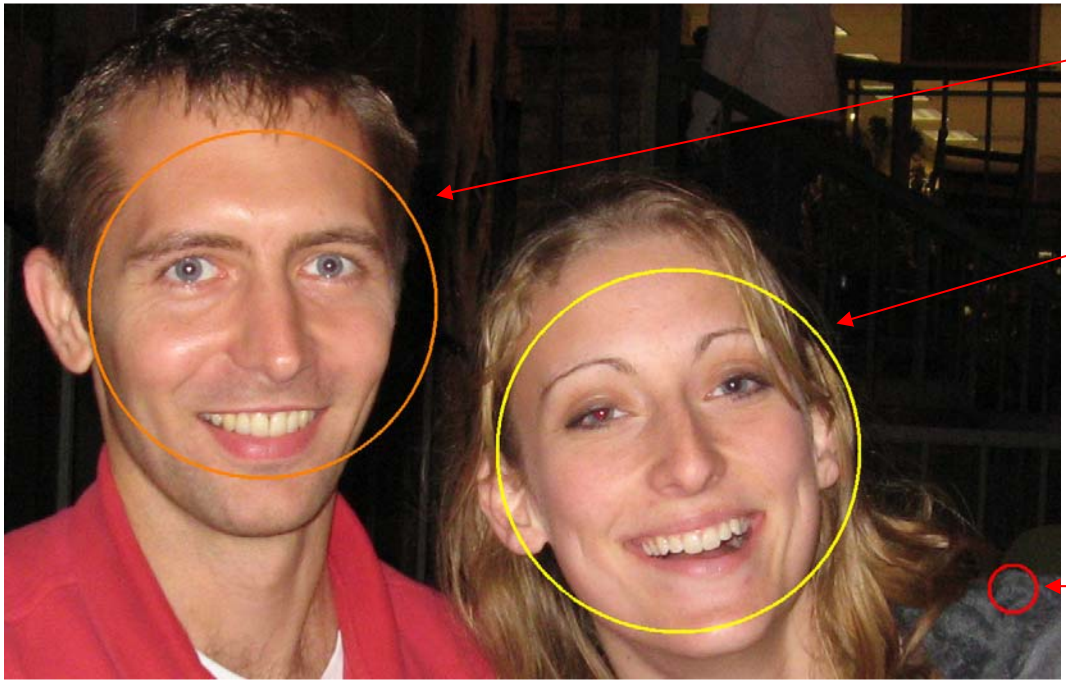
Faces

Newton Petersen

February 29, 2008

Introduction/Problem Statement



Tell me this is Newton

Don't tell me this is Newton

Minimize the 'what the #!@%s'

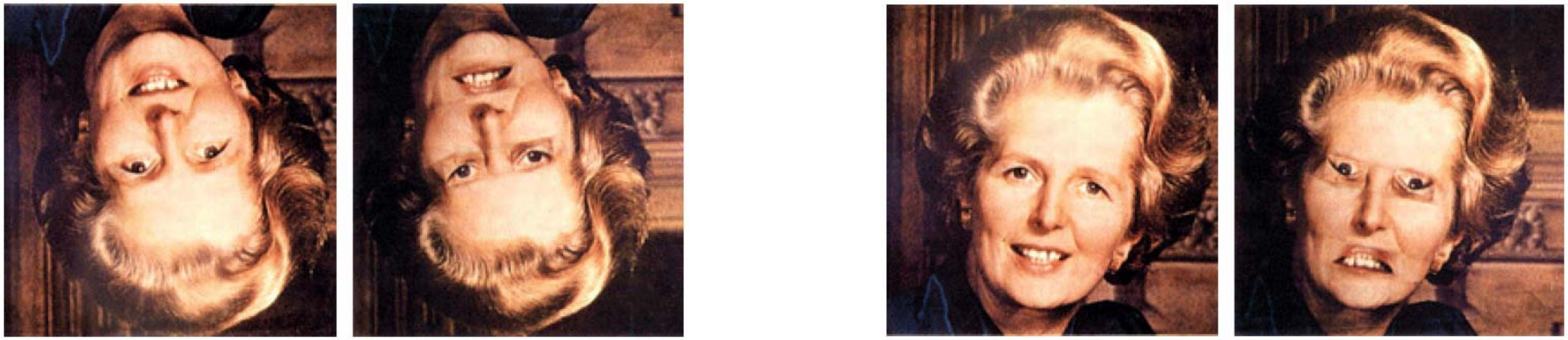
- Given still or video images identify or verify one or more persons using a stored database of faces
- Minimize false accepts and false rejects, maximize true accepts and true rejects
- Why, how, what works best, what's next?

