ECCV 2004 T4 Tutorial

Face Recognition and Modeling-Part I

Face Recognition: A Tutorial

Wen-Yi Zhao

wzhao@sarnoff.com

Sarnoff Corporation

Princeton, NJ 08543

Face Recognition: A Literature Survey. W. Zhao, R. Chellappa, A. Rosenfeld, and J. Phillips

University of Maryland Tech. Rep. CS-TR-4167R, 2002/ACM Computing Survey, Vol. 35, Dec, 2003

Application I: Robot Pets

Sony Aibo to spread more puppy love

By John G. Spooner
Staff Writer, CNET News.com
October 10, 2002, 7:09 AM PT

Sony is planning to train Aibo, its robot dog, to be able to pick you out of a crowd.

Sony's Entertainment Robot America division said Tuesday it will introduce Aibo Recognition, a new application for its newest Aibo ERS-210A and ERS-210 models. The software will grant the mechanical dog the ability to recognize its owner's name, voice and **face,** as well as automatically recharge itself.

Sony's Aibo will soon recognize its owner.

Application II: Cell Phone

March 27, 2002

Face Recognition, via Cell Phones

By Ryan Naraine

Chicago-based telecoms equipment maker Motorola, Inc. has announced plans to put face recognition technology into Java-enabled mobile phones.

In partnership with Visionics Corp. (Quote, Company Info) and Wirehound LLC, Motorola said the application was being developed specifically for law enforcement agencies.

The announcement, made at the JavaOne Developer Conference in San Francisco, said the application would use Visionics' FaceIt ARGUS as the delivery platform for facial recognition capabilities.

Motorola would also install Wirehound's Birddog software on the its i95cl phone, a J2ME technology-enabled mobile phone with a color display.

The FaceIt ARGUS system would automatically find **faces** in a field of view and search them against a mug shot database. "Upon finding a match, the Birddog component generates a wireless alert to the phones used by mobile law enforcement officials, who are then able to verify the identity of the subject," the company said.

Application III: Access Control

The SmartGate installation at Sydney International Airport made major use of Cognitec's software for automatic border control

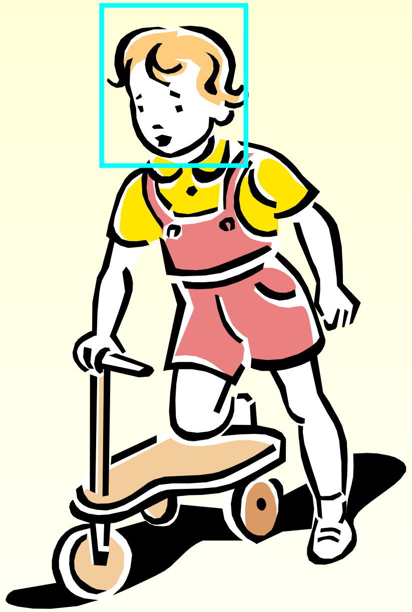
Samsung's Magicgate:
A Door Lock Control System using Face Verification Technology

Application IV: Entertainment

Image manipulation

Blanz & Vetter

Emily?



Family Photo Album
By Microsoft

Outline

- Introduction
- Relevant Psychophysics/Neuroscience Issues
- Still-Image Face Recognition
- Video-Based Face Recognition
- Illumination & Pose Problem
- 3D Face Recognition
- Performance Evaluation (FRVT 2002)
- Challenges and Directions

Introduction

Task of Face Recognition

Why Face Recognition?

- Face is the personal communication center
  - Carrying ID (Face recognition)
  - Speech recognition (enhanced by lip-reading)
  - Emotion through facial expression and voice

- Suitable for numerous applications
  - User-friendly (v.s. finger-print, iris recognition)
  - Natural human computer interface (virtual reality, games, robotic dogs)

How to Recognize Faces?

- How human perceives faces?
  - Psychophysics and neuroscience study (Darwin, 1872)
- How machine perceives faces?
  - From simple geometric profiles to complex features (Galton)
- We need to apply all relevant disciplines
  - Psychology: providing guidance/lesson
  - Image/video processing: pre-processing, feature
  - Pattern recognition/Learning/neural networks: classifier
  - Computer vision & Computer graphics: 2D-3D-2D

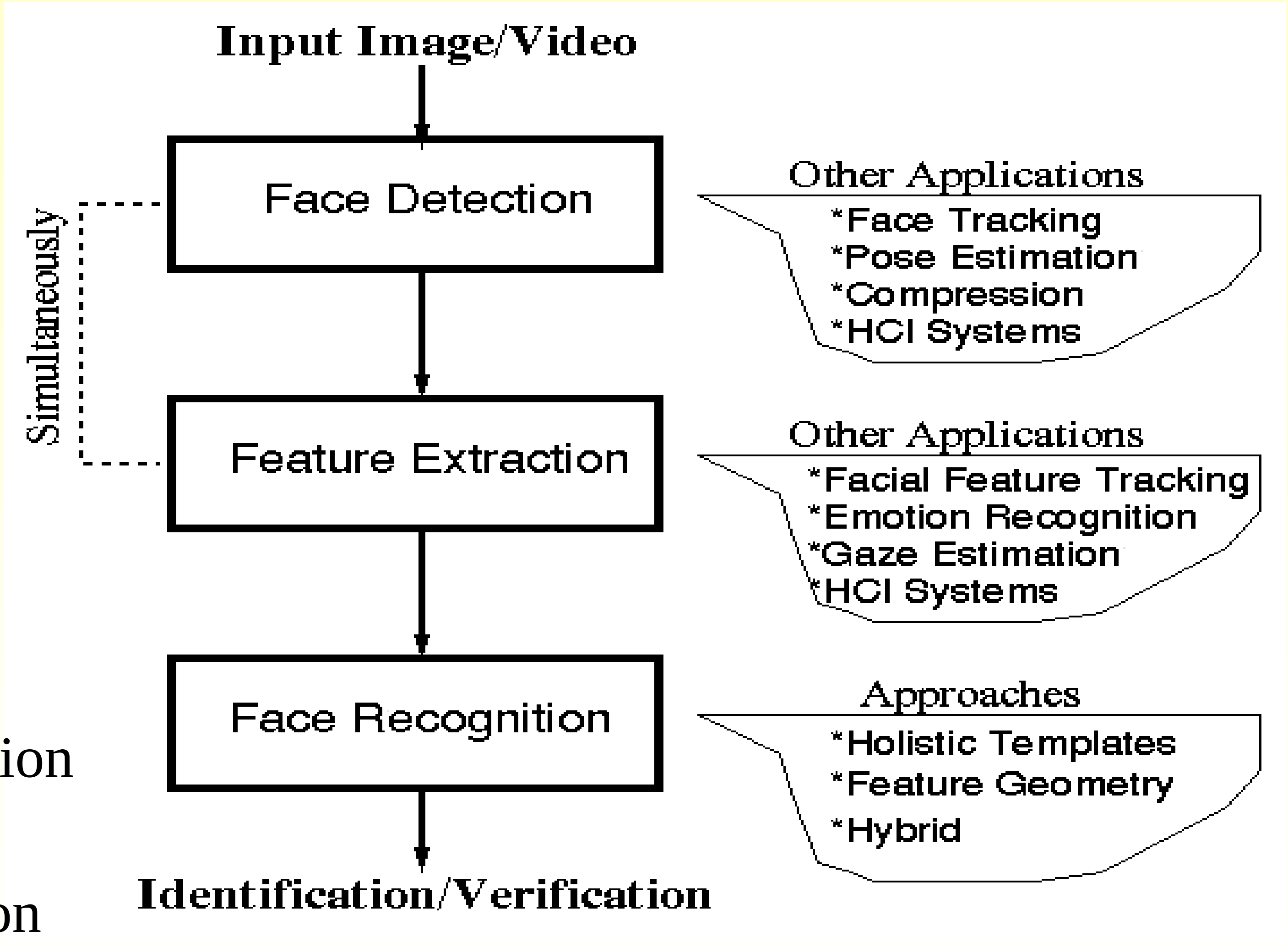
Available Commercial Systems

FaceIt from Visionics	www.FaceIt.com
Viisage Technology	www.viisage.com
FaceVACS from Plettac	www.plettac-electronics.com
FaceKey Corp.	www.facekey.com
Cognitec Systems	www.cognitec-systems.de
Keyware Technologies	www.keywareusa.com
Passfaces from ID-arts	www.id-artss.com
Image Ware Sofware	www.iwsinc.com
Eyematic Interfaces Inc.	www.eyematic.com
BioID sensor fusion	www.bioid.com
Visionsphere Technologies	www.visionspheretech.com/menu.htm
Biometric Systems, Inc.	www.biometrica.com
FaceSnap Recoder	www.facesnap.de/htdocs/english/index2.html
SpotIt for face composite	spotit.itc.it/SpotIt.html

Typical Applications

Areas	Specific Applications
Entertainment	Video Game/Virtual Reality/Training Programs Human-Computer-Interaction/Human-Robotics Family Photo Album
Smart Cards	Drivers' Licenses/Passports/Voter Registrations/Entitlement Programs Welfare Fraud/Passports/Voter Registration
Information Security	TV Parental control/Desktop Logon/Personal Device (Cell phone etc) Logon/Database Security/ Medical Records/Internet Access
Law Enforcement & Surveillance	Advanced Video Surveillance/CCTV Control Shoplifting/Drug Trafficking/Portal Control

Generic System Configuration

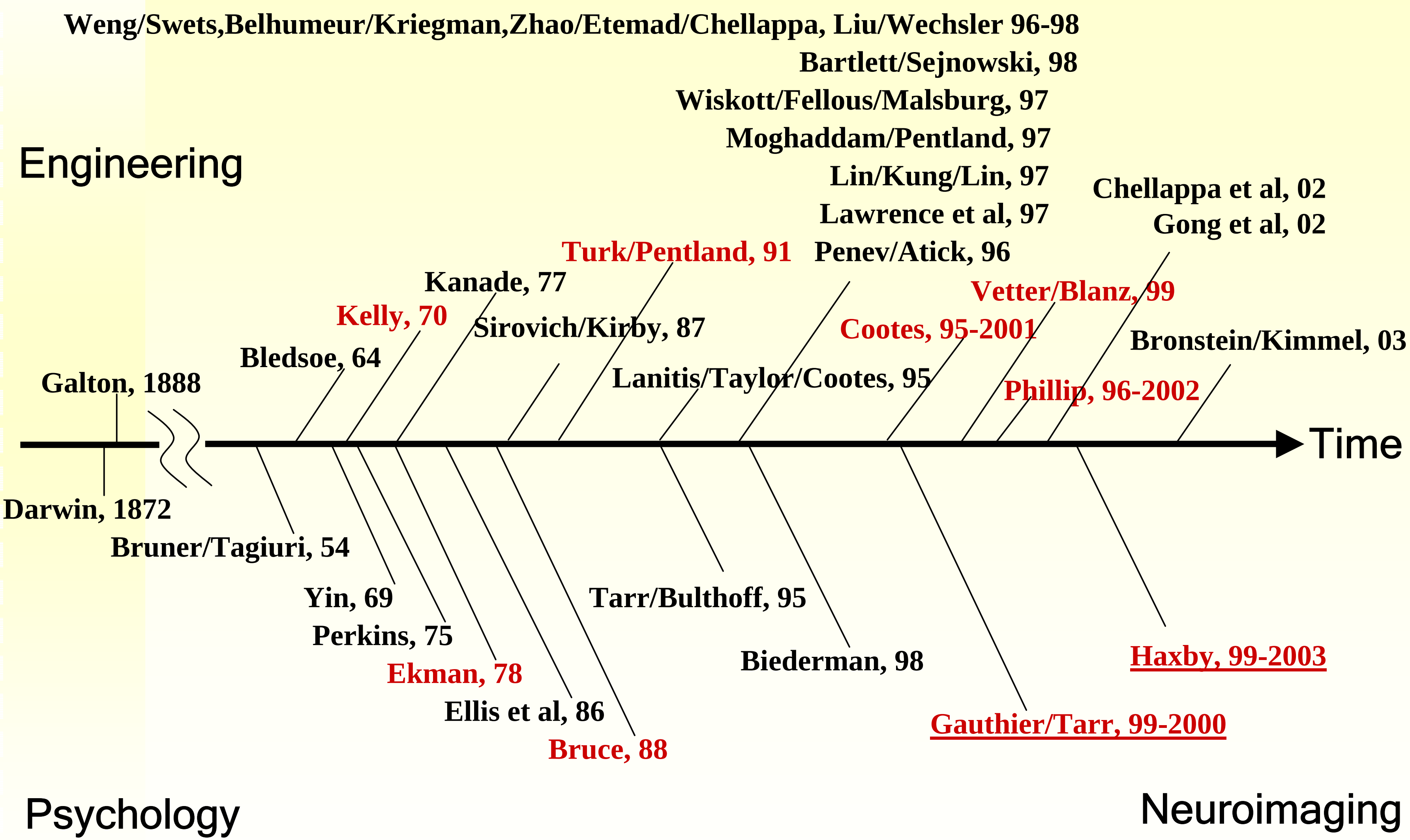


Identification

v.s.

Verification

Brief History

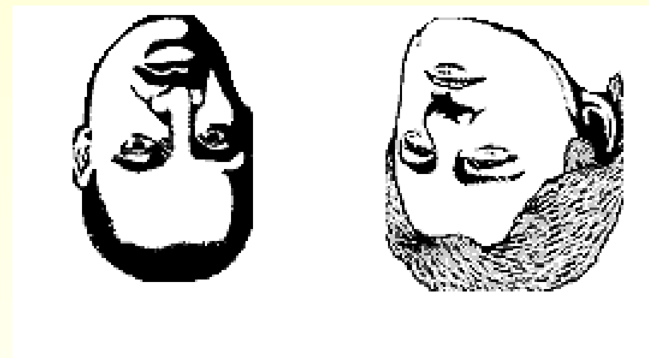


Relevant Psychophysics/Neuroscience Issues

Psychophysics Issues
Relevant to Machine Recognition

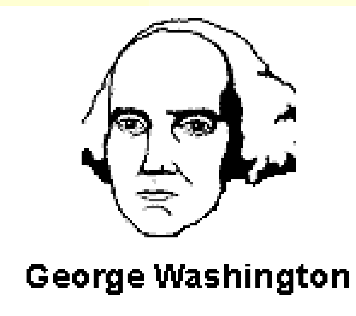
- Is face perception the results of holistic/configural or local feature analysis? Is it a dedicated process?
  - ➔Build a special or generic system
- Ranking of significance of facial features
  - ➔Different weights for features
- Viewpoint-invariant or view-dependent recognition
  - ➔Handle pose problem
- Effect of lighting change
  - ➔Shape from shading
- Movement and face recognition
  - ➔Favor video based recognition
- Role of race/gender/familiarity
  - ➔Suggest adaptive system

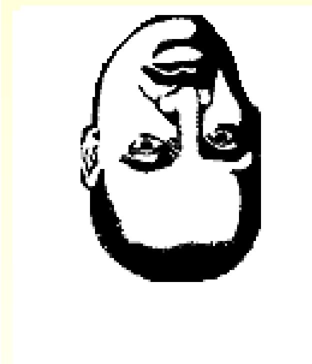
Inverted Faces

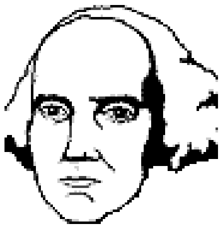
Prof. Eric H. Chudler
Dept. of Anesthesiology
University of Washingtonc

Inverted Faces



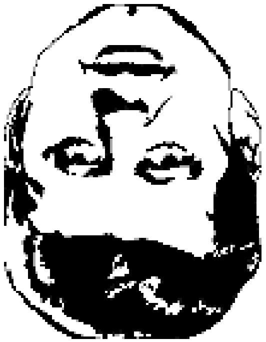


Inverted Faces

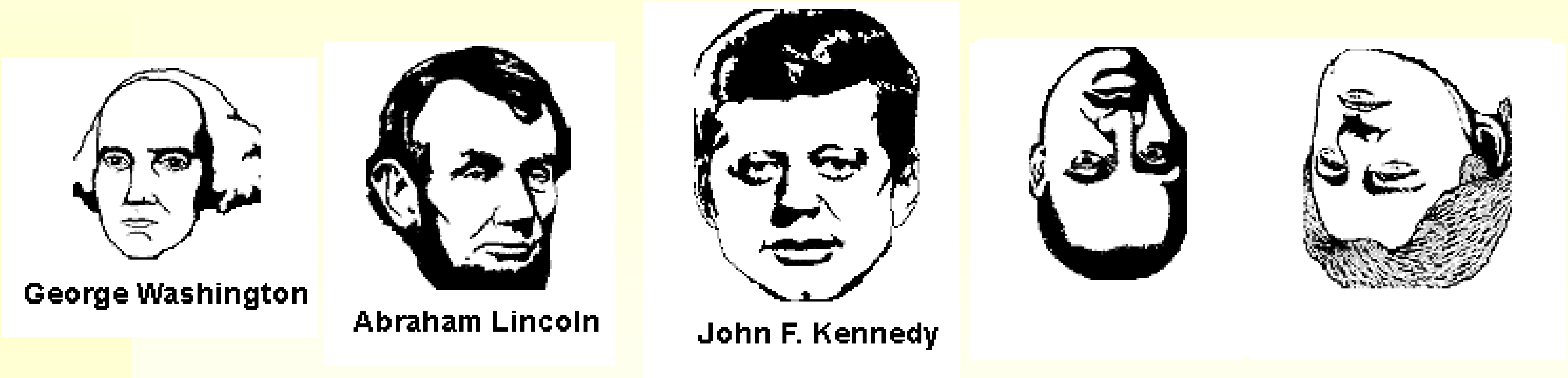


George Washington

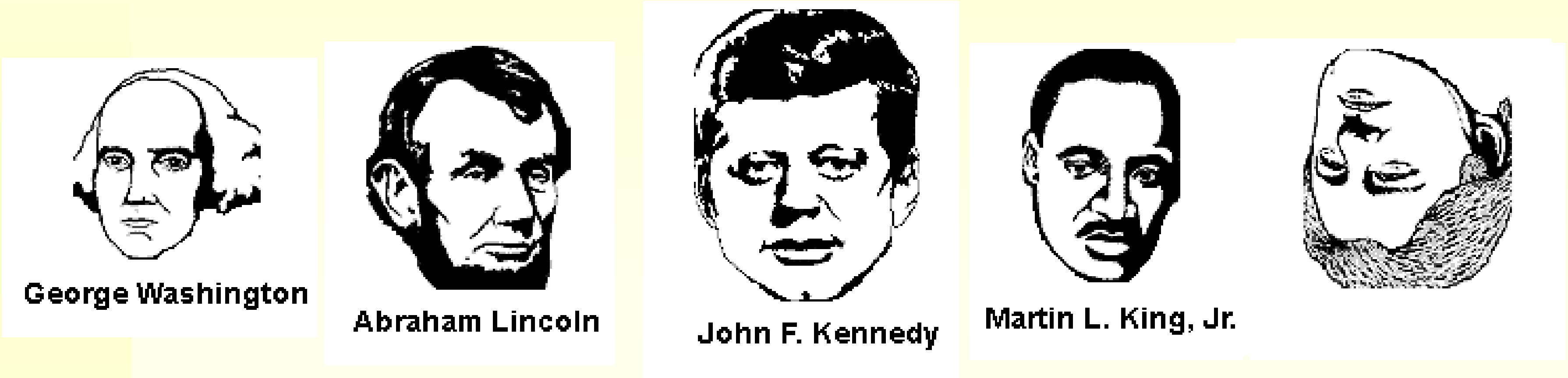
Abraham Lincoln



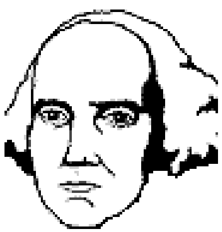
Inverted Faces



Inverted Faces



Inverted Faces



George Washington

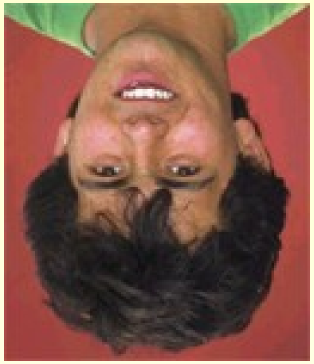
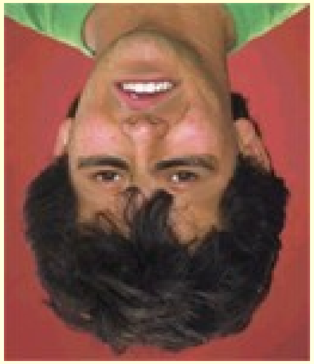
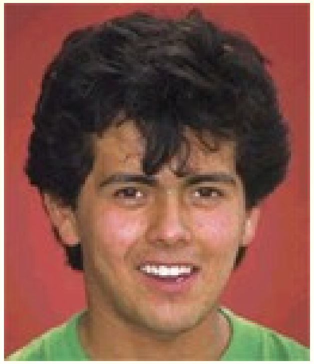
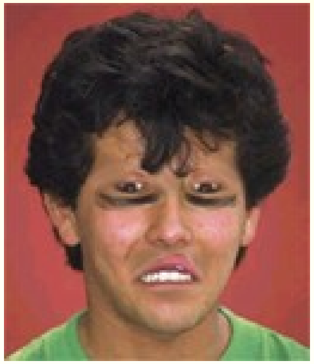
Abraham Lincoln

John F. Kennedy

Martin L. King, Jr.

