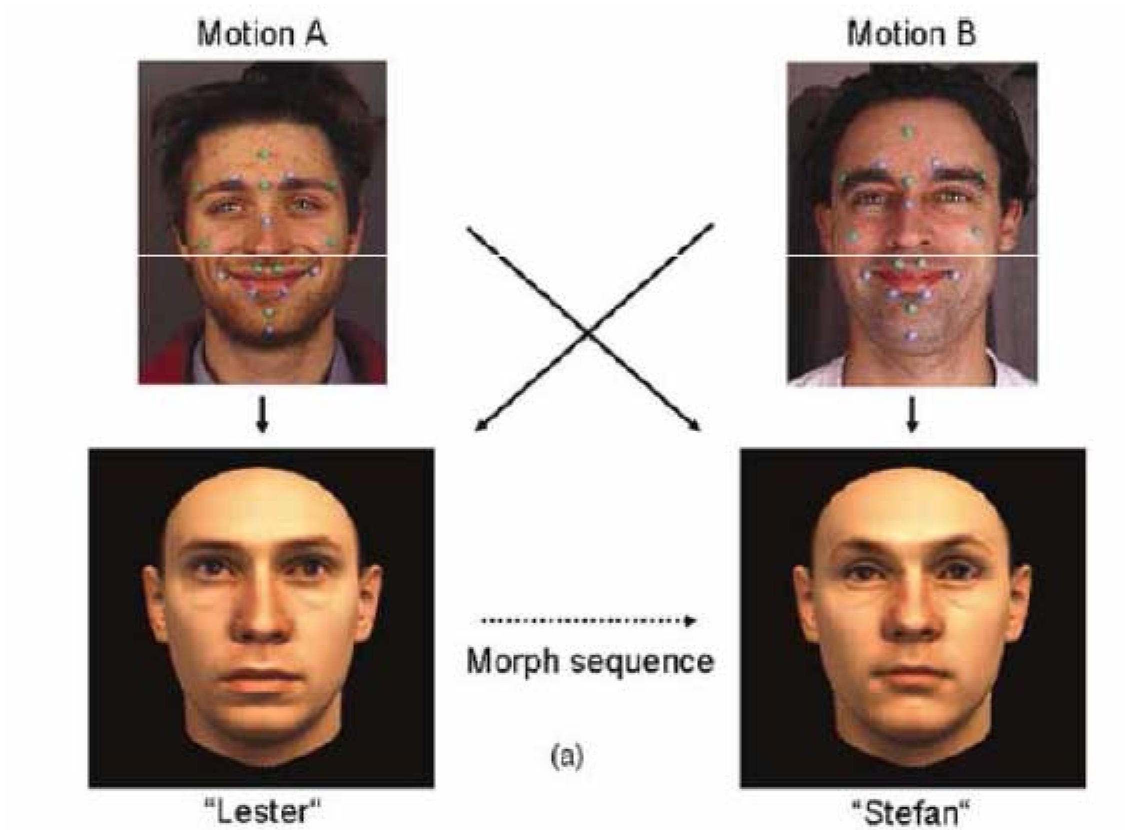


Result 12

- Contrast polarity inversion dramatically impairs recognition performance, possibly due to compromised ability to use pigmentation cues

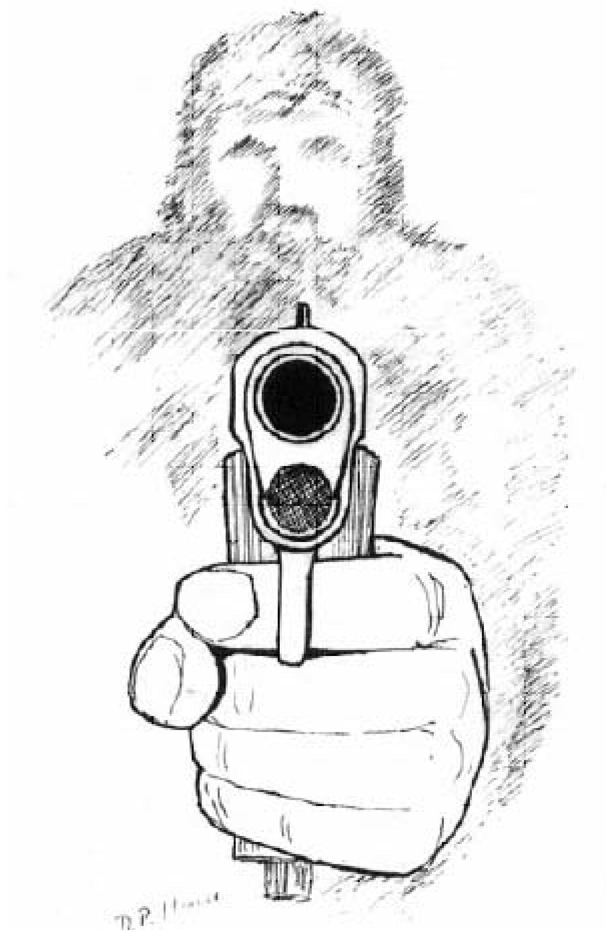
Result 15

- Motion of faces appears to facilitate subsequent recognition



Result 20

- Human memory for briefly seen faces is rather poor



Things to remember

- PCA is a generally useful dimensionality reduction technique
  - But not ideal for discrimination

- FLD better for discrimination, though only ideal under Gaussian data assumptions

- Computer face recognition works very well under controlled environments – still room for improvement in general conditions

Next class

- Image categorization: features and classifiers