Why Interesting

- Commercial applicability
  - Law enforcement
  - Security
  - Smart identification
  - Entertainment
  - Search
- Humans are good at it, why can't computers
  - Attracts diverse researchers from psychology, neuroscience, image processing, patter recognition, AI, computer vision
- Google is shopping

Image from Sinha, Balas, et al., "Face Recognition by Humans: 20 results all computer vision researchers should know about", 2005

Interesting Notes from Neuroscience



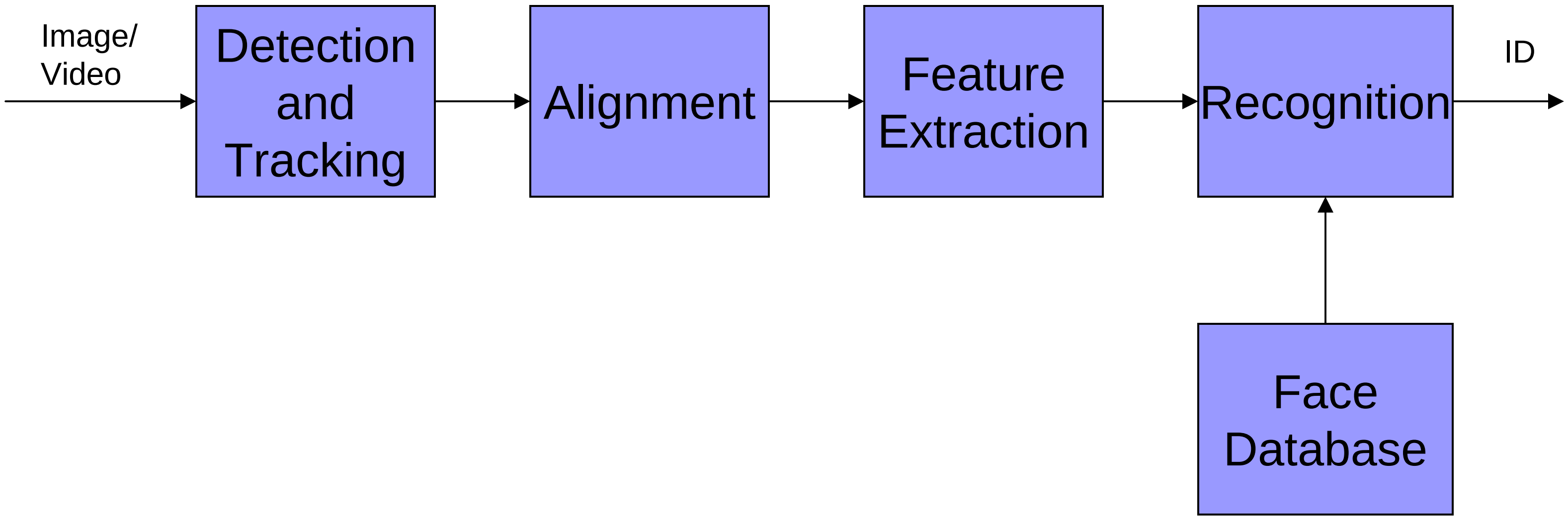
- Faces more easily remembered by humans than any other object when in upright orientation
- Evidence of holistic approach by human brain – inverted face more difficult to recognize
- Probably different circuits for detection and recognition
  - Distinctive faces easier to identify
  - Typical faces easier to detect
- Upper part of face more significant than lower
  - Oddly nose appears mostly insignificant
- Moving face easier to recognize

Image from Sinha, Balas, et al., "Face Recognition by Humans: 20 results all computer vision researchers should know about", 2005

Why Difficult

- Varying shape
- Varying illumination
- Varying pose
- Varying facial expressions (smiling vs frowning)
- Varying age and ethnicity
- Varying image resolution

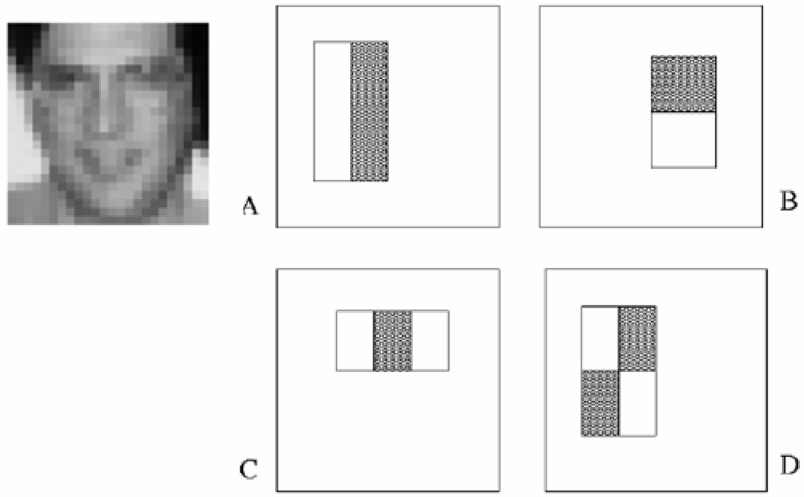
Key Technical Approach



Feature Based Detection - Viola and Jones '01

- Cascaded decisions based on 'rectangle features'
- AdaBoost used to select which features important
- Cascade of classifiers

‘Rectangle Features’