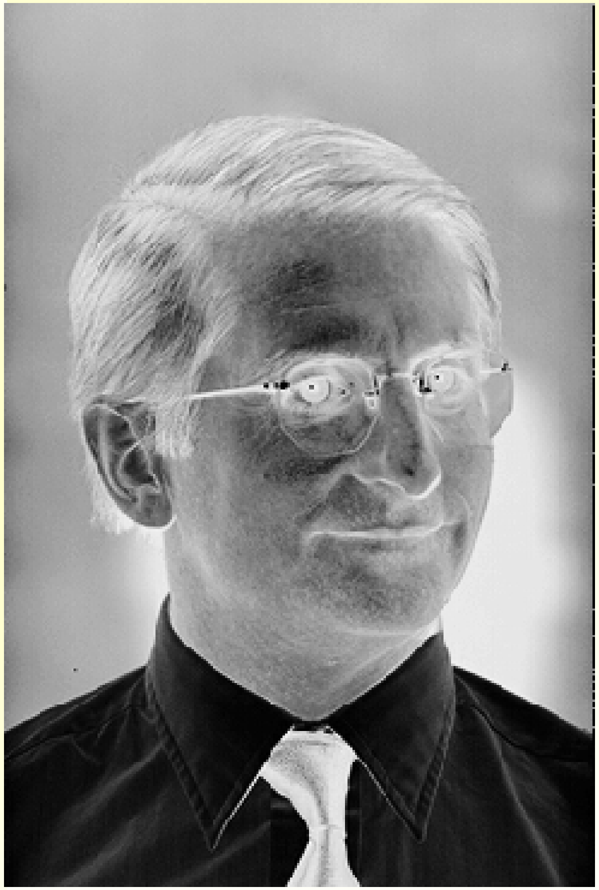
Bill Clinton

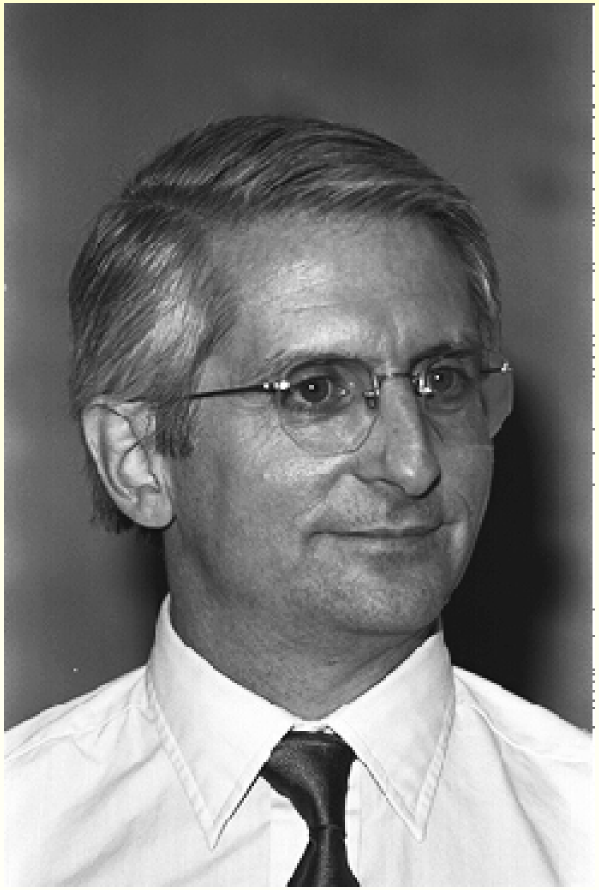
More Inverted Faces

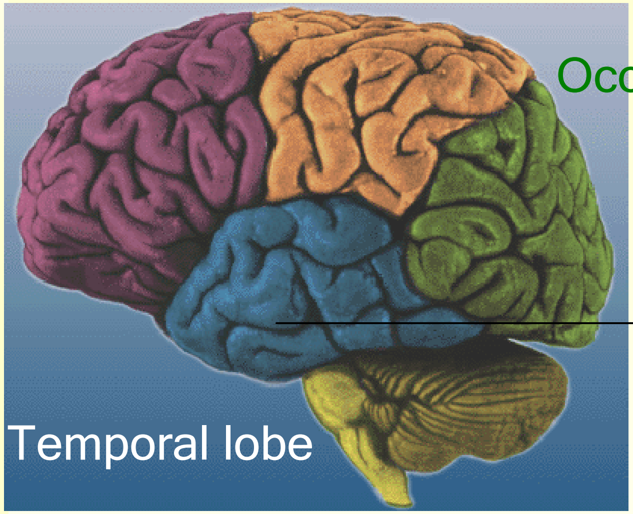
Prof. Eric H. Chudler
Dept. of Anesthesiology
University of Washington

Negated Faces





Neuroimaging Study



Cortical regions (lateral view)

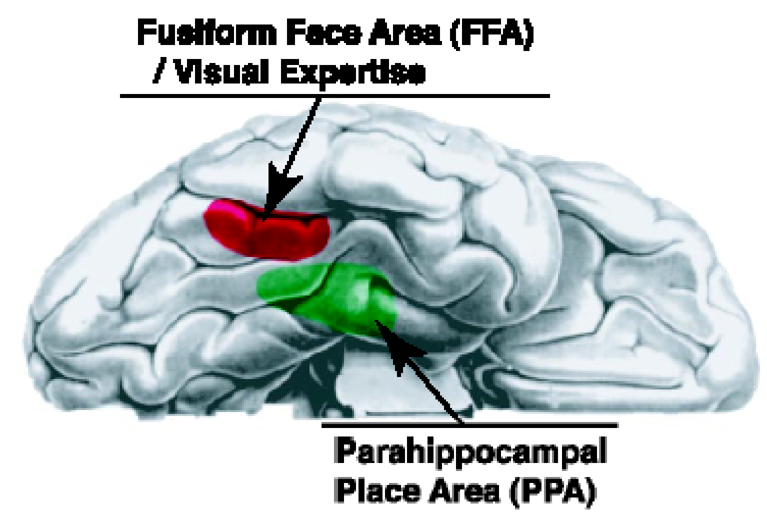
Single-cell recording studies

•Leads to the discovery of individual neurons that are tuned to faces and other objects

Structural Studies

•Leads to large-scale spatial organization for specialization within the ventral object vision pathway

The Functional Architectures



Schematic diagram illustrating the location of FFA and PPA on the ventral surface of the right temporal lobe

Model I:

•specialized areas for representing specific categories of stimuli: FAA and PPA

Model II:

greeble

•different areas are specialized for different types of perceptual processes: FAA is for expert recognition, not just face

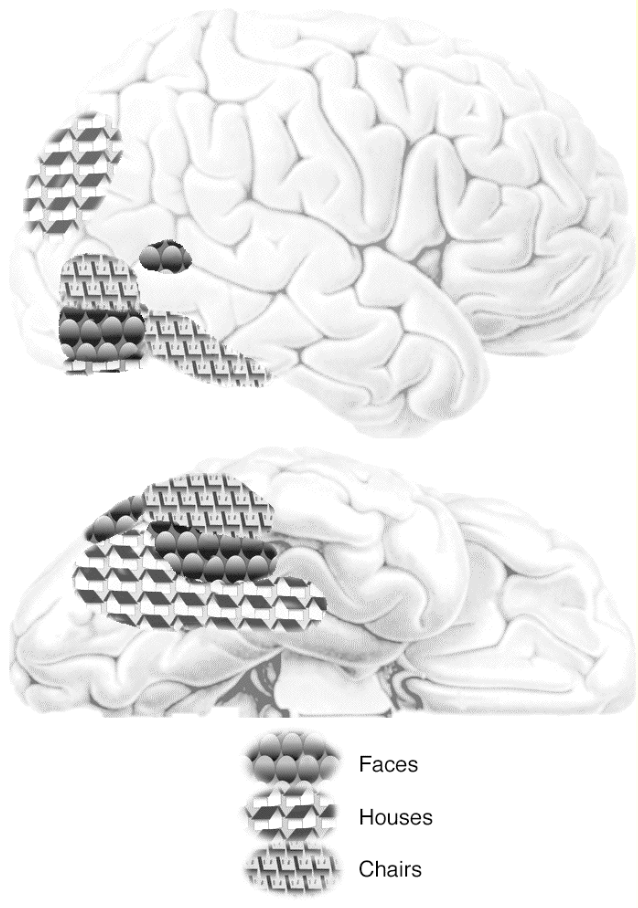
Model III:

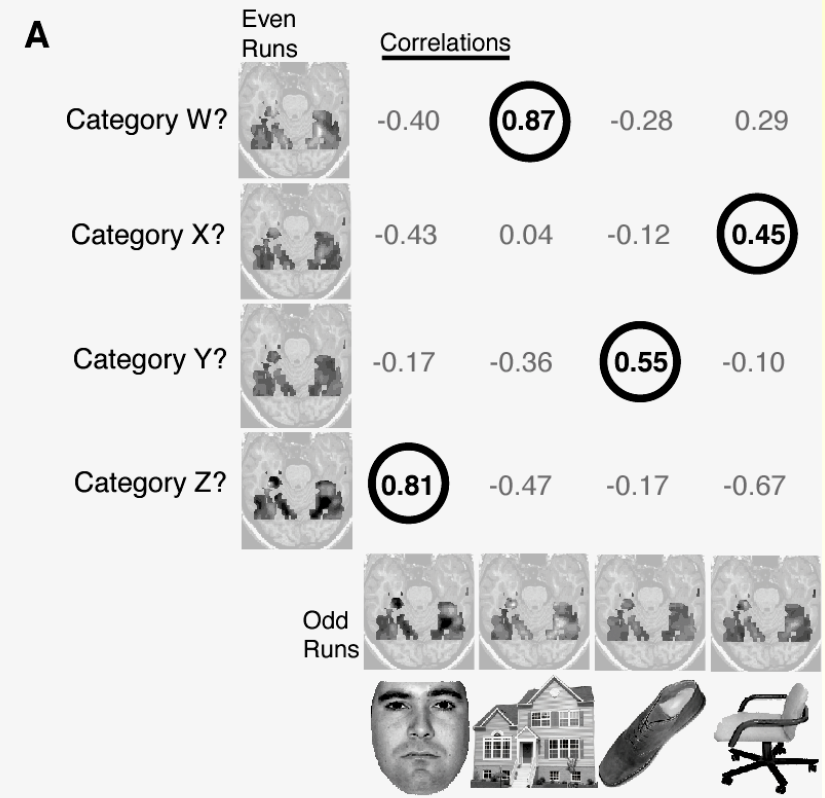
•representations of faces and different categories of objects are widely distributed and overlapping

"Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex," J. Haxby, M.I. Gobbini, M.L.Furey, A. Ishai, J.L. Schouten, and P. Pietrini, *Science*, Vol. 293, pp. 2425-430, 2001

Distributed Representations of Faces and Objects

(Haxby et al, 2003)





Locations of regions in occipitotemporal cortex that respond preferentially to faces, houses, and chairs

The category that a subject is viewing can be identified by the spatial pattern of response in ventral temporal cortex.

Still-Image Face Recognition

Face Recognition from Still Images

1. Face Detection and Feature Extraction
   - Summary
2. Recognition from Intensity Images
   - Categorization
   - Brief description of representative work
3. Summary/Discussion
   - Status of machine face recognition
   - Open research issues

Face Detection Summary

- Up to the mid 90's: single-face segmentation
  - Whole-face template
  - Deformable feature-based template
- More Recently
  - Image-/appearance-based methods:
    extensive training on large samples
  - Multiview-based detection of 3D rotated faces
- Improve detection systems
  - ↑ true positives (detection rate), ↓ false positives
  - Re-training on false positive samples
  - Real-time (Viola & Jones'01: AdaBoost)

Facial Feature Extraction Summary

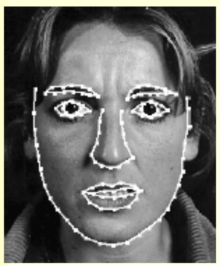
- Important for both detection and recognition
- Three types of approaches
  - Generic features: lines, curvatures, etc.
  - Feature templates: mouth, nose, eyes
  - Structural matching: e.g., ASM
- Challenging issues:
  - Distortion of features under large pose
  - “Restoration” of invisible features

Statistical Face Models

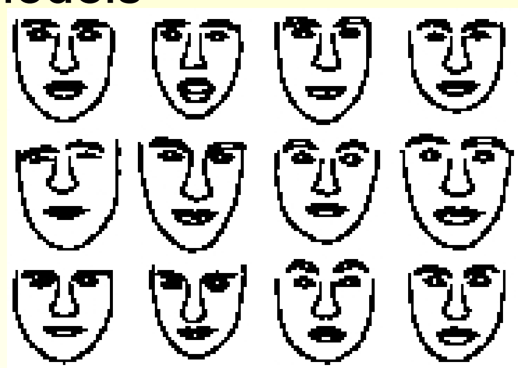
Active Shape Model/Active Appearance Model

(Cootes, Taylor, et. al.’95/00/01)

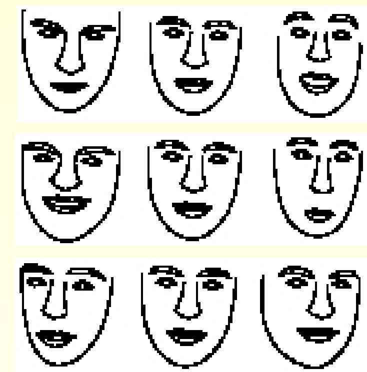
- Statistical Shape Models



Extracting shape

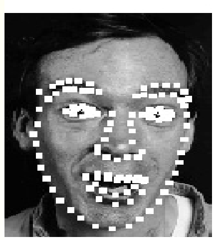


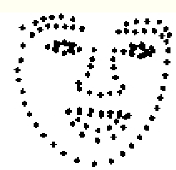
Example shapes from training set



Shape Synthesis

- Statistical Appearance (combined) Models





Set of Points

Shape Free Patch

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{P}_s \mathbf{b}_s$$

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g$$

$$\begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_g \end{bmatrix} \xrightarrow{\text{PCA}} \mathbf{c}$$

Active Statistical Face Models

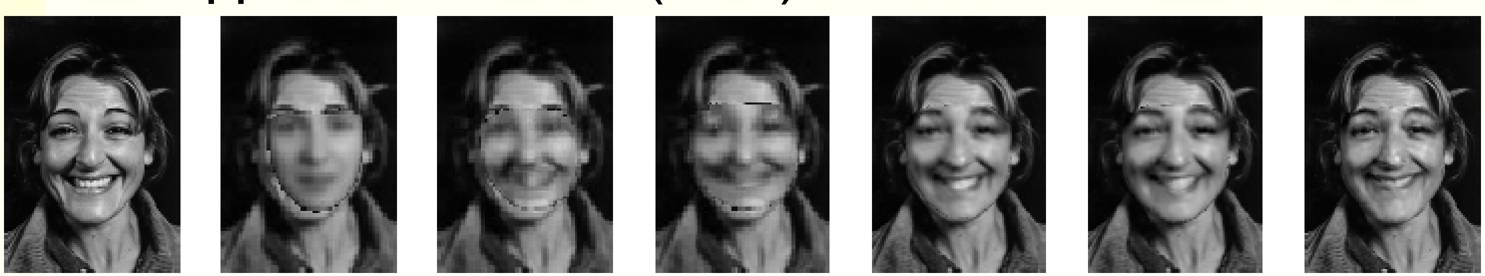
Searching & model fitting (active)

- Active Shape Model (ASM)

Normal example of search using ASM

Failure example of search using ASM

- Active Appearance Model (AAM)



original

Multi-resolution search using AAM

Recognition from Intensity Images

High-level Categorization

1. Holistic matching methods
   - Classification using whole face region
2. Feature-based (structural) matching methods
   - Structural classification using local features
3. Hybrid Methods
   - Using local features and whole face region

Still Face Recognition Systems

Approach	Representative Works
Holistic methods	
Principal Component Analysis (PCA)	
Eigenface	Direct application of PCA
Fisherface/Subspace LDA	FLD on eigenspace
SVM	Two-class problem based on SVM
ICA	ICA-based feature analysis
Other Representations	
LDA/FLA	FLD/LDA on raw images
PDBNN	Probabilistic decision based NN
Feature based methods	
Pure geometry methods	Earlier methods, recent methods
Dynamic Link Architecture	Graph matching methods
Convolution Neural Network	SOM learning based CNN methods
Hybrid methods	
Modular eigenface	Eigenface & eigenmodules
Hybrid LFA	Local & global feature method
Component-based	Face region and components

Holistic Approaches
PCA for face images

Why principal component analysis (PCA)?