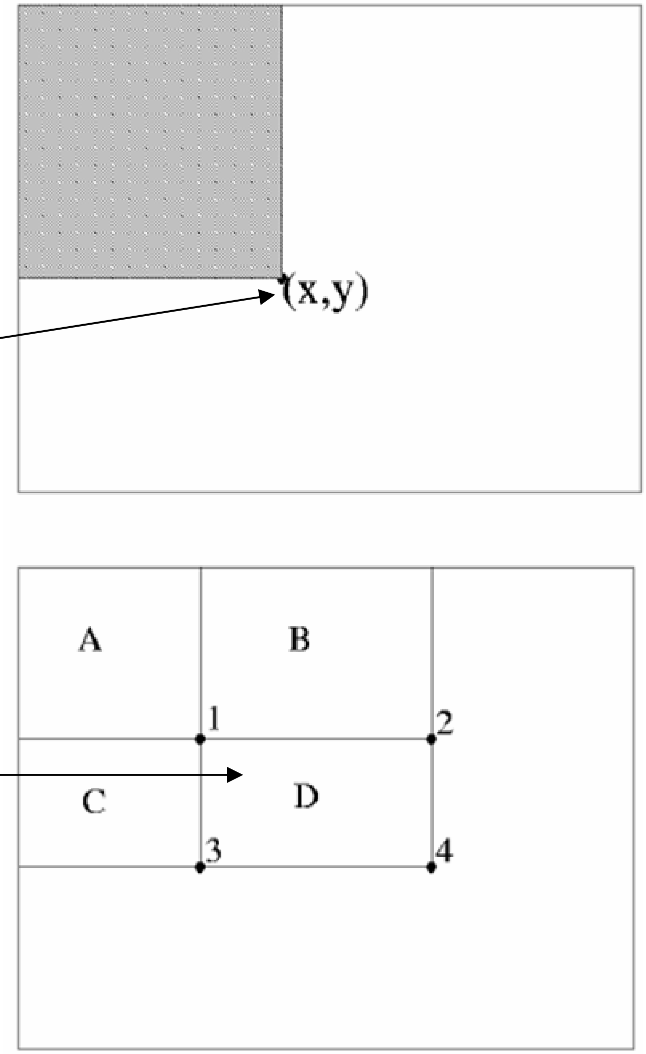


- Rectangle filters
- Sum of pixel values in white regions subtracted from pixel values in grey regions
- Efficiently computed using concept of an ‘integral image’

Viola, Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", 2001

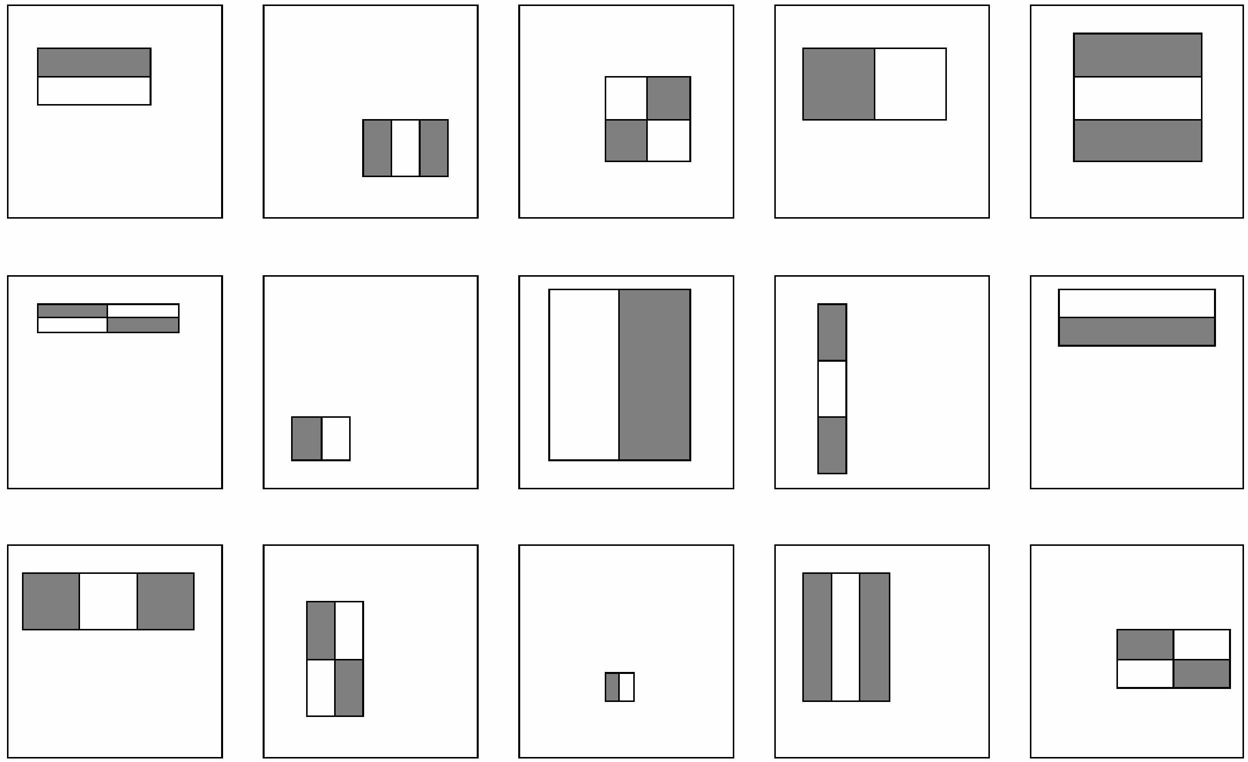
Integral Image

- Value at (x,y) is sum of pixels above and to the left of (x,y)
- Built with single pass through image
- Now a rectangle sum like D = 4 – 2 – 3 + 1
- Any size rectangle sum computed in constant time



Viola, Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", 2001

Rectangle windows



- 180,000+ rectangle windows possible
- How do you choose the windows that are most important?

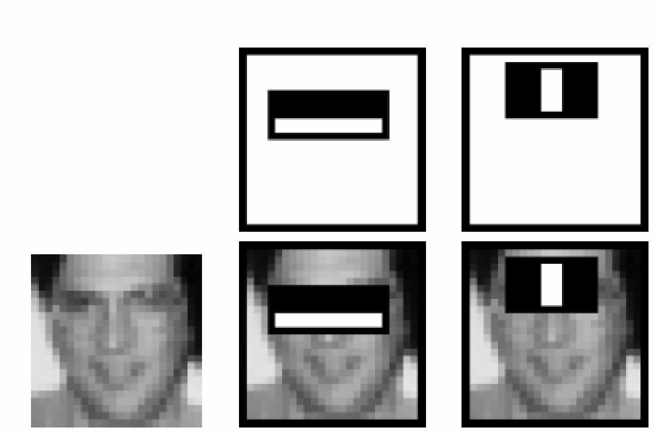
Viola, Jones, "Rapid Object Detection
using a Boosted Cascade of Simple
Features", 2001

AdaBoost for Feature Selection

- Image Features = Weak classifiers
- For each round of boosting
  - Evaluate each rectangle filter on each example
  - Select best threshold for each filter (minimize error)
  - Select best filter/threshold combination
  - Weight on ‘feature’ is just a function of error rate

Viola, Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", 2001

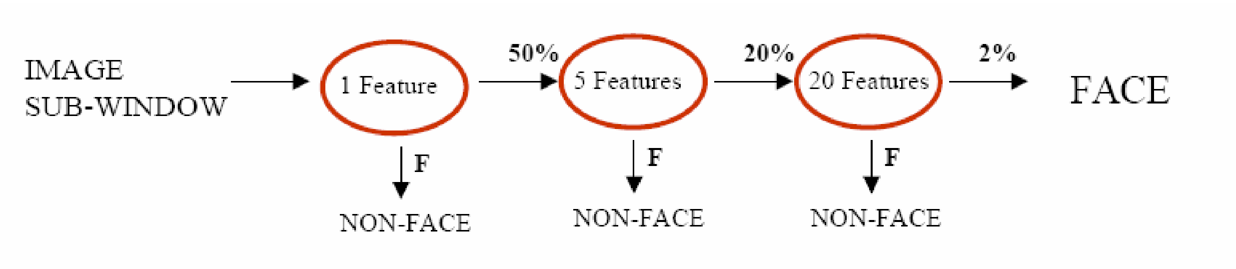
Most telling features



- First common feature – Eyes darker than nose and cheeks
- Second common feature – Eyes darker than bridge of nose
- …
- Now how is this used for detection?

Viola, Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", 2001

Cascaded Classifier for Detection



- Most telling feature checked first, if fail => no face
- Check less telling / more computationally intensive features next
- Continue until reach desired accuracy

Viola, Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", 2001

Discussion Point

- Perhaps different features are better identifiers for different people -> merge detection and identification?

Basic approaches to identification

- Holistic - Eigenfaces
- Feature geometry – Elastic Bunch Graph Matching
- Active Appearance Models
- Video/Multi-view

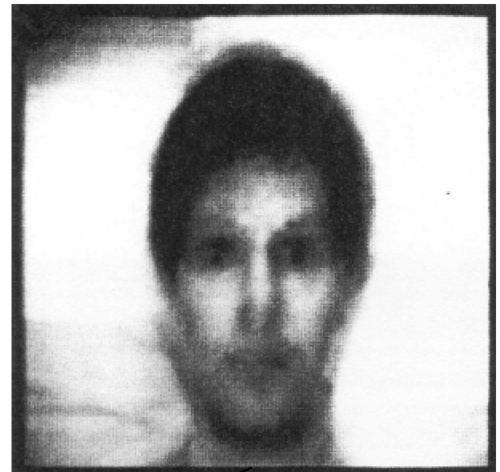
EigenFaces – Turk and Pentland '91

- Holistic Approach
- Attempts to find 'face space' automatically from training set of faces
- Basic idea: linear combination of eigenvectors can compose any face and capture important variability