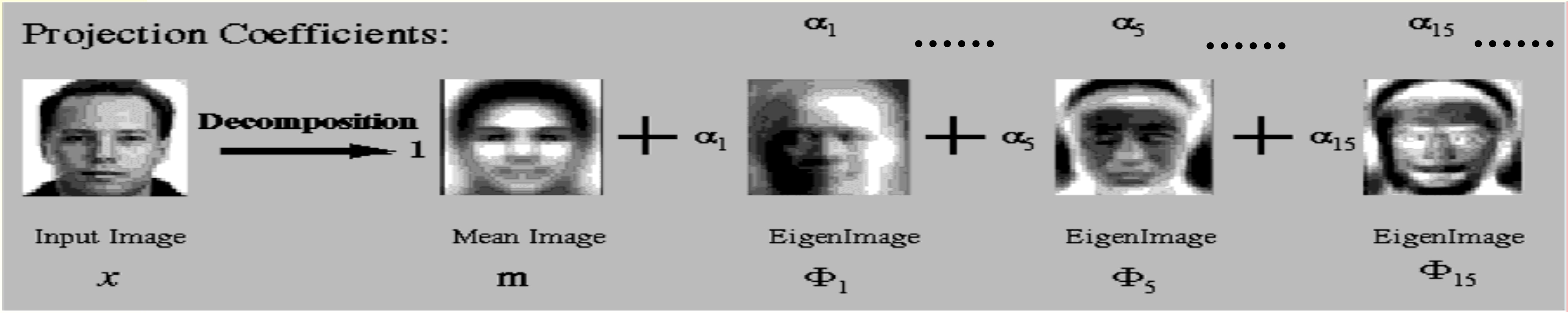
- **Statistical redundancy for natural images** (Ruderman 94)
  Especially for normalized face images
- **Reduced sensitivities to noise: geometric and photometric**

How to compute principal components?

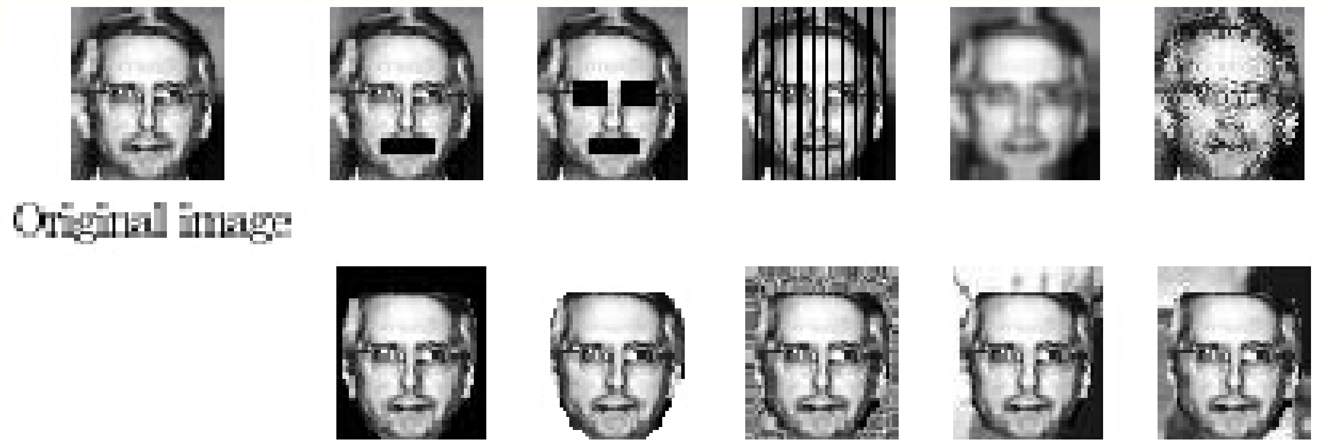
$$C\Phi = \Phi\Lambda$$

PCA reconstruction

$$\mathbf{x} \approx \mathbf{m} + \sum_{i=1}^{M} \alpha_i \Phi i$$



Holistic Approaches
PCA for face images



Electronically Modified images

Holistic Approaches
PCA for face images

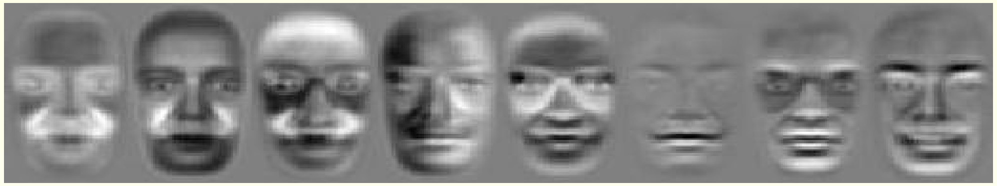
PCA reconstructed images

Holistic Approaches
Probabilistic eigenface

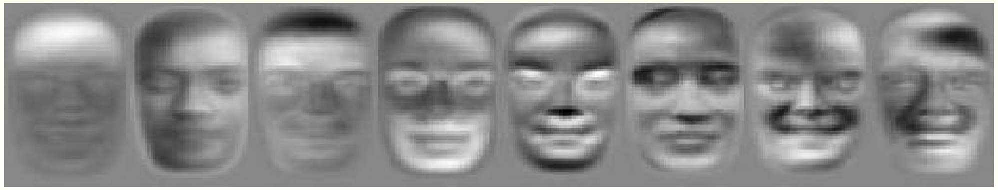
Probabilistic Eigenface (Moghaddam and Pentland'97)

- Making Bayesian practical by converting a multi-class problem into a two-class problem
- Intrapersonal class and extrapersonal class based on image difference $\Delta = I_1 - I_2$
- Performance: top 3 in the FERET test

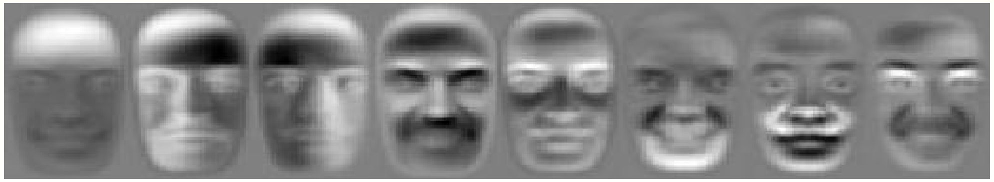
Intrapersonal



Extralpersonal



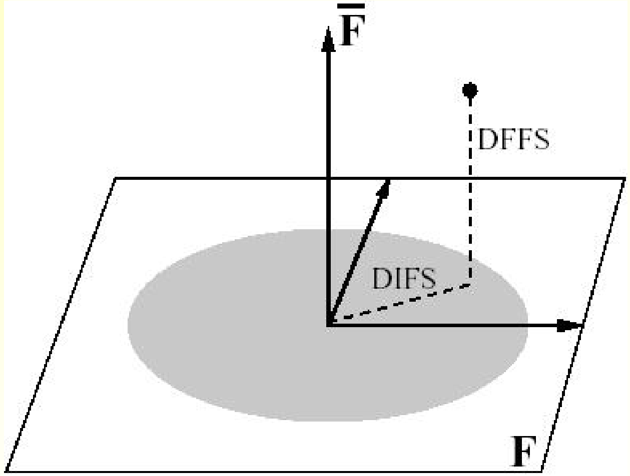
Standard

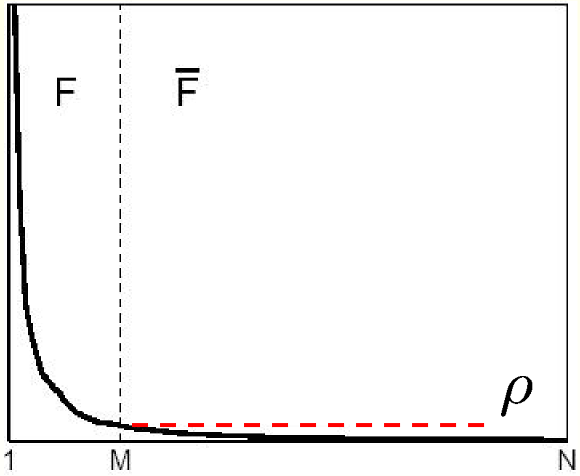


Holistic Approaches
Critical parameter estimation

Efficient Technique for Probability Estimation
(Moghaddam and Pentland'97, Sung ang Poggio'97)

- Decomposition into the principal subspace F and its orthogonal complement subspace $\bar{F}$
- Estimating covariance in F
- Estimating a stable average eigenvalue ρ in $\bar{F}$





Holistic Approaches
FLDA for Face Recognition

LDA Face Recognition

(Swets and Weng'96: PCA + LDA)
(Etemad and Chellappa'97: Linear discriminant analysis)
(Belhumeur, Hespanha and Kriegman'97: Fisherface)
(Zhao and Chellappa'98: Regularized subspace LDA)
(Liu and Wechsler'98: Enhanced FLDA)
(Chen et al'00: LDA with singular within-class scatter matrix)
(Yu and Yang'01: Direct LDA removing null space of between-class scatter matrix)

Why LDA?

- Improve classification performance when more than one samples are available per class

How to carry out LDA?

Objective function

$$J(W) = \frac{|W^T S_b W|}{|W^T S_w W|}$$

Solution

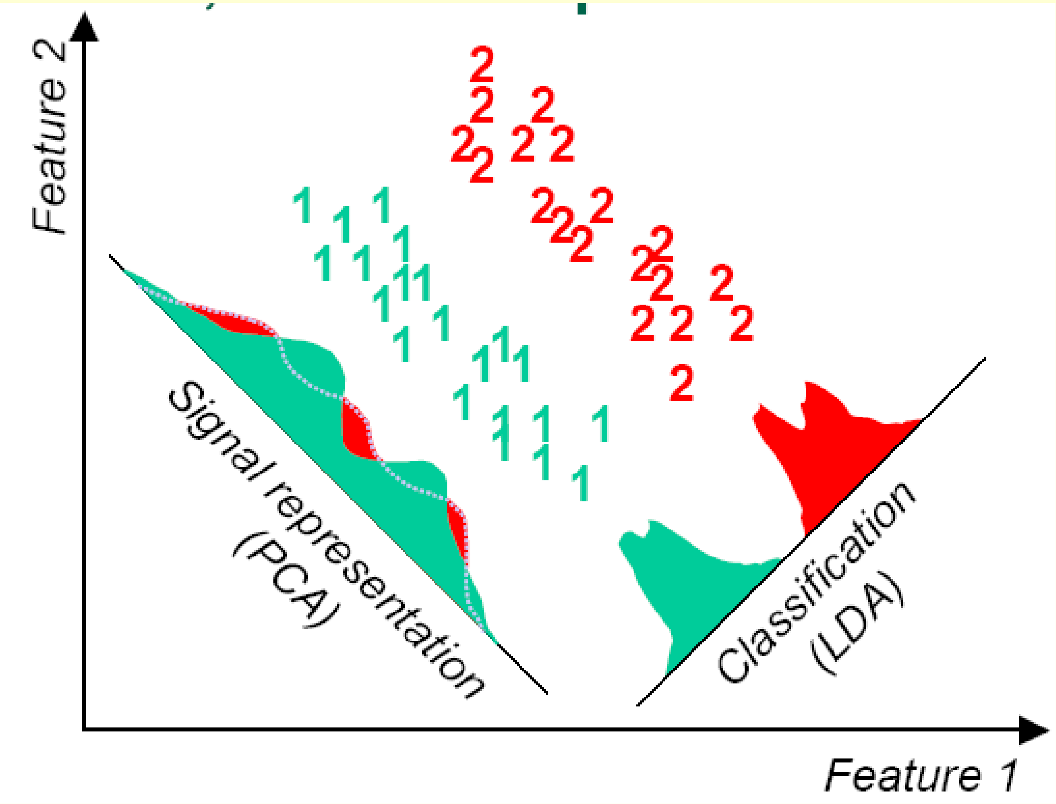
$$S_b W = S_w W \Lambda_L$$

$$\mathbf{z} = W^T \mathbf{x}$$

$$S_w = \sum_{i=1}^{K} Pr(\omega_i) C_i$$

$$S_b = \sum_{i=1}^{K} Pr(\omega_i)(\mathbf{m}_i - \mathbf{m}_0)(\mathbf{m}_i - \mathbf{m}_0)^T$$

Holistic Approaches
Why LDA for Classification?



Holistic Approaches
Regularized Subspace LDA

Regularized Subspace LDA (Zhao and Chellappa'98)

- Unique choice of a universal subspace dimension
- Weighted distance
- Regularized S_W matrix (two reasons) $\hat{S}_w = S_w + \delta I$
- Performance: top 3 in the FERET test

The Sequential Procedure

- Removing within-class variations (whitening)

$$S_w W_w = W_w \Lambda_w \Rightarrow \mathbf{y} = \Lambda_w^{-1/2} W_w^T \mathbf{x}$$

- Maximizing class separation (eigenvalue)

$$S_b^y W_b = W_b \Lambda_b \Leftarrow S_b^y = \sum_{i=1}^K Pr(\omega_i)(\mathbf{m}_i^y - \mathbf{m}_0^y)(\mathbf{m}_i^y - \mathbf{m}_0^y)^T$$

- The final mapping matrix (not orthogonal)

$$W = W_w \Lambda_w^{-1/2} W_b \Leftarrow \mathbf{z} = W_b^T \mathbf{y} = W^T \mathbf{x}$$

(Fugana'91)

Holistic Approaches
Global Face Dimension

The use of a face subspace is not just for dimensionality reduction!

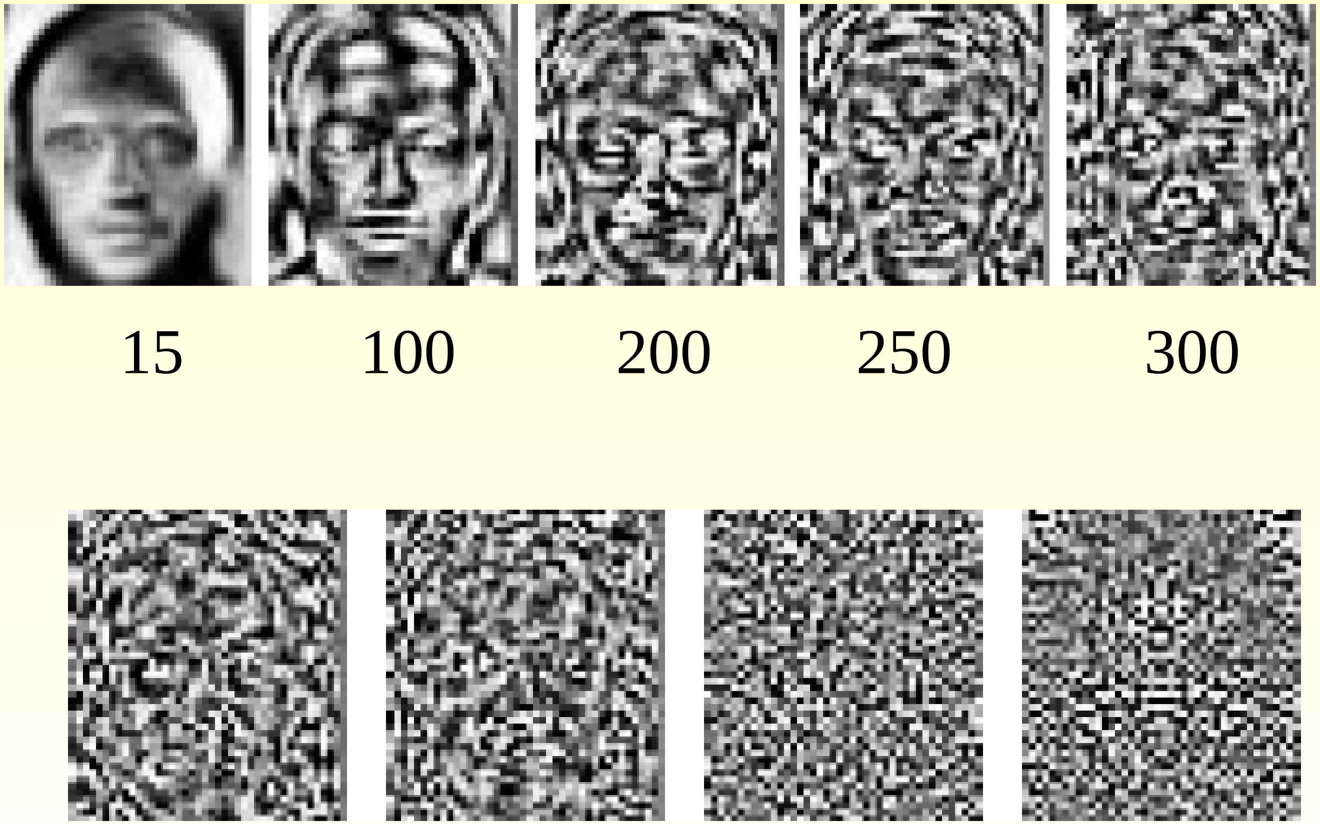
A fixed universal subspace 300 is chosen based on the characteristics of eigenimages based on 1078 FERET images. (It was reported to be around 400 based on 5000 images. Penev and Sirovich, AFGR'00)

One interesting implication: image size could be any size as long as it is larger then 300.