EigenFaces - Example

Training Faces

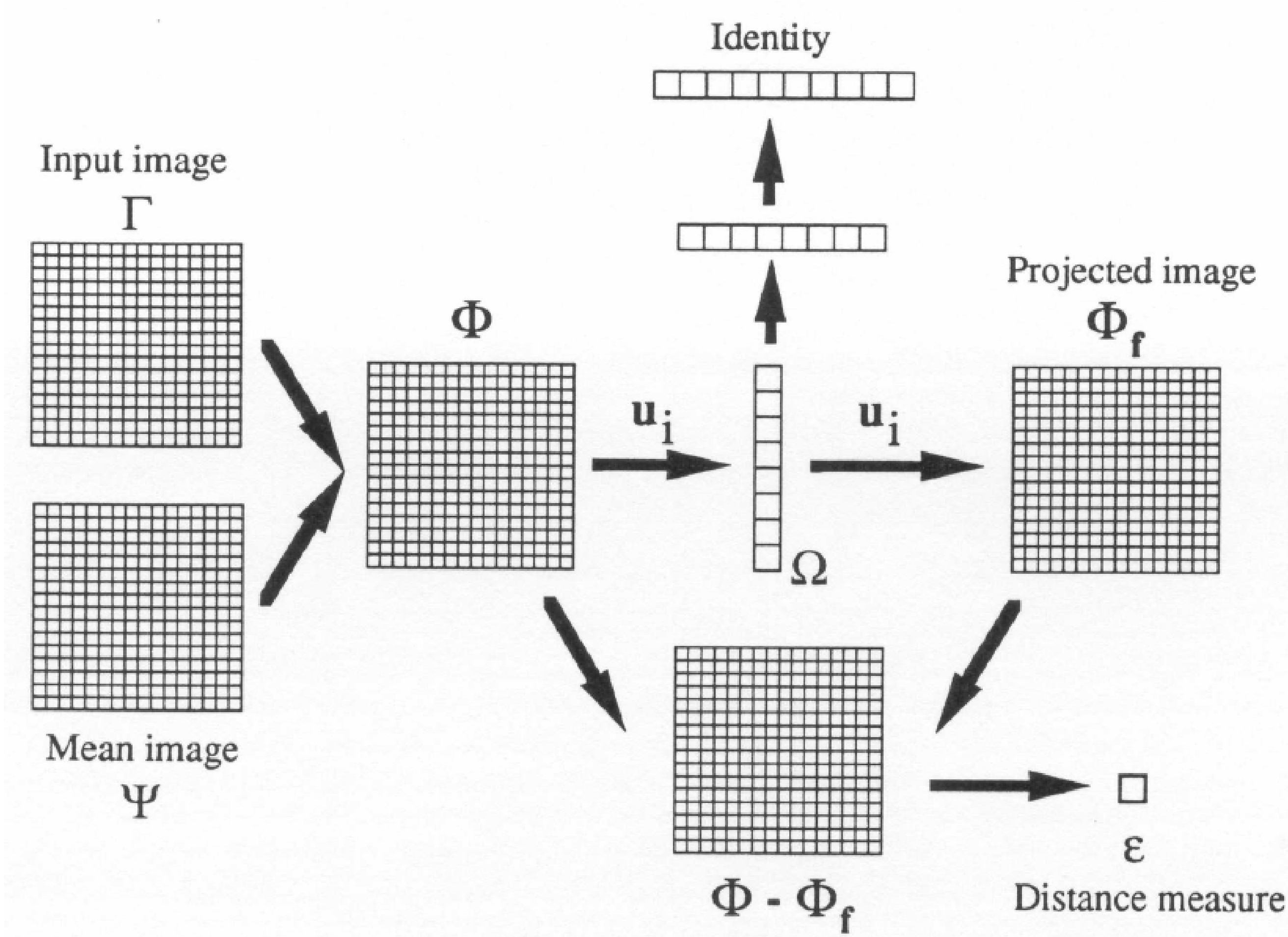
Average



7 highest weighted eigenvectors of covariance matrix

Turk and Pentland, "Eigenfaces for Recognition", 1991

EigenFaces – For Identification



Turk and Pentland, "Eigenfaces for Recognition", 1991

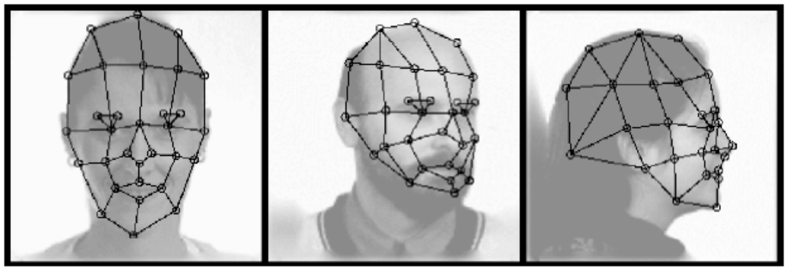
EigenFaces – Pros and Cons

- Pros
  - Training automatic
  - Agnostic to the object even being a face
  - Adequately reduces statistical redundancy in face image representation
- Cons
  - Difficult to capture things like expression changes
  - Sensitive to illumination and pose changes
  - Also sensitive to just pixel misalignment
  - Occlusion causes problems

Local Feature Analysis

- Uses face domain knowledge
- Models size and distance (shape) between geometric features on the face
- Can also model appearance, textures, etc

Elastic Bunch Graph Matching



Manually choose fiducial points

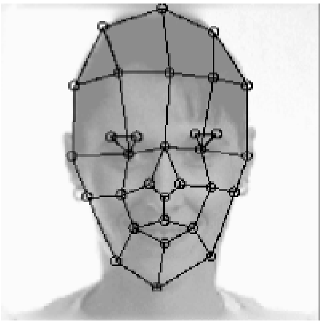
$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} \exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right)\left[\exp\left(i\vec{k}_j \vec{x}\right) - \exp\left(-\frac{\sigma^2}{2}\right)\right]$$

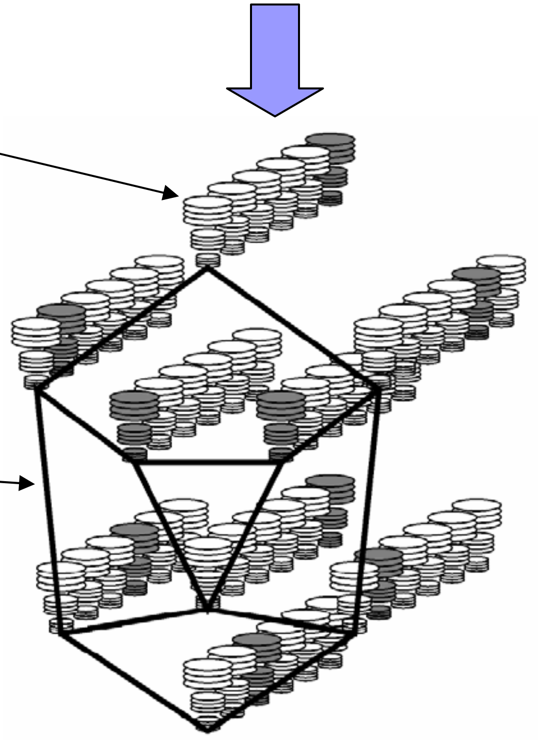
Apply Gabor wavelet kernel at points of interest to get local frequency and phase information

Why is frequency content more meaningful than intensities?

Wiskott, Fellous, Kruger, Malsburg, "Face Recognition by Elastic Bunch Graph Matching", 1997

Elastic Bunch Graph Matching





Each ‘bunch’ in graph holds wavelet coefficients, ‘jets’, for population of interest at particular fiducial points

Edges in graph contain length between fiducial points

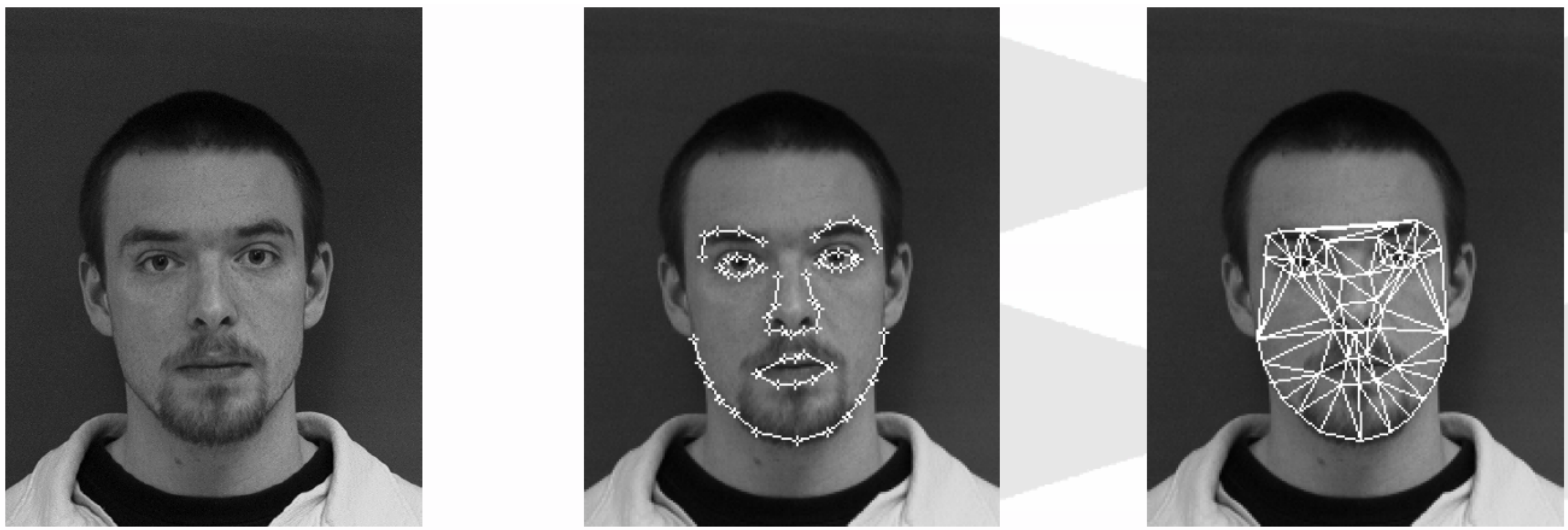
Find best match for identification

Wiskott, Fellous, Kruger, Malsburg, "Face Recognition by Elastic Bunch Graph Matching", 1997

Elastic Graph

- Pros
  - Encodes domain knowledge
  - Search over multiple scales relatively straight forward
  - Occluded fiducial points don't necessarily cause problems
- Cons
  - Points of interest manually identified
  - Pose manually labeled
  - Somewhat sensitive to rotations (supposedly not big deal < 22 degrees)
  - Illumination changes cause problems

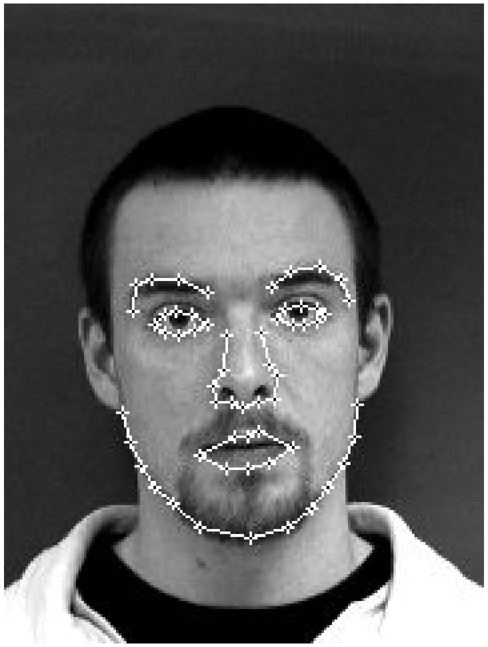
Active Appearance Models



Training example | Annotated training example | AAM mesh model

Cootes, Edwards, Taylor, "Active
Appearance Models", 2001

Active Appearance Models



$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s\mathbf{c} \quad \text{Shape}$$