Dim/Size	*96x 48*	*48 x 42*	*24 x 21*	*19 x 17*	*12 x 11*
132	x	x	x	x	84.34
200	68.69	71.30	70.43	76.52	x
300	**81.74**	**85.21**	**86.08**	**90.43**	x
400	68.69	71.30	69.56	x	x
500	67.83	66.96	71.30	x	x

Experimental Verification

Holistic Approaches
Global Face Dimension



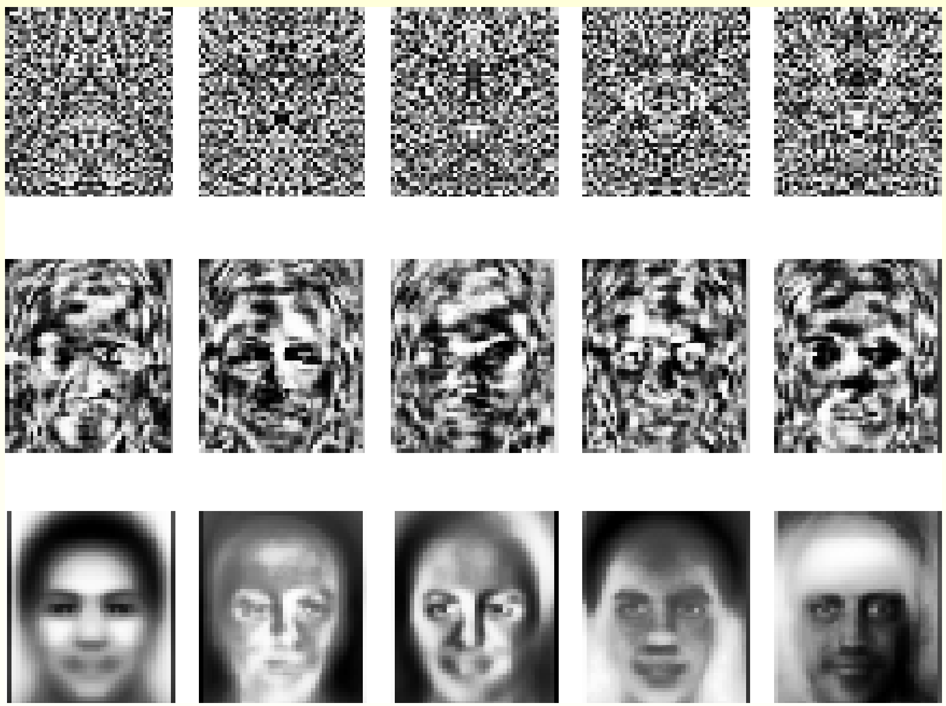
15 100 200 250 300

400 450 1000 2000

Holistic Approaches

Linear Subspace Projection Algorithms

$$\mathbf{z} = \mathbf{P}_{roj}^{T}\mathbf{x}$$



$\mathbf{P}_{roj} = W$ LDA Bases

$\mathbf{P}_{roj} = \Phi W$ Subspace LDA Bases

$\mathbf{P}_{roj} = \Phi$ PCA Bases

Holistic Approaches
LDA Variants for Face Recognition

Handling small size sample: the null space of S_w contains important discriminant information

- LDA in the null space of S_w only (Chen et al 00)

$$S_w W_w^{null} = 0 \qquad \mathbf{y} = {W_w^{null}}^T \mathbf{x}$$

- Efficient direct LDA (Yu and Yang'01): Removing null space of S_b

$$S_b W_b = W_b \Lambda_b \longrightarrow \mathbf{y} = \Lambda_b^{-1/2} W_b^T \mathbf{x} \longrightarrow S_w^y W_w = W_w \Lambda_w \longrightarrow \mathbf{z} = W_w^T \mathbf{y}$$

- Augment S_w (Zhao and Chellappa'98, Hong and Yang'91):

$\widehat{S}_w = S_w + \delta I$ OR replacing zero eigenvalue with a small value

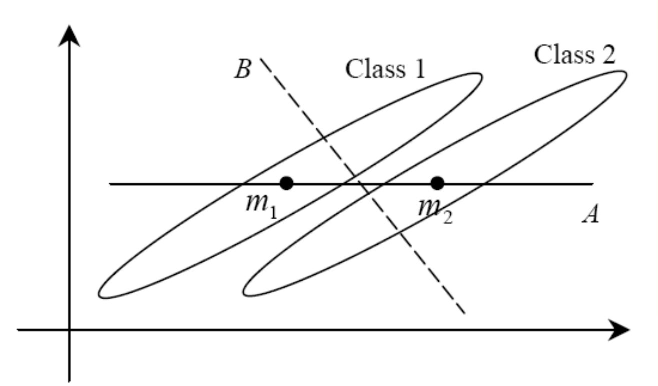
Handling small number of classes: taking full information

- Discriminant Component Analysis (Zhao'00,Okada and Tomita'85)
  - Starting from the full signal space, we first select the optimal linear vector/basis W_1 based on LDA criterion. Then we project the signal into the complementary subspace and select the optimal linear vector in that subspace. Repeat this until full bases are constructed.

Handling non-linearity: kernel methods

Holistic Approaches
What matters for LDA Face Recognition?

1. Face subspace is important for extracting stable features; hence, better generalization capability.
2. Pay attention to details:
   - How to compute matrices and solve eigenvalues?
3. Understand the limitations of different algorithms



Information lost in Direct LDA

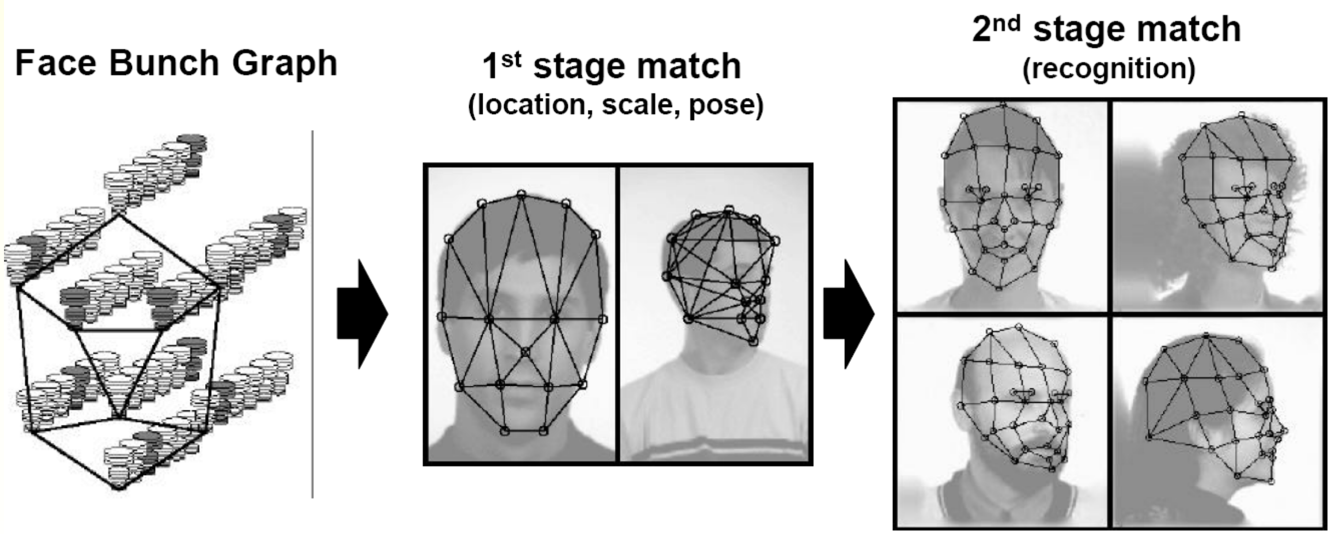
Feature based Approaches
DLA, EGM and EBGM

DLA, EGM, EBGM (Lades, Wiskott, and Malsburg, etc.'93,97,98,99)

- Comparing jets: $$\mathcal{S}_\phi(\mathcal{J},\mathcal{J}') = \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j - \vec{d}\vec{k}_j)}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}}$$
- Graph based on local feature matching & global structure

$$S_\mathcal{B}(\mathcal{G}^I,\mathcal{B}) = \frac{1}{N}\sum_n \max_m(\mathcal{S}_\phi(\mathcal{J}_n^I,\mathcal{J}_n^{\mathcal{B}_m})) - \frac{\lambda}{E}\sum_e \frac{(\Delta\vec{x}_e^I - \Delta\vec{x}_e^\mathcal{B})^2}{(\Delta\vec{x}_e^\mathcal{B})^2}$$ where $\Delta\vec{x}_e = \vec{x}_n - \vec{x}_{n'}$

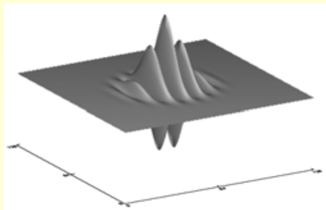
- Recognition: $$S_\mathcal{G}(\mathcal{G}^I,\mathcal{G}^M) = \frac{1}{N'}\sum_{n'} \mathcal{S}_a(\mathcal{J}_{n'}^I,\mathcal{J}_{n'}^\mathcal{M})$$
- Performance: top 3 in the FERET test



Feature based Approaches

Gabor wavelets and image graphs

Gabor wavelets: biologically motivated convolution kernels in the shape of plane waves restricted by a Gaussian envelope function (Daugman'88)



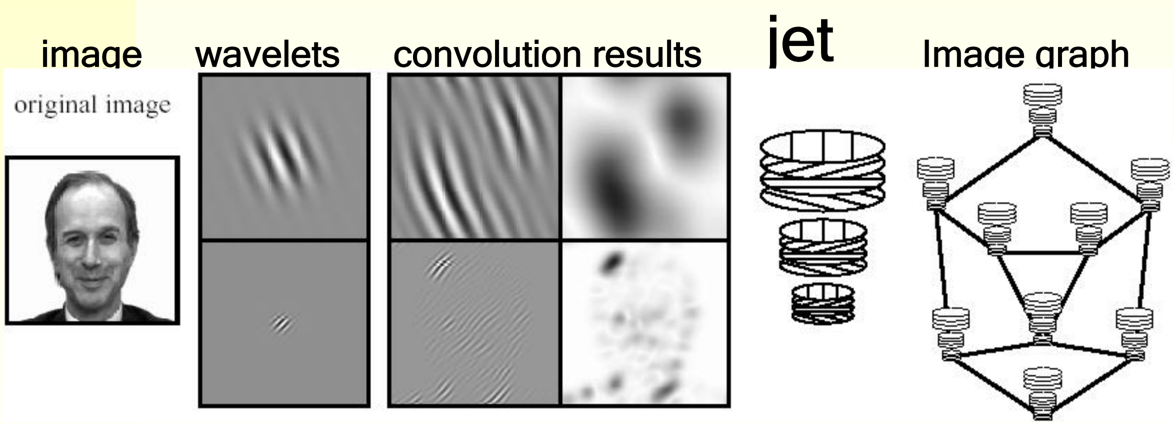
$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} \exp\left(-\frac{k^2 x^2}{2\sigma^2}\right)\left[\exp(i\vec{k}\vec{x}) - \exp(-\sigma^2/2)\right]$$

$$(\sigma = 2\pi)$$

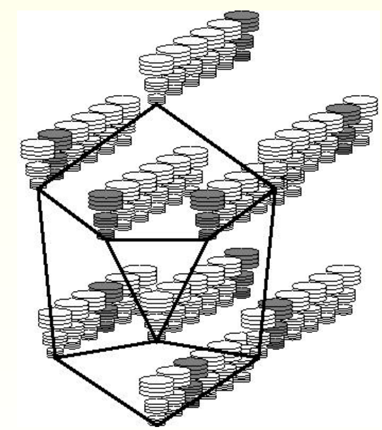
Jets $\{J_i\}$: $J_i(\vec{x}) = \int I(\vec{x}')\psi_j(\vec{x}' - \vec{x})d\vec{x}' = a_j \exp(i\phi_j)$

wave vectors: $\vec{k}_j = \begin{pmatrix} k_\nu \cos\theta_\mu \\ k_\nu \sin\theta_\mu \end{pmatrix}, k_\nu = 2^{-\frac{\nu+2}{2}}, \theta_\mu = \mu\frac{\pi}{8}$

$$(\nu = 0, \cdots, 4, \mu = 0, \cdots, 7, j = \mu + 8\nu)$$



Graph representation



Stack-like bunch graph

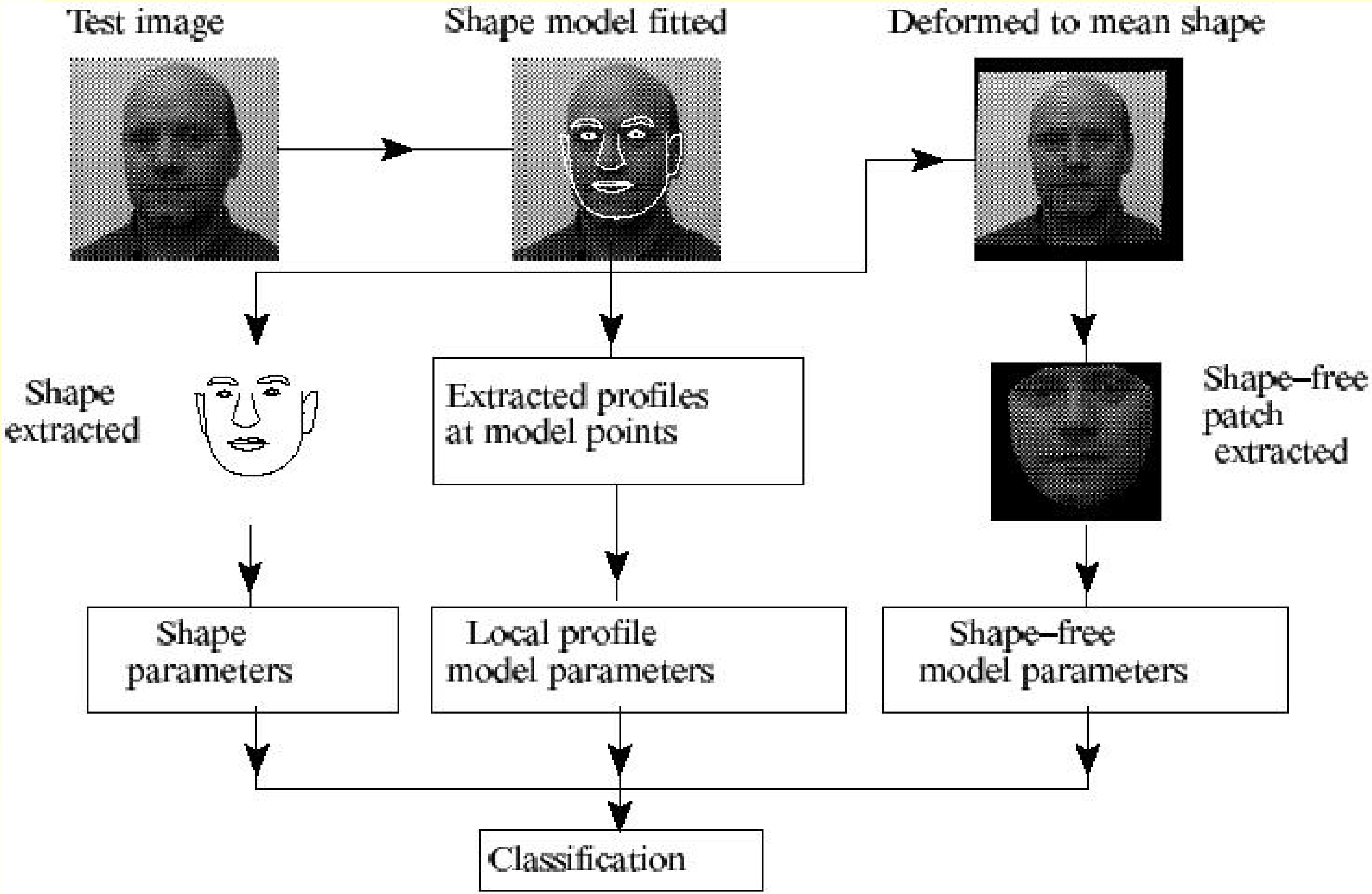
Hybrid Approaches
Flexible appearance models

FAM based method (Lanitis, Taylor, and Cootes'95)

- Use of shape-free texture information
  - Applying PCA to normalized texture
- Use of Discriminative Active Shape Model (ASM)
  - Separating shape variations due to inter-class from those due to intra-class
- Use of local profiles perpendicular to shape boundary
  - Enhancing robustness of the system against local changes
- Performance

  92% for 10 normal images, 48% for 3 difficult images
  (Each of 30 individuals: training on 10 and testing 13 images)

Hybrid Approaches
Flexible appearance models

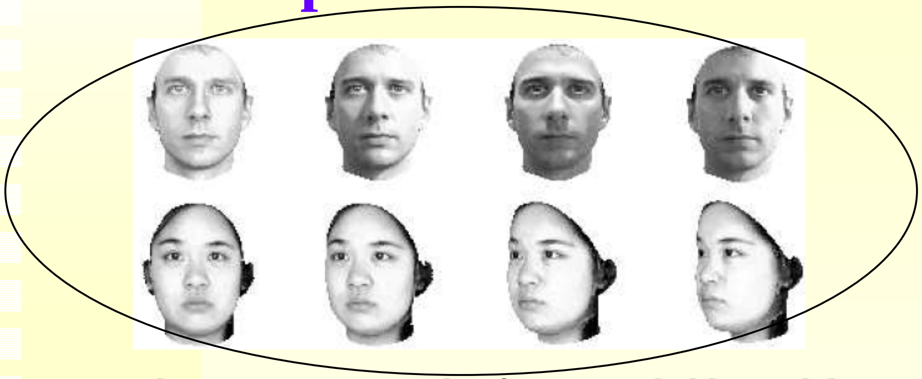


Hybrid Approaches

Component-based methods

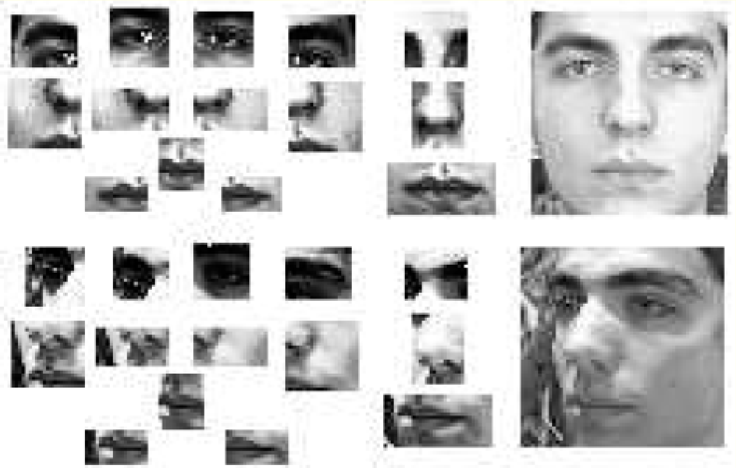
Component-based face recognition with morphable models

(Huang, Heisele, and Blanz'03)

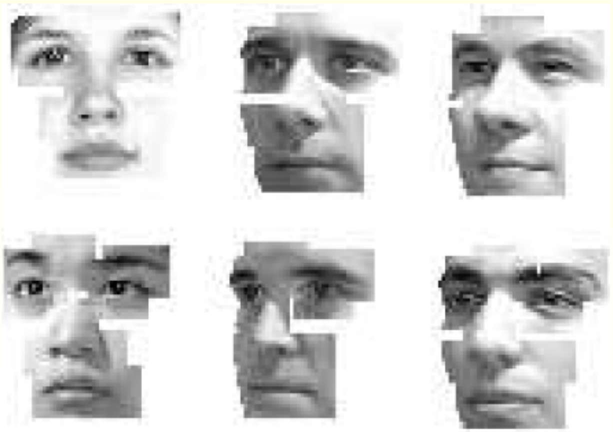


Synthetic training samples from morphable models under different lightings and poses

Component-based methods require large number of samples



Fourteen components from face detection



Nine components + whole face for face detection

Classifiers are SVM

Performance: 90% hybrid method vs 10% global method based on 6 subjects (3 training, 200 testing)

Still Face Recognition: Facts/Lessons

1. Reached a significant level but still far away from the capability of human perception
2. Take advantage of domain knowledge
3. How big should be the face image size: 128 or 12?
4. Accurate feature localizations are critical
5. Face recognition is probably not unique compared to other object recognition (psychology ← → engineering)
6. Comparing different systems is important but difficult
7. Choose appropriate systems based on particular applications
8. Building blocks for video-based methods

Still Face Recognition: Open Issues

1. Addressing the issue of recognition being too sensitive to inaccurate facial feature localization

2. Robustly recognizing *faces*
   - Small and/or noisy images
   - Images acquired years apart
   - Outdoor acquisition: lighting, pose

3. What's the limit on the number of faces that can be distinguished?

4. What's the principal and optimal way to arbitrate and combine local features and global features

Video-based Face Recognition

Video Based Face Recognition

1. Challenges of video-based face recognition

2. Basic techniques employed in video-based face recognition

3. Video-based face recognition methods
   - 3D recognition will be discussed separately

4. Summary/Discussion

Video Based: Challenges

1. Quality of video is low
   - Under non-ideal acquisition condition
   - Objects are not co-operative
2. Face images are too small
   - Not suitable for many methods, e.g., local feature methods
3. Characteristics of faces
   - Relatively easy for detection, but hard for recognition