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g\mathbf{c} \quad \text{Texture}$$

Mean

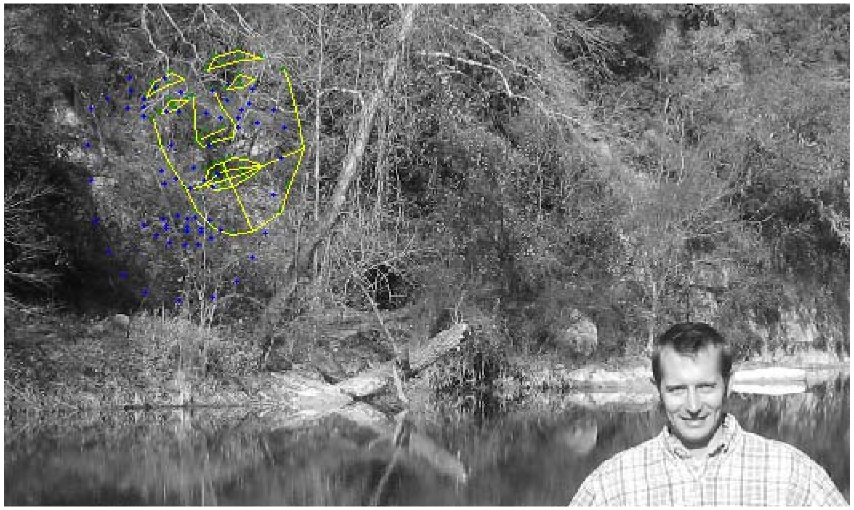
Modes of Variation

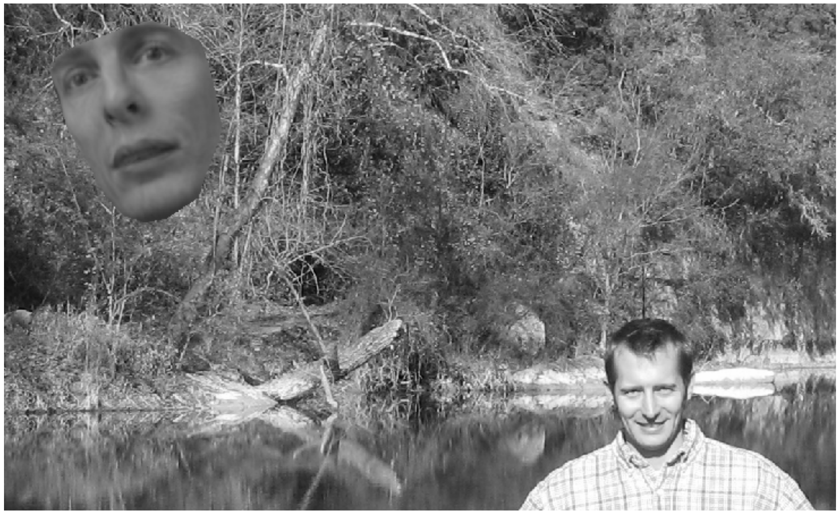
Face specific parameters

- Manually select points defining main features
- Statistical shape model built
- Texture model then built – By Eigen-analysis

Cootes, Edwards, Taylor, "Active Appearance Models", 2001

Active Appearance Models - Hmm





- Gradient based search
- Other detectors (Viola-Jones) seem to work better for just detection
- Then the active appearance model can be applied to get identifying parameters

Cootes, Edwards, Taylor, "Active Appearance Models", 2001

Active Appearance Models

- Pros
  - Again encoding domain knowledge
  - Mesh model more accurately captures shape than elastic graph
  - Shape model somewhat allows applying scaling, rotation, and translation
- Cons
  - Detection is sensitive to local minima
  - Illumination still causes problems (although they try to normalize intensity)