Video Based: Basic Techniques

1. Face Segmentation & Pose Estimation
   - Use motion/color to speed up
   - Learning for detection and pose estimation

2. Face and Feature Tracking
   - Head tracking
   - Facial feature tracking
   - Complete tracking

3. Face Modeling (important!)
   - Shape modeling and texture modeling
   - Tracking plus bundle adjustment
   - Multi-view 2D modeling, deformable 3D modeling and direct 3D modeling

Video-based Face Recognition

Categorization

1. Still-image methods
   - Basic methods
   - Tracking-enhanced (voting schemes, depth for detecting head and/or virtual view synthesis)
2. Multi-modal methods
   - Video- and Audio-based
   - Face- and Gait-based
   - Face- and Iris-based
3. Spatio-temporal methods
   - Feature trajectory based
   - Video-Video based

Still Image Methods
Tracking plus appearance model

Tracking + appearance model (Edwards, Taylor, and Cootes'98)

- Building an appearance model for each image
- LDA decomposition of the model parameters $\mathbf{c}$
 into the identity subspace and residual subspace (light, expression,pose)

  LDA projection $\mathbf{c} = \bar{\mathbf{c}} + \mathbf{Dd} + \mathbf{Rr}$

- Class-specific linear correction $\mathbf{A}_j$ to the results of global LDA

  $\mathbf{d}_t$ is the true proj.: $\mathbf{d} - \mathbf{d}_t = \mathbf{A}_j(\mathbf{r} - \bar{\mathbf{r}})$

- Performance: enhanced face tracking and visual enhancement are demonstrated, but no recognition results

Varying the most significant identity parameters $\mathbf{d}$

Varying residual parameters without affecting identity

Multi-modal Methods
An ATM application

Multi-modal person recognition system

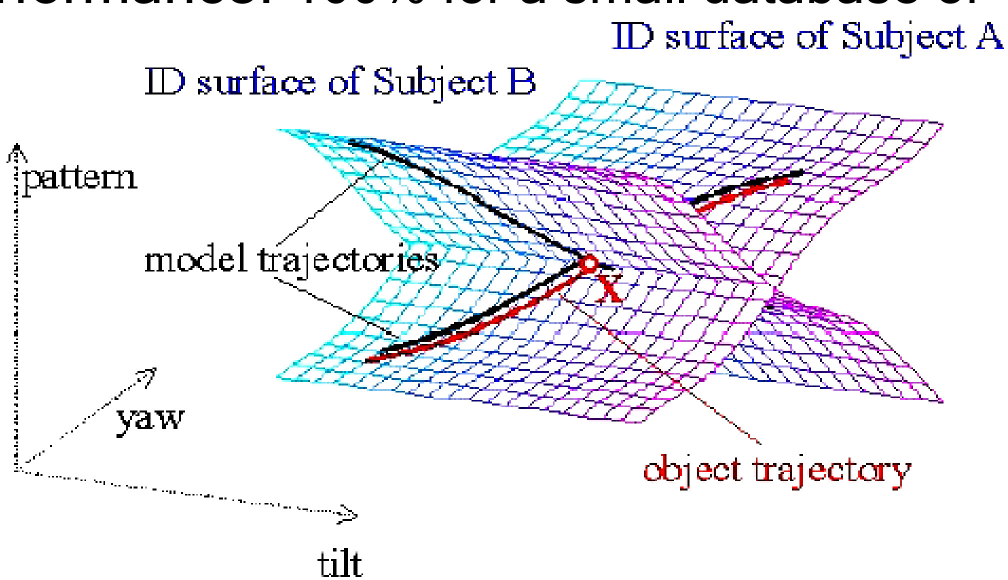
(Choudhury, Clarkson, Jebara, and Pentland'99)

- Three modules:
  - Face recognition (FR) module
  - Speaker identification (SI) module
  - Classifier fusion module
- Characteristics:
  - FR module can detect and compensate for pose variation based on feature tracking
  - Select reliable video and audio clips for recognition
  - Use SfM to detect the presence of actual people as opposed to face images
- Performance: 100% for a small database of 26 people

Spatio-temporal Methods
Identity surface method

Identity surface method (Li, Gong, and Liddell'01)

- Identity surface
  - Yaw and tilt axes: head pose
  - Other axes: discriminating 3D geometry and shape-free-texture features obtained from PCA+KDA
- Matching trajectories (synthesized for known subjects based on estimated poses) in the identity surface
- Performance: 100% for a small database of 12 subjects



Spatio-temporal Methods
Video-to-video method

Exemplar-based probabilistic method (Krueger,Zhou,and Chellappa'02)

- Representing gallery video using a online algorithm to learn exemplars as mixture centers (6-20 ones out of 300 frame video)
- Match gallery video against probe video via a condensation method for computing *Posteriors* of identity and affine motion

$$p(i_t|Z_1, \cdots, Z_t) = \int_{\alpha_t} p(\alpha_t, i_t|Z_1, \cdots, Z_t) \equiv p_t(\alpha_t, i_t)$$

- Performance: range from 88% to 100%
  (25 individuals, 4 videos/individual, one gallery, three as probe)

Success: sample frames 1, 9, 40, 72

Failure: sample frames 1, 35, 81, 100

Video-based Face Recognition: Summary

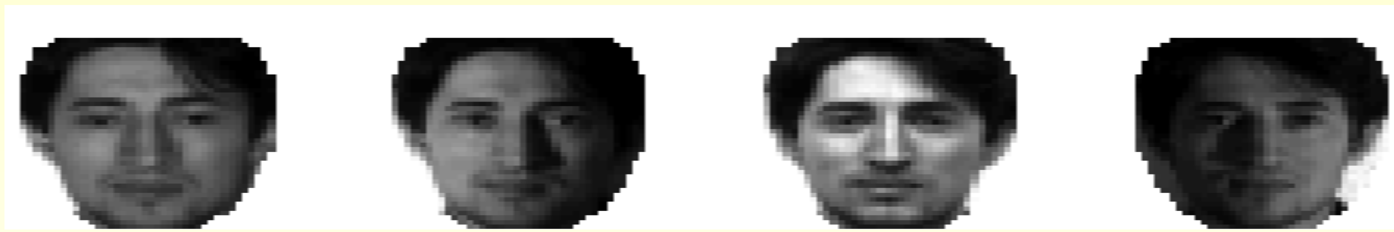
1. Good news:
   - Psychology study suggests motion helps recognition.
   - Additional temporal information
     - Select or reconstruct good still frame
     - Multi-view images, recovered 3D information

2. Issues:
   - Very small testing database (e.g., 20 subjects)
   - How to better use joint spatial and temporal information?
   - 3D recovery accurate enough for recognition?
   - Demand huge space for storing videos

Illumination & Pose Problem

Illumination and Pose Problem

Face I



Face II




A Simple Reflectance Model

View-independent Lambertian Model

$$I[x],y] = \rho[x,y]\mathbf{n}[\mathbf{x},\mathbf{y}]^T\mathbf{s}$$

A Simple Reflectance Model

Varying Albedo Lambertian Model

Varying Albedo

$$I = \rho \frac{1 + pP_s + qQ_s}{\sqrt{1+p^2+q^2}\sqrt{1+P_s^2+Q_s^2}}$$

Light Source

Surface Normal

$$P_s = \tan\alpha \cos\tau \qquad p = \partial z / \partial x$$

$$Q_s = \tan\alpha \sin\tau \qquad q = \partial z / \partial y$$

A Simple Reflectance Model

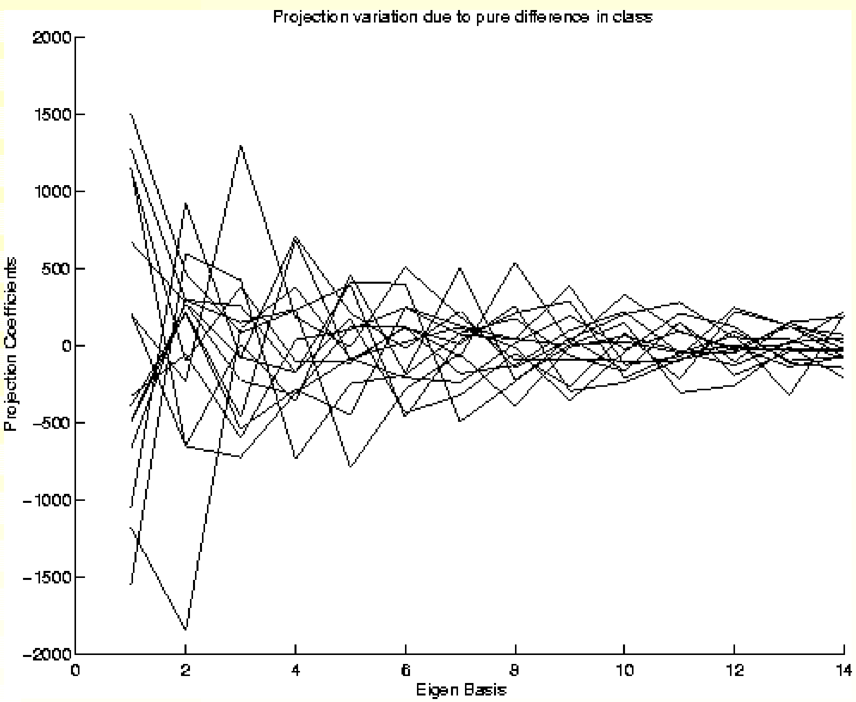
How illumination change affect recognition?

$$a_i = I_p \odot \Phi_i - I_A \odot \Phi_i$$
$$\tilde{a}_i = \tilde{I} \odot \Phi_i - I_A \odot \Phi_i,$$

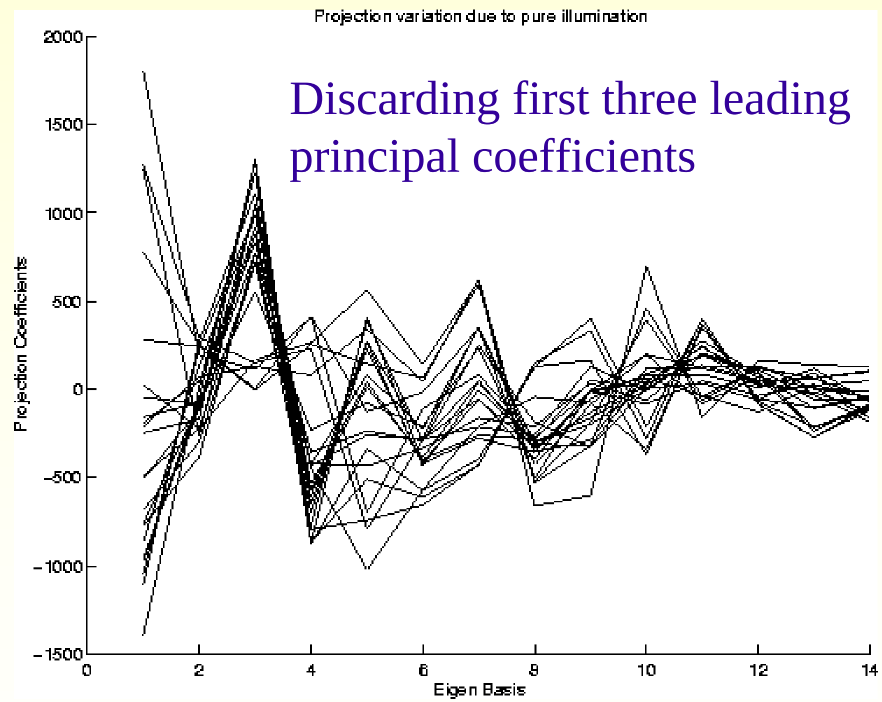
+ bilaternal symmetric image (Zhao, 1999)

$$\tilde{\mathbf{a}} = (\frac{1}{K}\mathbf{a}) + \frac{Q_s}{K}[f_1^a, f_2^a, \cdots, f_m^a]^T - \frac{K-1}{K}\mathbf{a}_A.$$

where $f_i^a = 2(I_p^L[x,y]q^L[x,y]) \odot \Phi_i^L[x,y]$



Class-only Variation

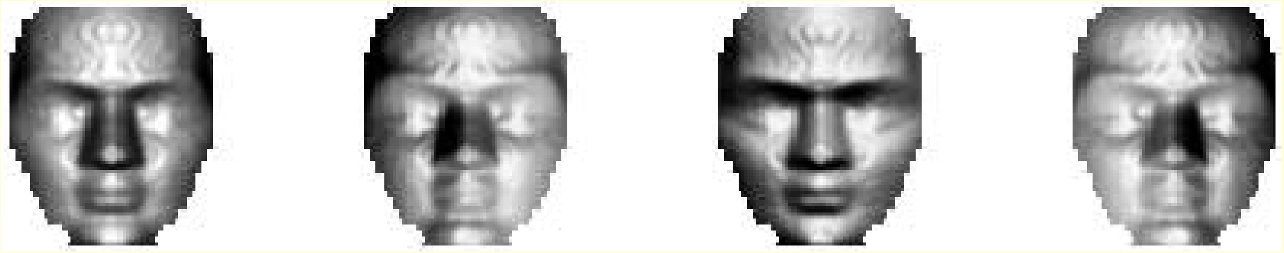


Illumination-only Variation

A Simple Reflectance Model

Image synthesis: constant vs. arbitrary albedo

Constant Albedo



Arbitrary Albedo

Arbitrary Albedo & Rotation

Solving Illumination Problem

1. Heuristic approaches
   - Disregarding leading principal components
   - Use of frontal-face symmetry
2. Image comparison approaches
   - Different image representations (edges, filtered, image-ratios)
   - Different distance measures
3. Class-based approaches
   - Using multiple images per class (at least during training)
4. Model-based approaches
   - Employing 3D face model

Class-based Approaches
Quotient Image

Quotient Image (Riklin-Raviv and Shashua'99)

- 3D linear illumination subspace $s = x_1 s_1 + x_2 s_2 + x_3 s_3$
- Basic assumption: same shape but different texture
- Q is the ratio of albedo functions given two objects
- Using Q and a training/bootstrap set to synthesize images of a novel object under various lighting
- New cost function for better result

$$(y_s - \sum_{i=1}^{N} \alpha_i A_i x)^2 \qquad \sum_{i=1}^{N} (\alpha_i y_s - A_i x)^2$$

Original images

Quotient images

Image synthesis

Class-based Approaches
Albedo-integrated normal map

Illumination Invariant Method

(Sim and Kanade'01, Zhou and Chellappa'03)

- Lambertian assumption

$$I = \rho \mathbf{n}^T \mathbf{s}$$

- Training images under different lightings

$$\mathbf{b}^T = \rho \mathbf{n}^T \mathbf{S} = \mathbf{r}^T \mathbf{S}$$

- Albedo-integrated normal map

$$\mathbf{r}^T = \mathbf{b}^T \mathbf{S}^+$$

- Linear generalization from the training set to the gallery and probe sets via *rank constraints* on the albedo and surface normal

Model-based Approaches
Direct 2D-2D method

Direct 2D-2D method (Zhao and Chellappa'00)

- Basic assumption: same & symmetric shape but different texture
- Obtain the prototype (frontal view + normal illumination) image directly from the single input image

Input | Local SFS | Direct Method | Original

Model-based Approaches
Spherical Harmonics Method

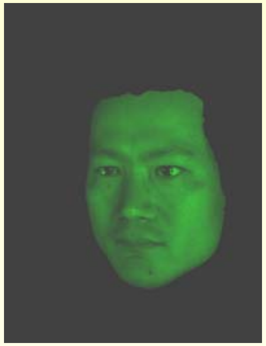
Second-order Spherical Harmonics Method
(Basri and Jacobs'01)

- The set of all reflectance functions can be approximated using spherical harmonics expansion
- Decompose input image into harmonic images which are illuminated by *harmonic light*
- The 3D linear illumination subspace methods are just first-order harmonic approximation without the DC components.

Forming Harmonic images

$$b_{nm}(p) = \lambda r_{nm}(X, Y, Z)$$

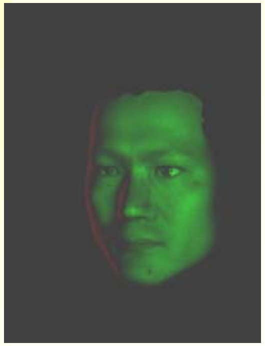
λ



λZ

λX

λY



$2\lambda(Z^2 - X^2 - Y^2)$

$\lambda(X^2 - Y^2)$

λXY

λXZ

λYZ

Solving Pose (and Illumination) Problem

1. Multi-view approaches
   - Multi-view database images of each person
2. Hybrid approaches
   - Multi-view images available only during training
3. Single image based approaches
   - Invariant feature based methods
   - 3D model based methods

Multi-view Approaches
Extending Illumination Cone

Illumination cone (Belhumeur and Kriegman'97)

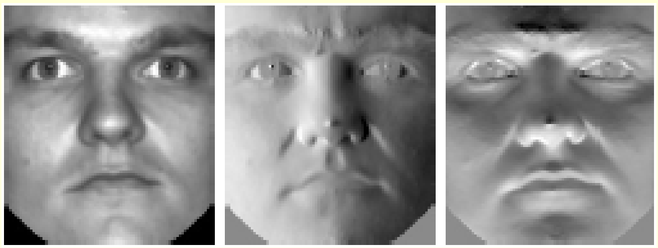
- Based on 3D linear illumination subspace
- Extending the linear subspace to convex representation to include attached shadow $b = \max(B\mathbf{s}, 0)$

Extending illumination cone for pose/illumination problem

(Georghiades, Belhumeur and Kriegman'01)

- Resolve GBR ambiguity to recover 3D Euclidean shape using domain knowledge (symmetry, chin & forehead, height)
- Originally, build a illumination cone for each pose
- Speed up by sub-sampling illumination cone and approximating it with linear subspace

Training images



Base images

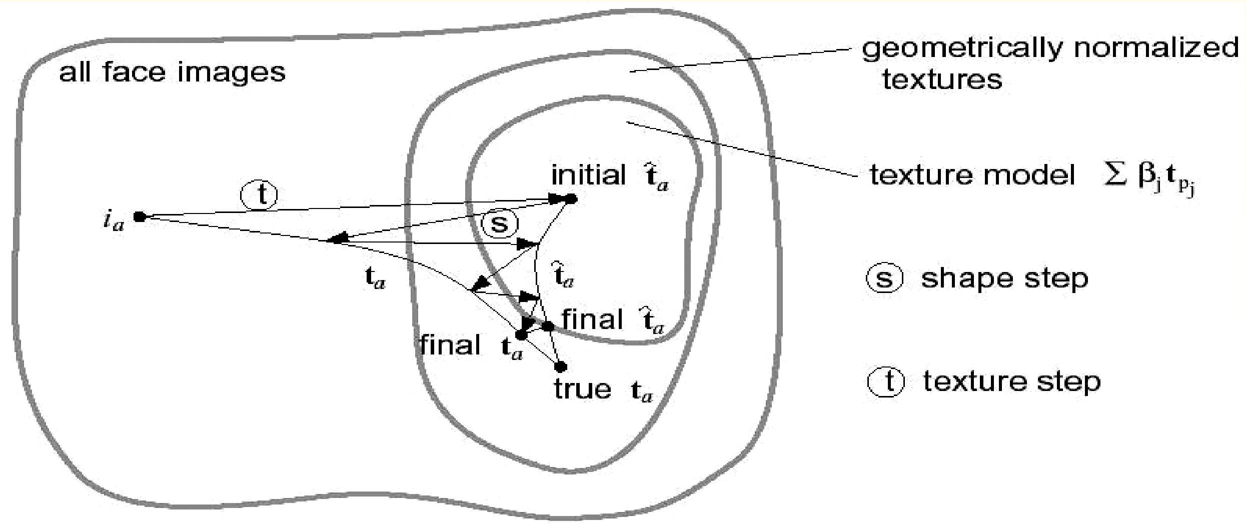
Synthesized images under different illumination

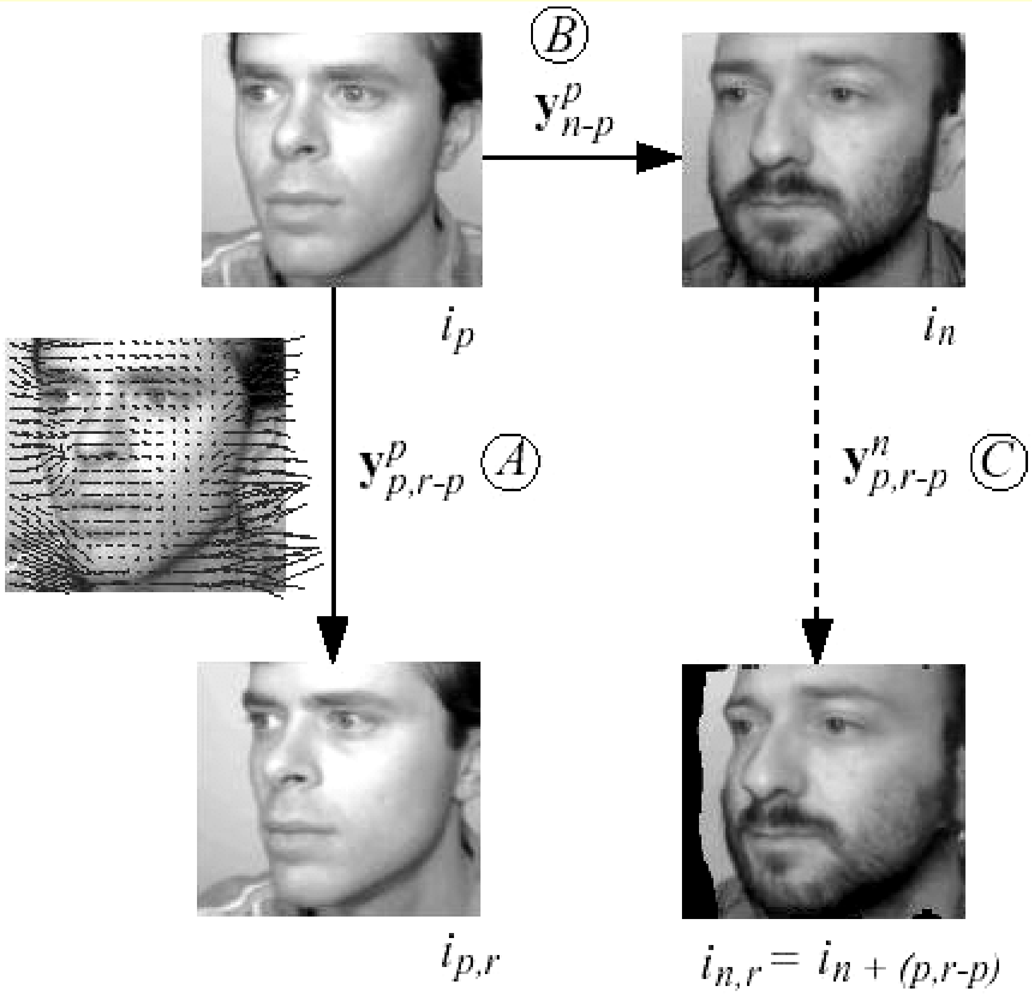
Synthesized images under different poses

Hybrid/Class Approaches
Vectorized Representation of Images

Vectorized representation (similar to AAM) (Beymer'95)

- A vectorized representation at each pose consists of both shape and texture that are mapped into the reference shape
- For a new image, a vectorization procedure is invoked that iterates between shape step and texture step



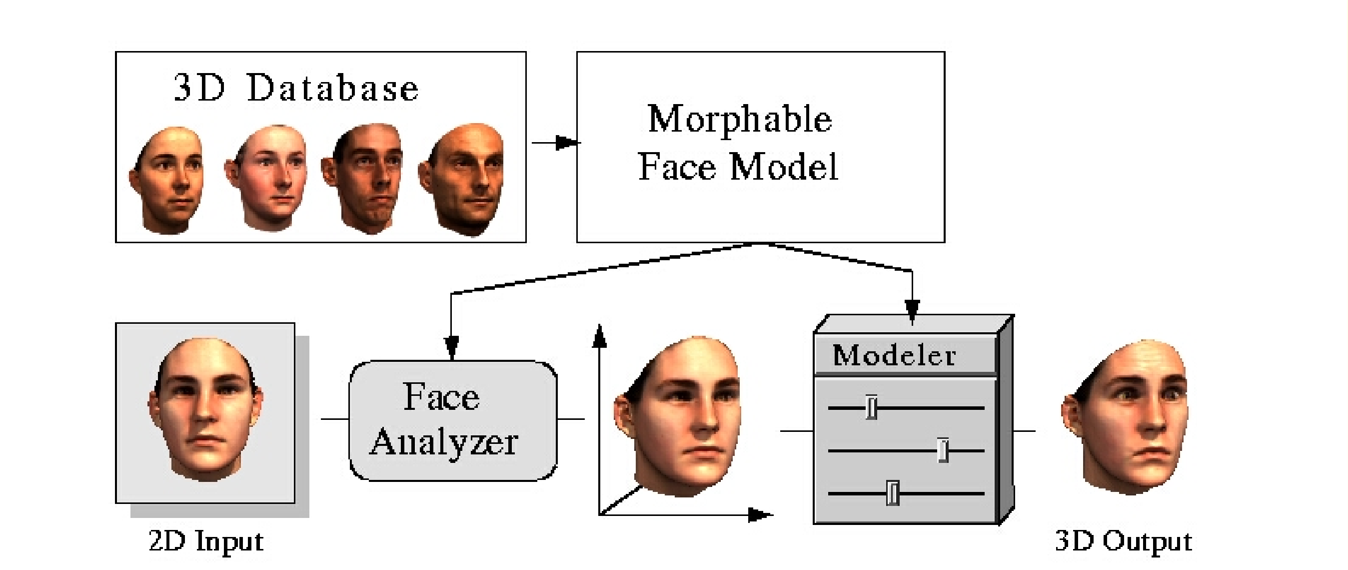


Parallel deformation

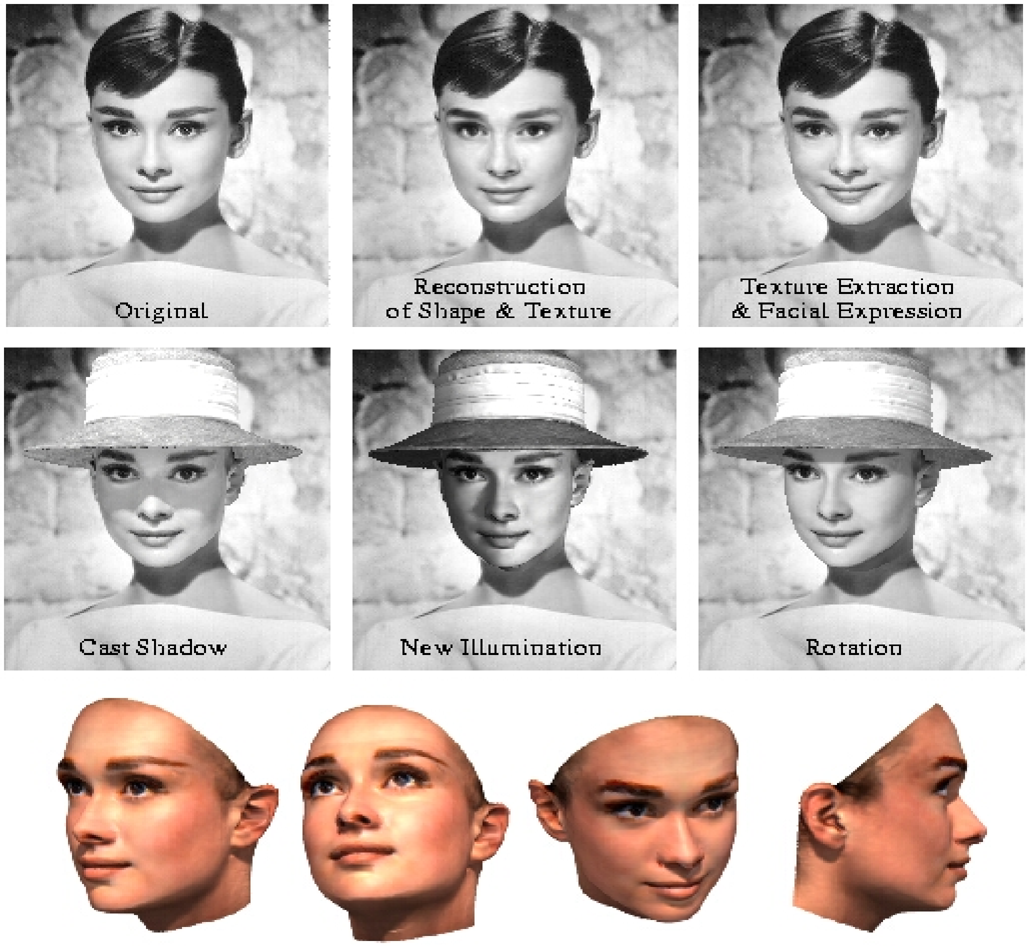
Hybrid/Class Approaches
Linear 3D Class Methods

Linear 3D Class Methods (Vetter and Poggio'97, Blanz and Vetter'99)

- Similar to AAM and vectorized representation
- After manual initialization, align a novel 2D image to a morphable 3D model learnt from a set of training samples



See the second part of this tutorial



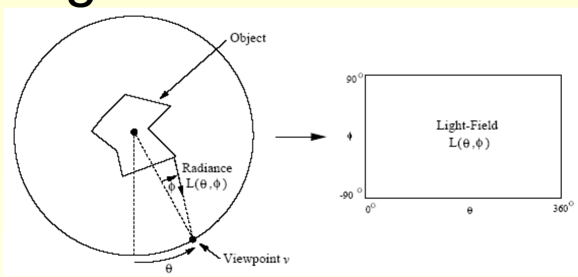
Recovered 3D shape and synthesized images

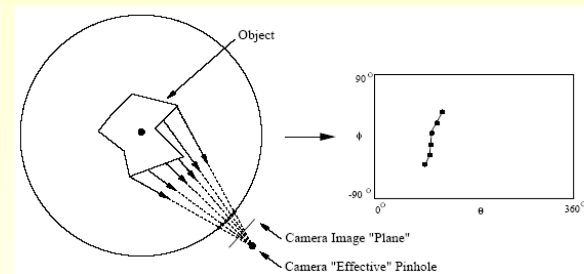
Hybrid/Class Approaches
Eigenfield Methods

Eigenfield Methods (R. Gross, I. Matthews, and S. Baker, 2002)

Obtaining a universal subspace

- Light fields





- Eigen Light-Fields (occlusion issue)

$$L(\theta,\phi) \approx \sum_i^d \lambda_i E_i(\theta,\phi)$$

Pose variation in CMU PIE database

Single Image Approaches
Front-view synthesis

Front-view synthesis **(Zhao and Chellappa'99/00)**

- An interesting question to ask:
    - Considering illumination model, can we directly apply optimal discriminant mapping learnt from frontal images to warped face images?

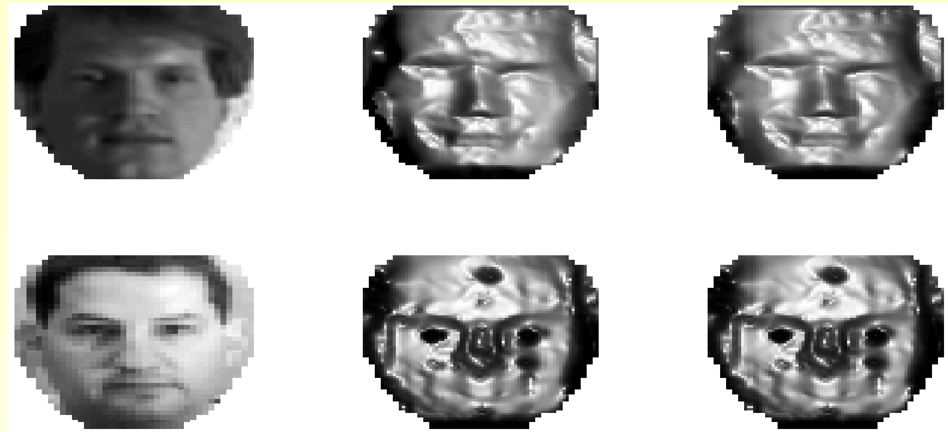
- How to use 3D information?
    - Recovery of 3D shape: SFS and symmetric SFS
        - Elegant but difficult to obtain accurate 3D shape
    - Or use a generic 3D shape that is deformable
        - More practical solution
        - Has been used in face/head tracking and modeling

Synthesis based on image warping and a generic shape

Original Input Synthesis Original Input Synthesis

Recovering 3D shape from a single face image

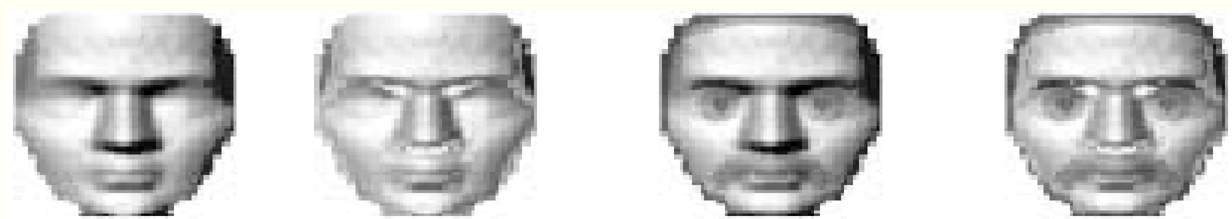
Shape-from-Shading(SFS)



Input synthesized synthesized prototype

Symmetric SFS (Does not work for arbitrary albedo yet)

Case of constant albedo Case of piece-wise albedo



Input synthesized Input synthesized

(Zhao and Chellappa, 2001)

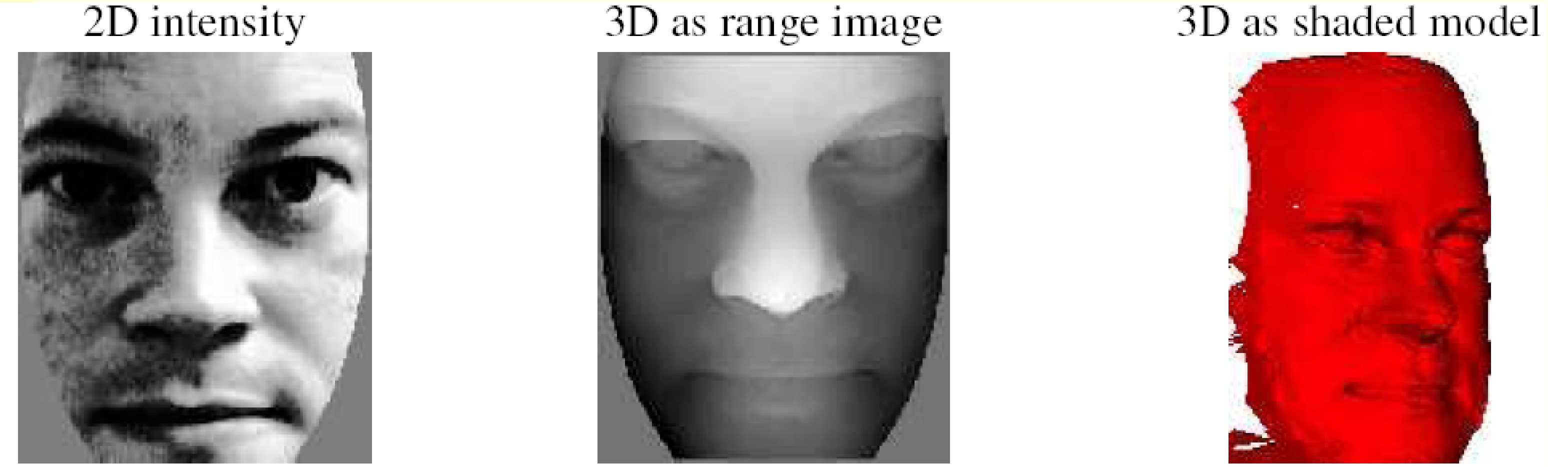
3D Face Recognition

3D Face Recognition

1. 3D face recognition methods
   - Single-modal 3D
   - Multi-modal 3D + 2D
2. How to obtain 3D structure?
   - Stereo imaging
   - Active projection of structured light
3. How to compare non-ideal 3D structure?
4. Still in the early development stage
   - Most reported results are NOT based on large datasets

Why 3D Face Recognition?

1. 3D model is invariant to illumination change
2. 3D model is invariant to pose change (after alignment)
3. Cheap, accurate, and close-to real-time 3D scanner is available
    - Active projection of structured light



Survey of 3D Face Recognition

Reference	number of persons	number of images	image size	3D face data	reported performance	size variation	expression variation
Face Recognition Algorithms Using Only 3D Data							
Cartoux 1989 [7]	5	18	?	profile, surface	100%	yes	no
Lee 1990 [12]	6	6	256x150	EGI	none	no	some
Gordon 1992 [9]	26 train 8 test	26 train 24 test	?	feature vector	100%	yes	no
Nagamine 1992 [18]	16	160	256x240	multiple profiles	100%	yes	no
Achermann 1997 [3]	24	240	75x150	range image	100%	yes	no
Tanaka 1998 [20]	37	37	256x256	EGI	100%	no	no
Achermann 1997 [2]	24	240	75x150	point set	100%	yes	no
Hesher 2003 [10]	37	222 (6 expr. ea.)	242x347	range image	97%	yes	no
Medioni 2003 [14]	100	700 (7 poses ea.)	?	surface mesh	98%	yes	no
Moreno 2003 [17]	60	420 (3 expr., 2 poses)	avg 2,200 point mesh	feature vector	78%	yes	some
Lee 2003 [13]	35	70	320x320	feature vector	94% at rank 5	yes	no

K. Bowyer, K. Chang, and P. Flynn, A Survey Of 3D and Multi-Modal 3D+2D Face Recognition Notre Dame Department of Computer Science and Engineering Technical Report, January 2004.

Survey of 3D+2D Face Recognition

• Multi-modal 3D+2D face recognition is better than 3D only face recognition

Multi-Modal 3D + 2D Face Recognition Algorithms							
Lao 2000 [11]	10	360	480x640	surface mesh	91%	yes	no
Beumier 2001 [4]	27 gallery 29 probes	240 2D	?	multiple profiles	1.4% EER	yes	no
Wang 2002 [22]	50	300	128x512	feature vector	>90%	no	yes
Bronstein 2003 [6]	157	?	2250 avg. vertices	range image	not reported	yes	yes
Tsalakanidou 2003 [21]	40	80	100x80	range image	99% 3D+2D 93% 3D only	yes	no
Chang 2003 [8]	200 (+ 75 in training)	951	480x640	range image	99% 3D+2D 93% 3D only	yes	no

K. Bowyer, K. Chang, and P. Flynn, A Survey Of 3D and Multi-Modal 3D+2D Face Recognition Notre Dame Department of Computer Science and Engineering Technical Report, January 2004.

An Exemplar Procedure for 3D + 2D Face Recognition

Dataset (x 275)

200 galleries, 676 probes, 275 training set

1 – 13 weeks between gallery and probes

Methods

1. normalization

   intensity & pose

2. identification

3. fusion (score-based)

   score normalization – decision

K. Bowyer, K. Chang, and P. Flynn, A Survey Of 3D and Multi-Modal 3D+2D Face Recognition Notre Dame Department of Computer Science and Engineering Technical Report, January 2004.

Normalization

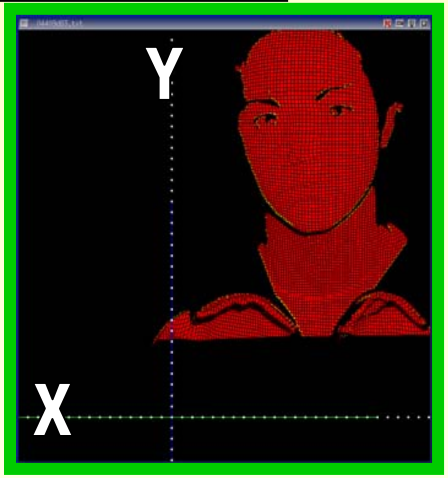
- Procedure to minimize the variations in data

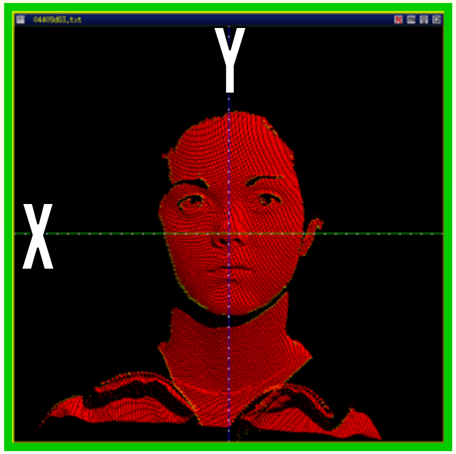
  Intensity in 2D

  - Rotation around Z axis in X-Y plane
  - Histogram equalization

  Pose in 3D

  - Rotation around X, Y and Z
  - Translation nose tip to the origin



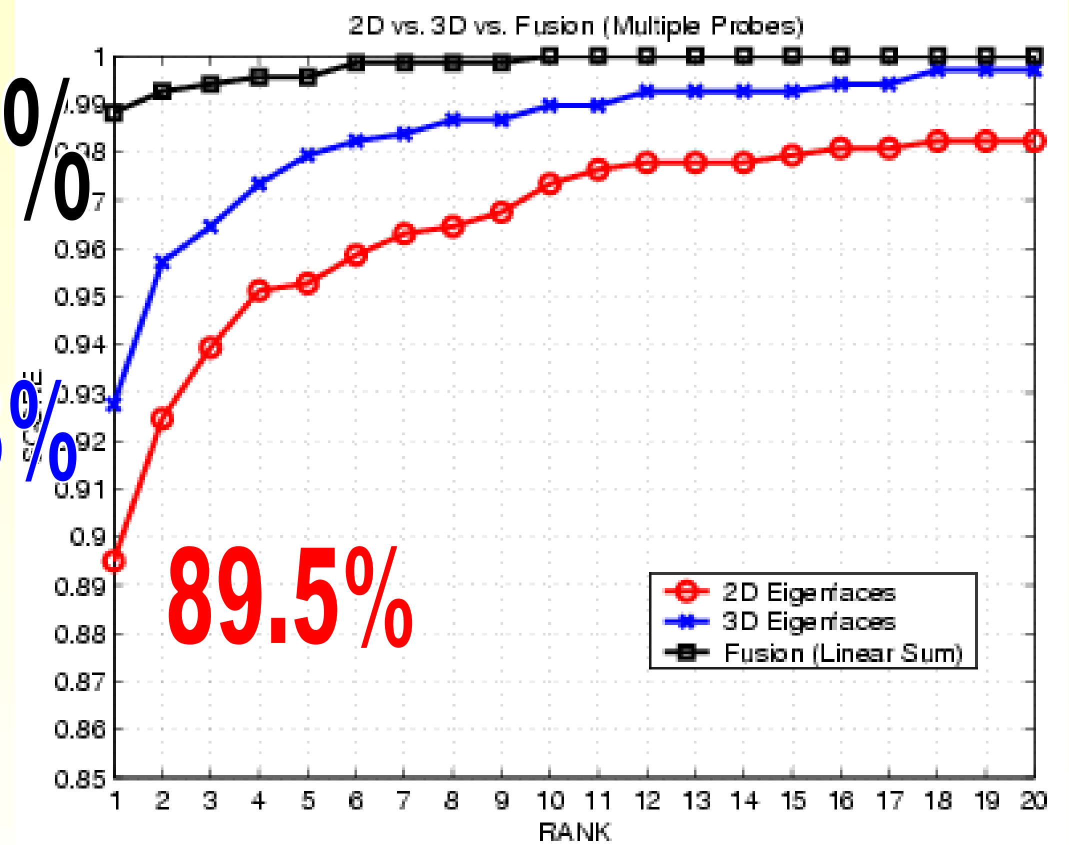


K. Bowyer, K. Chang, and P. Flynn, A Survey Of 3D and Multi-Modal 3D+2D Face Recognition Notre Dame Department of Computer Science and Engineering Technical Report, January 2004.

Performance Comparison

98.8%

92.8%

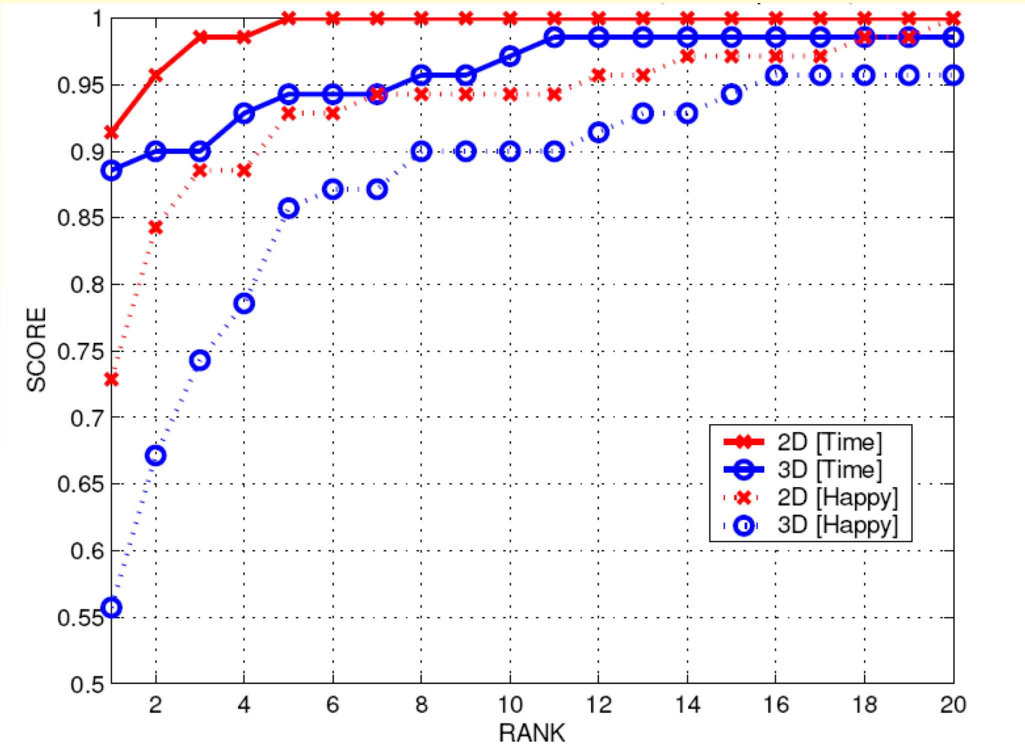


89.5%

- Training 275
- Time lapse
  1 - 13 weeks
- 200 galleries
- 676 probes

K. Bowyer, K. Chang, and P. Flynn, A Survey Of 3D and Multi-Modal 3D+2D Face Recognition
Notre Dame Department of Computer Science and Engineering Technical Report, January 2004.

Issue of 3D Face Recognition

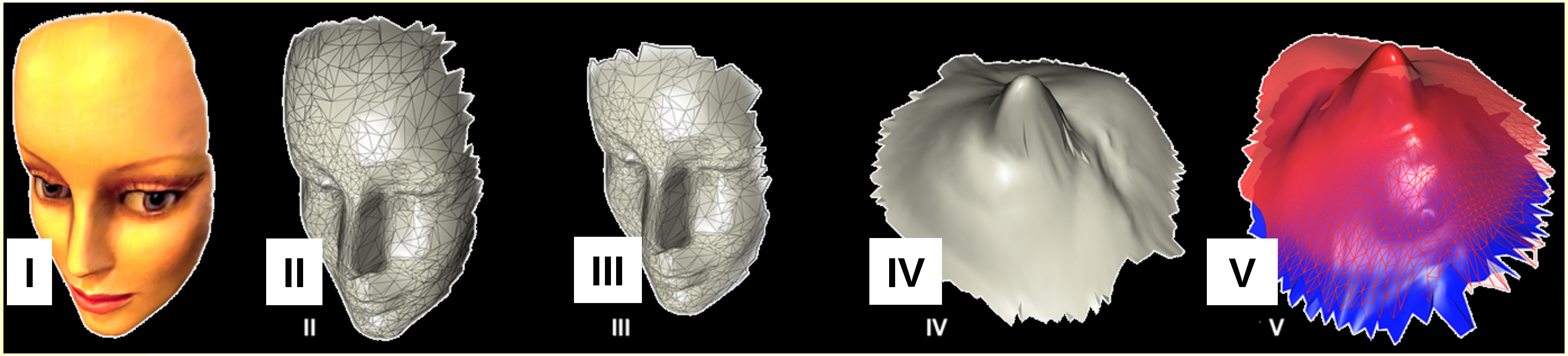


- Eigenface
- Time lapse
  2 - 13 weeks
- 70 galleries (Neutral)
- 70 probes (Smile)

Facial expression is a challenge!

K. Bowyer, K. Chang, and P. Flynn, A Survey Of 3D and Multi-Modal 3D+2D Face Recognition Notre Dame Department of Computer Science and Engineering Technical Report, January 2004.

Expression-Invariant 3D Face Recognition



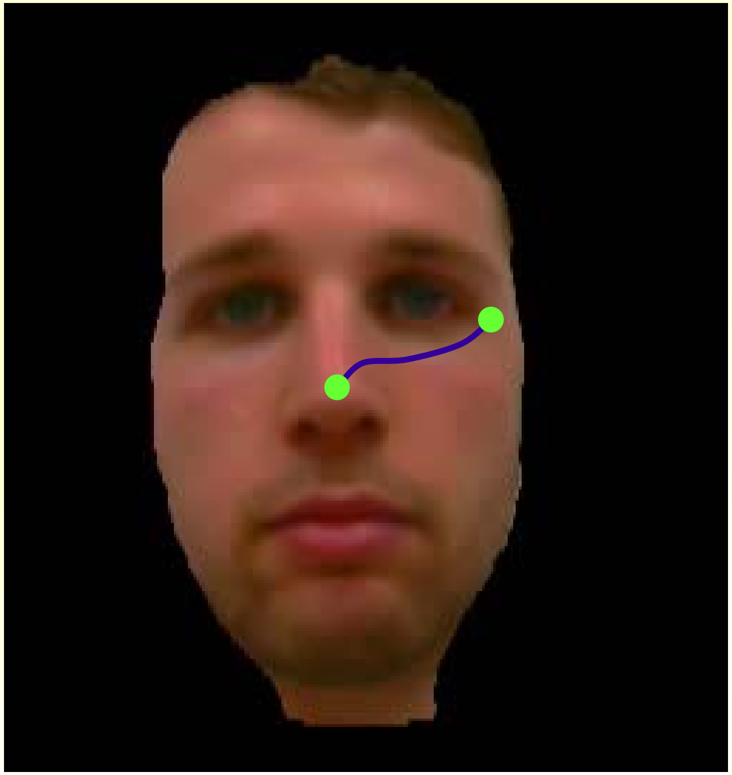
- Range camera acquires facial surface (**I**).
- The surface is smoothed (**II**), subsampled and cropped (**III**).
- Fast marching computes geodesic distances on the surface.
- Facial surface is flattened via MDS (multi-dimensional scaling) (**IV**).
- Rigid surface matching using the canonical surfaces (**V**).

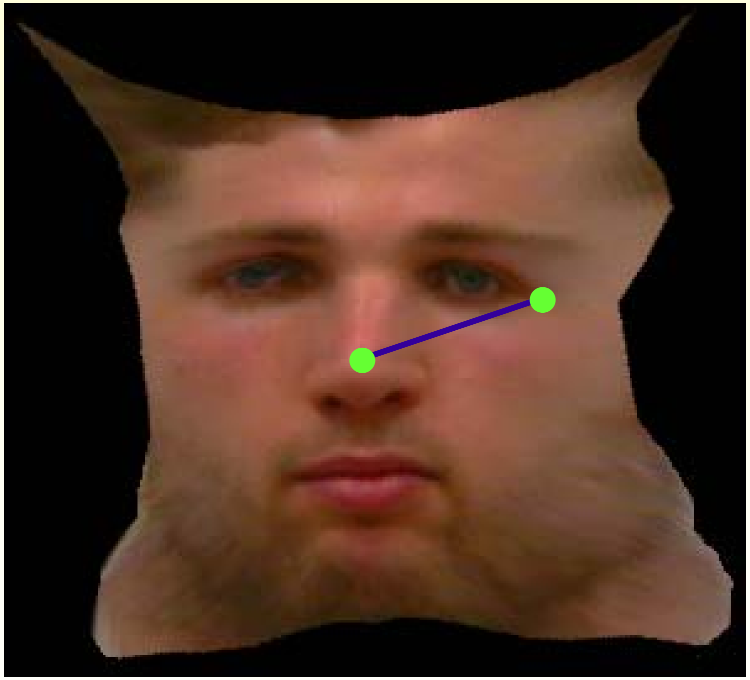
A. Bronstein, M. Bronstein, E. Gordon. And R. Kimmel, Expression-Invariant 3D Face Recognition, Audio and Video based Person Authentication (AVBPA), 2003.

R. Kimmel and J.A Sethian, “Computing geodesic on manifolds”. *Proc. of* ***National Academy of Science*** 95, pp. 8431–8435, 1998.

Texture Flattening via MDS

Flatten curved surfaces so that the geodesic distances between two points are well approximated by Euclidean distances in the embedding surface



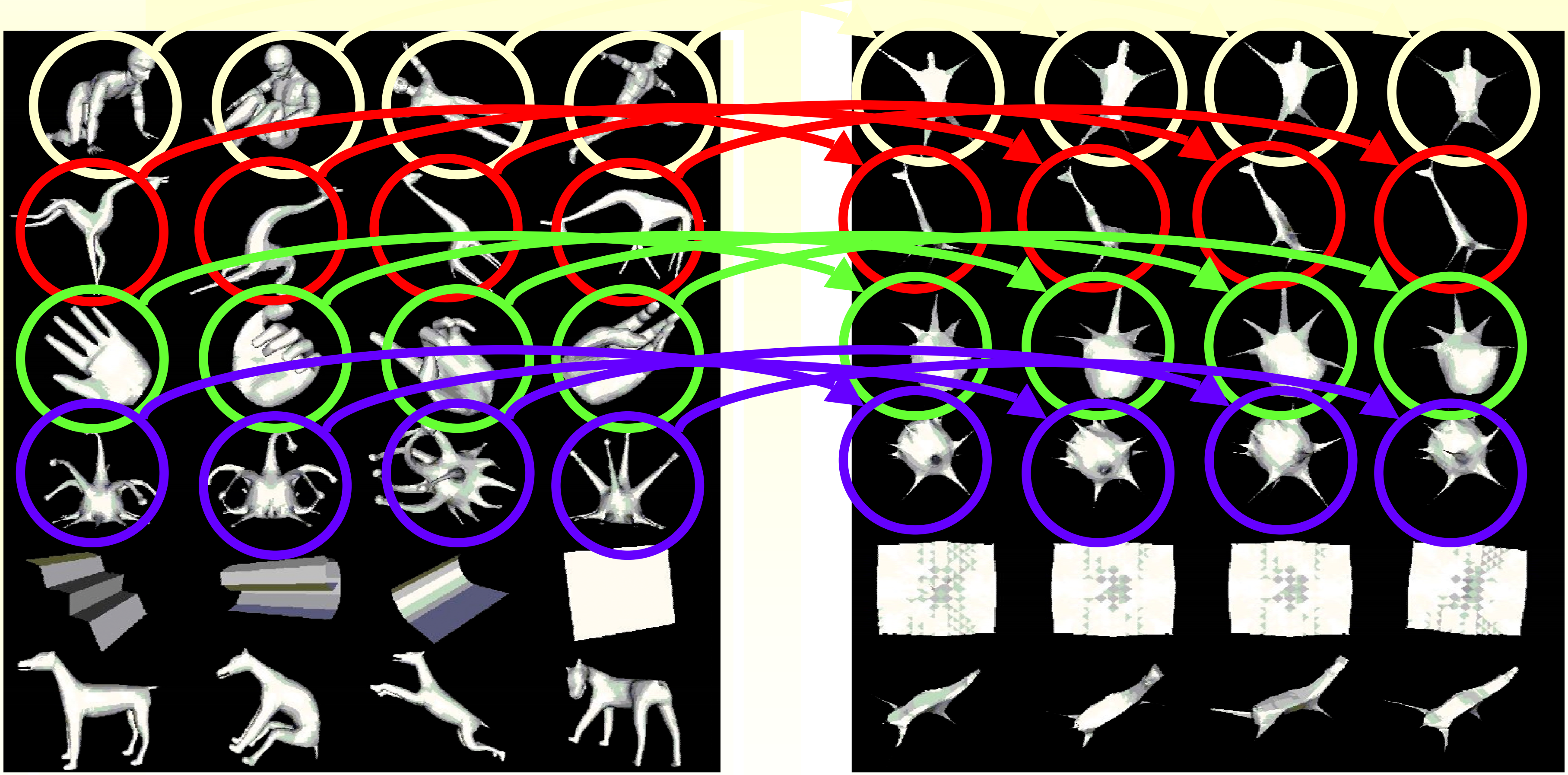


G. Zigelman, R. Kimmel, and N. Kiryati, “Texture Mmapping using surface flattening via multi-dimensional scaling, IEEE Trans. Visualization and Comp. Gracphics, 8. pp. 198-207, 2002.

Bending-Invariant Canonical Signatures

The flattened manifold via MDS

Original surfaces — Canonical surfaces in $\mathbf{R}^3$

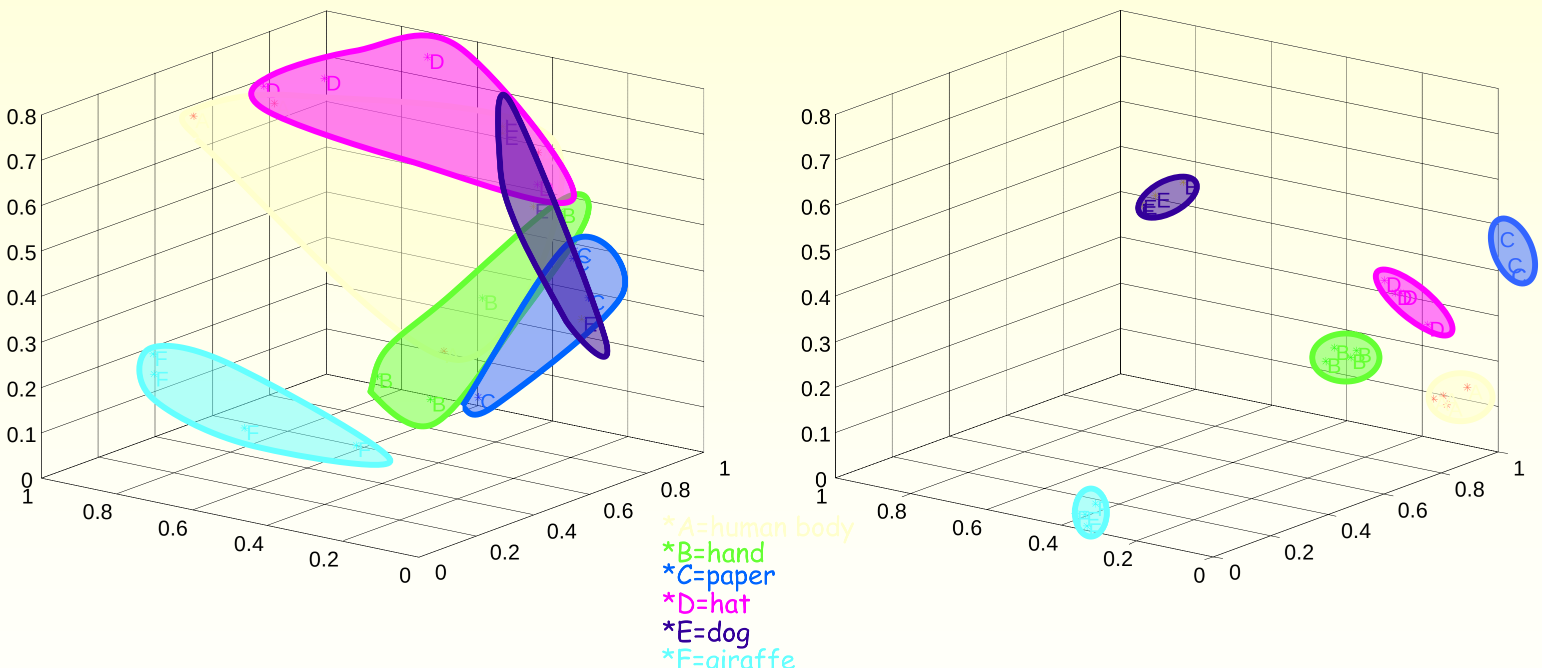


Elad and Kimmel, *CVPR'2001/PAMI'2003*

Clustering of Canonical Signatures

2nd moments based MDS for clustering

Original surfaces   Canonical forms

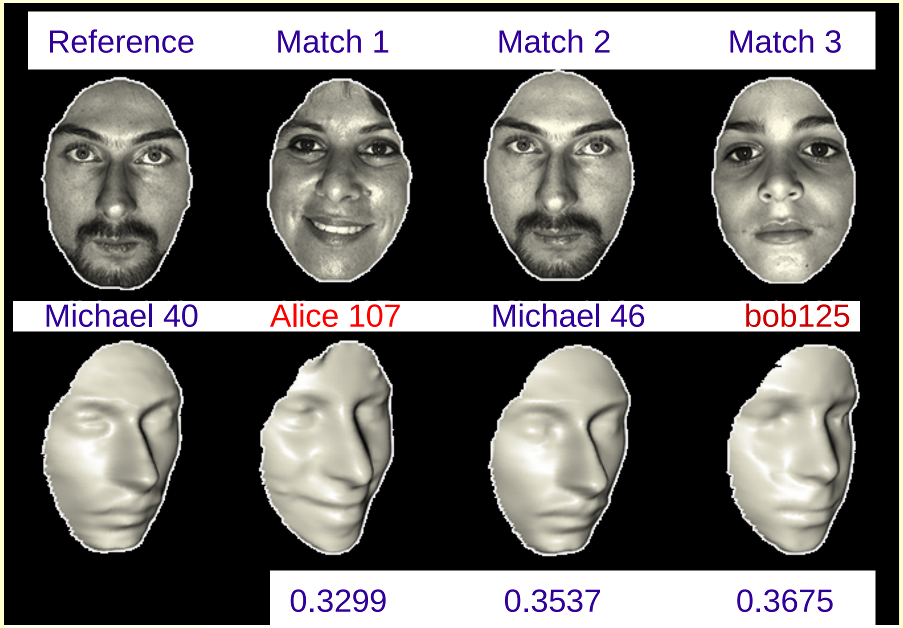


Elad and Kimmel, *CVPR'2001/PAMI'2003*

Twins Test I: 3D surfaces

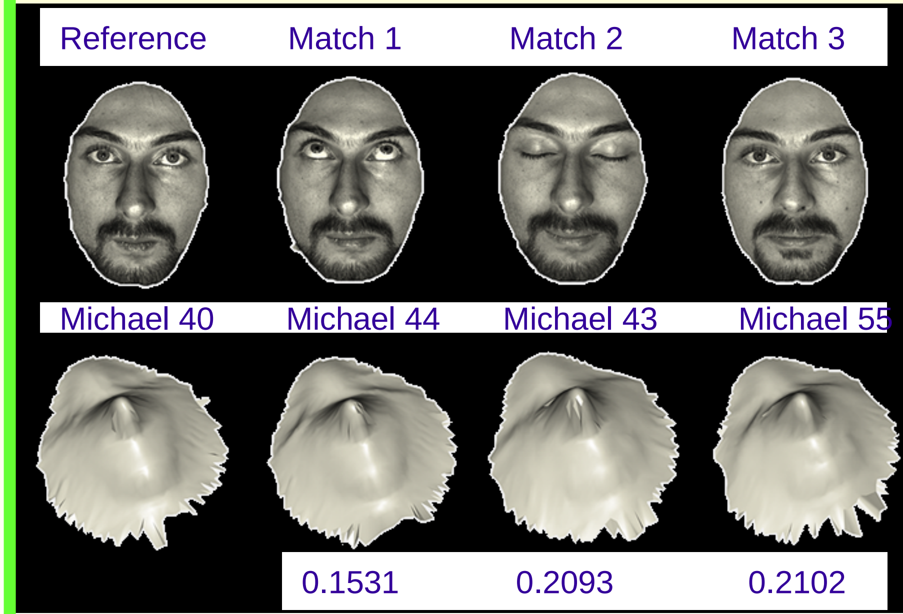
- Recognizing **twins**, a challenging test for face recognition.

SURFACE MATCHING



Facial surfaces as rigid objects -> inaccurate.

CANONICAL FORM MATCHING

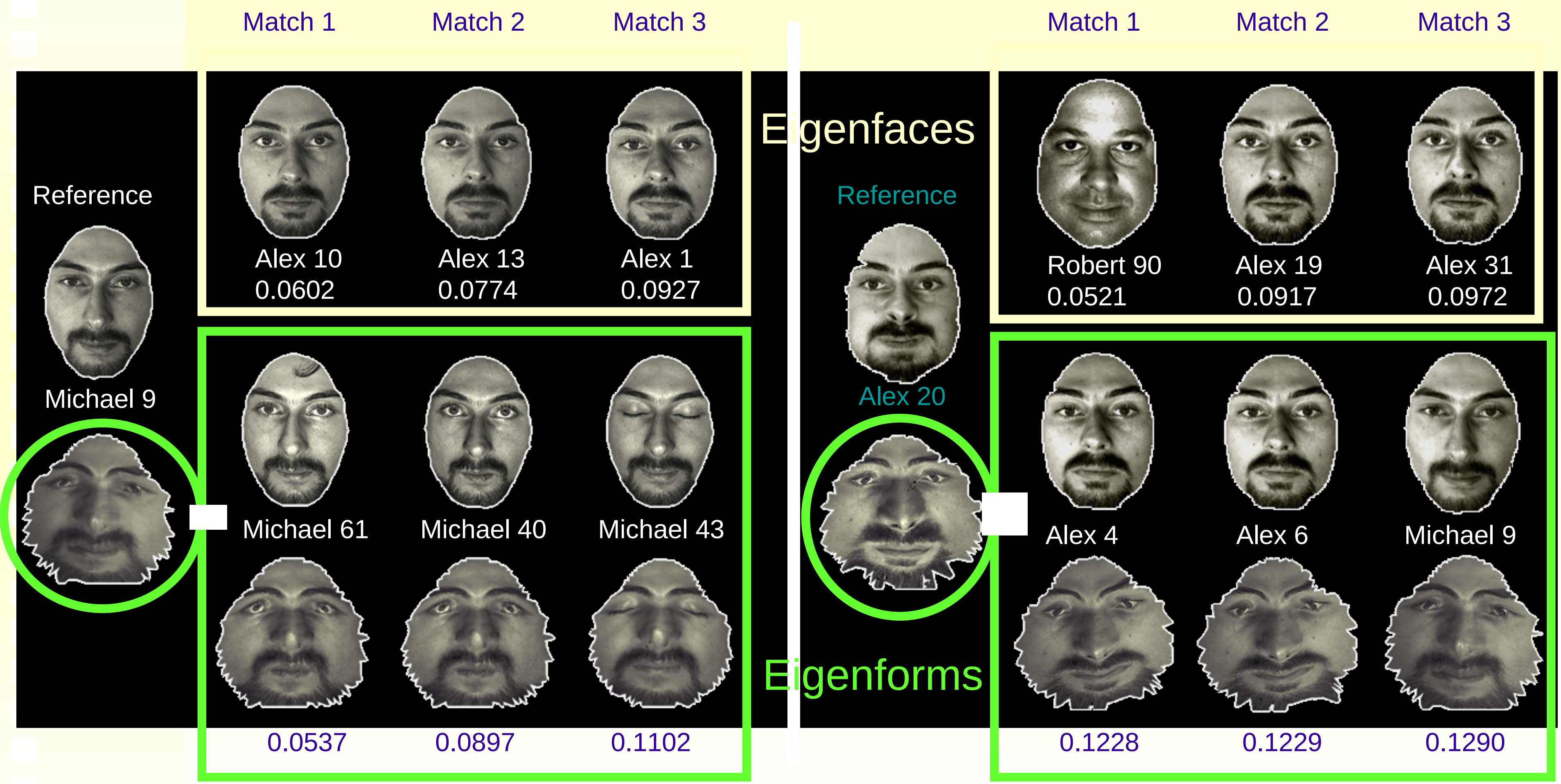


Canonical forms tell apart identical twins.

A. Bronstein, M. Bronstein, E. Gordon. And R. Kimmel, Expression-Invariant 3D Face Recognition, Audio and Video based Person Authentication (AVBPA), 2003.

A. Bronstein, M. Bronstein and R. Kimmel, "3D face recognition using geometric invariants", submitted, 2003

Twins Test II: Eigenforms (2D+3D)



A. Bronstein, M. Bronstein, E. Gordon. And R. Kimmel, Expression-Invariant 3D Face Recognition, Audio and Video based Person Authentication (AVBPA), 2003.

A. Bronstein, M. Bronstein and R. Kimmel, "3D face recognition using geometric invariants", submitted, 2003

Performance Evaluation

Performance Evaluation

1. FERET 93-97→ FRVT 2000 → FRVT 2002

   http://www.frvt.org

2. M2VTS → XM2VTS → BANCA

   http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb

   http://banca.ee.surrey.ac.uk.

3. Colorade State University Web Site

   http://www.cs.colostate.edu/evalfacerec/

FRVT 2002


FACE RECOGNITION
VENDOR TEST 2002

Dr. Jonathon Phillips
DARPA & NIST

http://www.frvt.org

September, 2003

Authors: Jonathon Phillips, Patrick Grother, Ross Micheals, Duane M. Blackburn, Elham Tabassi, and Mike Bone.

Background

FRVT 2002

United States Patriot Act:

"…develop and certify a technology standard…that can be used to verify the identity of persons applying for a US VISA…"

- Motivation:
  - Entry/exit United States
  - Airport Security
- Goals:
  - Assess performance on large real-world data sets
  - Identify new promising approaches
  - Measure progress on difficult face recognition problems
    - Pose variation
    - Images taken months/years apart
    - Video sequences

FRVT 2002 - History

FR 2 0 0 2 VT

FERET 1993-95

- Tech Agent: J. Phillips
- Face Recognition
- Established Face Database
- Established Standardized Evaluation Methodologies
- Basis of all Face Recognition Technology on HumanID

Facial Recognition Vendor Test 2000

- Assessed improvements since FERET
- Evaluated commercial state-of-the-art

FACE RECOGNITION VENDOR TEST 2002

- Assessed improvements since FRVT 2000
- Large scale, operational database
- Difficult problems

FRVT 2002

- Independent evaluation of face recognition systems
- Administered July and August 2003
- Sequestered data
- Open to:
  - Mature proto-types
  - Commercial Systems
- From:
  - Academia
  - Research Labs
  - Industry

Test Design

FRVT 2002

Two tests:

- High Computational Intensity Test
  - Measure performance on very large data sets
  - U.S. Dept of State Mexican non-immigrant visa
  - 121,589 still images
  - 37,437 individuals

- Medium Computational Intensity Test
  - 7,500 images
  - Pose variations
  - Months/years between images
  - Illumination

High Computational Intensity Test

Example Photos

Medium Computational Intensity Test

FRVT 2002

Indoor and Outdoor Images

Medium Computational Intensity Test

3D Morphable Models

Conclusions

FRVT 2002

- Indoor performance has improved since FRVT 2000.
- Performance decreases approximately linearly with elapsed time.
- Better systems are not sensitive to indoor lighting changes.
- Three-dimensional morphable models improve performance.
- Males are easier to recognize than females.
- Older people are easier to recognize than younger people.
- Outdoor face recognition performance needs improvement.

http://www.frvt.org

Challenges and Directions

Technical Challenges: System Performance

IEEE Computer Society Magazine Cover Feature
"An Introduction to Evaluating Biometric Systems"

"On the basis of media hype alone, you might conclude that biometric passwords will soon replace their alphanumeric counterparts with versions that cannot be stolen, forgot-ten, lost, or given to another person. But what if the performance estimates of these systems are far more impressive than their actual performance?"

Three major testing protocols:

- FERET (http://www.itl.nist.gov/iad/humanid/feret/)
- XM2VTS (http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/)
- FRVT (http://www.frvt.org)

Technical Challenges: System Requirement

- In many applications, devices are very small and needs very low power.

- Many of the existing algorithms are complex enough and still need to be refined to improve the performance.

“A PDA-based Face Recognition System”

at WACV 2002

by Jie Yang, Xilin Chen, William Kunz

School of Computer Science, Carnegie Mellon University

Technical Challenges: Algorithms

- There exists an important question
  - Theoretically, how unique and invariant faces could be?
  - Practically, how many people can be distinguished?
- The current focus is on how to recognize face under
  - Varying illumination (outdoor)
  - Varying pose (naturally)
  - Aging (missing children)
  - Cosmetic/makeup (?)
  - Degenerated quality (surveillance video)
  - Compressed images (to some extend)

Industry Challenges: Killer Applications

Challenges of finding killer-applications

- The technology is hardly perfect, especially compared with human beings. But under limited conditions it is quite successful. For example, verification is quite good under indoor, controlled environment. It can be integrated into Biometrics systems.
- Video game, Virtual reality.
- Cameras embedded in cell phone/palm pilot/watch: intelligent processing including face recognition.
- Surveillance/monitoring cameras
- …….

Conclusion and Directions

- Machine perception of faces has numerous applications.
- The choice of an appropriate method should be based on the specific requirement (e.g., image size) of a given task.
- Video-based face recognition is rapidly evolving while 3D face recognition is still in its early development stage.
- Face recognition plays an important role in biometrics.
- Comparison to human perceptual system:
  - Machine recognition of faces has reached a certain level of maturity, especially in terms of the gallery size. Meanwhile human perceptual system has limitations on the number and types of faces that can be easily distinguished.
  - Machine systems are still far away from the capability of human perceptual system, especially under realistic outdoor-like environment.

Resources

Main reference

Face Recognition: A Literature Survey

W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld

ACM Computing Survey Vol. 35, Dec. 2003/
University of Maryland Tech. Rep. CS-TR-4167R 2002/
http.www.cfar.umd.edu

Baseline Code

Colorado State University Evaluation Web site

www.cs.colostate.edu/evalfacerec

Relevant Conferences

- AFGR 1995-2002 (Automatic Face And Gesture Recognition)
- AVBPA 1997-2003 (Audio- and Video-Based Person Authenti.)
- AMFG 2003 (Analysis and Modeling of Faces and Gestures)
- ICIP/ICPR/ICME/CVPR/ICCV/WACV
- Biometric Consortium Conference (government and industry)

Internet Resources

Face Recognition Home Page

http.www.cs.rug.nl/~peterkr/FACE/frhp.html

Face Databases

- UT Dallas www.utdallas.edu/dept/bbs/FACULTY_PAGES/otoole/database.htm
- Notre Dame database www.nd.edu/~cvrl/HID-data.html
- MIT database ftp://whitechapel.media.mit.edu/pub/images
- Edelman ftp://ftp.wisdom.weizmann.ac.il/pub/FaceBase
- CMU PIE www.ri.cmu.edu/projects/project_418.htm
- Stirling database pics.psych.stir.ac.uk
- M2VTS multimodal www.tele.ucl.ac.be/M2VTS
- Yale database cvc.yale.edu/projects/yalefaces/yalefaces.htm
- Yale databaseB cvc.yale.edu/projects/yalefacesB/yalefacesB.htm
- Harvard database hrl.harvard.edu/pub/faces
- Weizmann database www.wisdom.weizmann.ac.il/~yael
- UMIST database images.ee.umist.ac.uk/danny/database.html
- Purdue rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html
- Olivetti database www.cam-orl.co.uk/facedatabase.html
- Oulu physics-based www.ee.oulu.fi/research/imag/color/pbfd.html