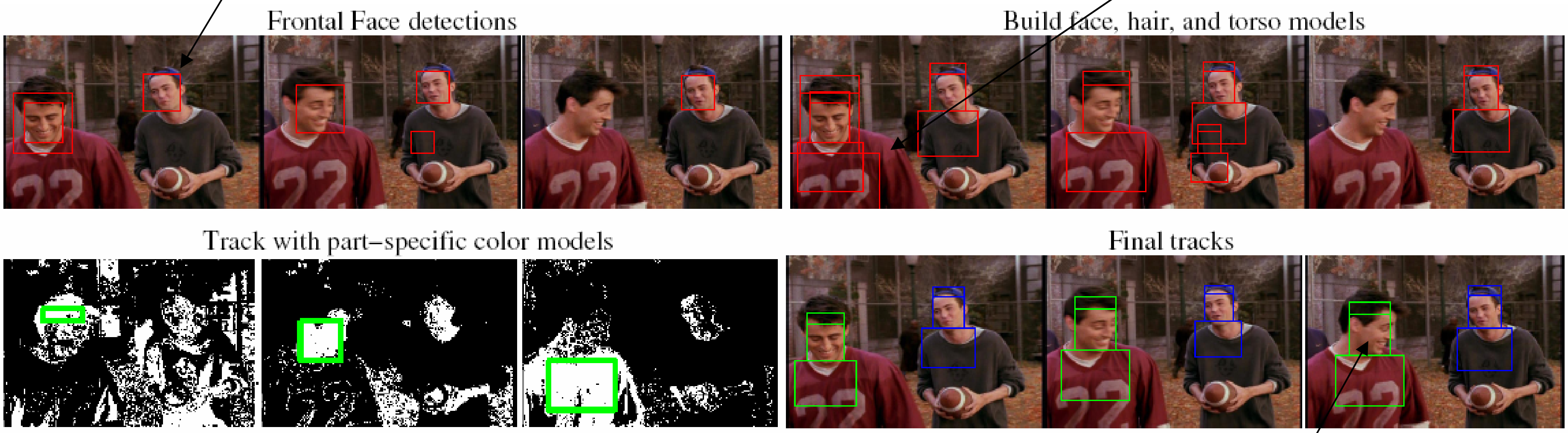
Video

- Quality usually lower than still images
- Faces are typically small
- Occlusion common
- But many different poses connected by minor motion shift
- And many context clues exist: clothing, etc

Video Cues

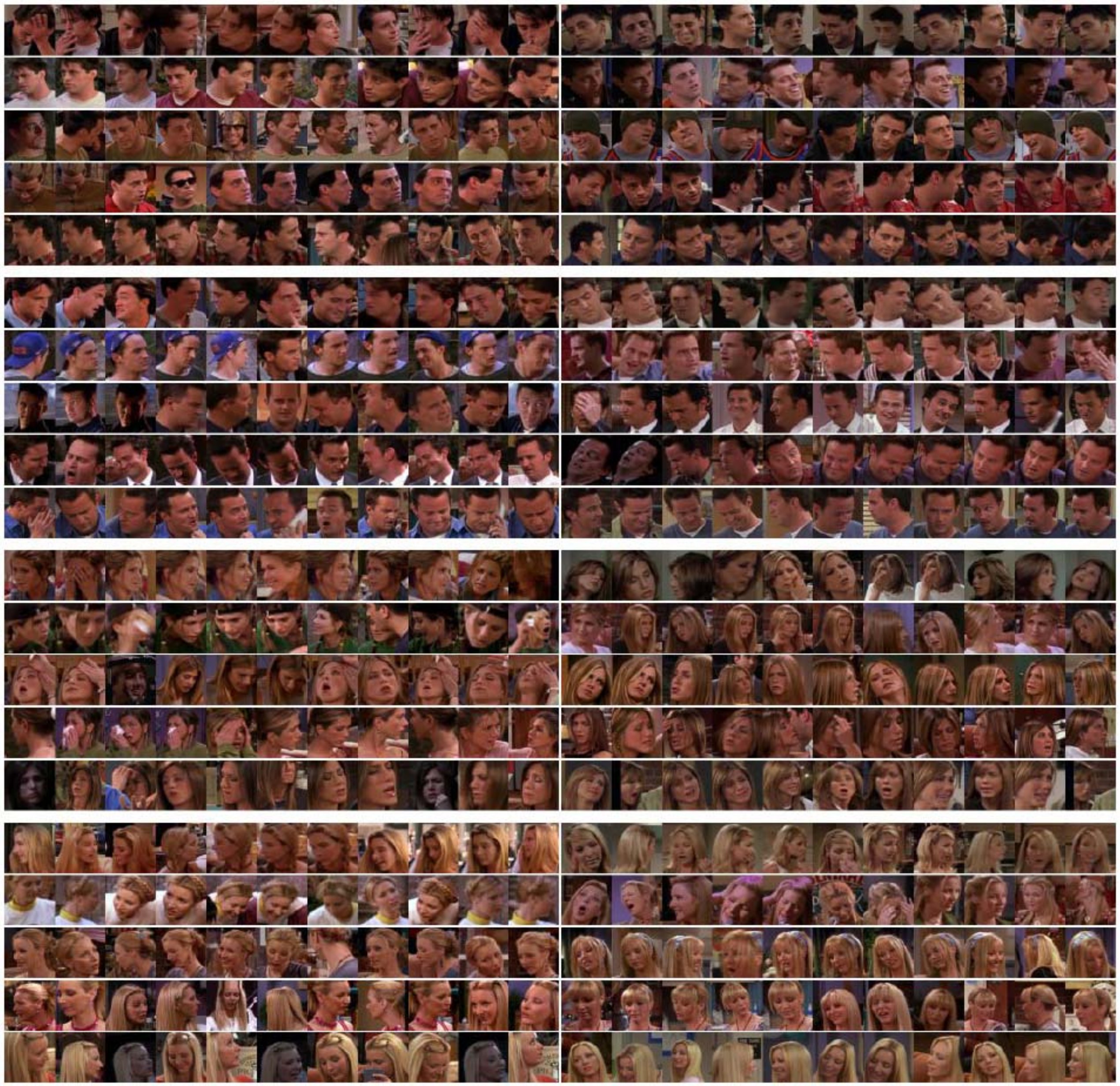
1. Viola-Jones Frontal Face Detection
2. Representation of face: simple eigenfaces

3. Part based color tracking



4. Using strict head torso models now we can detect and recognize non frontal faces

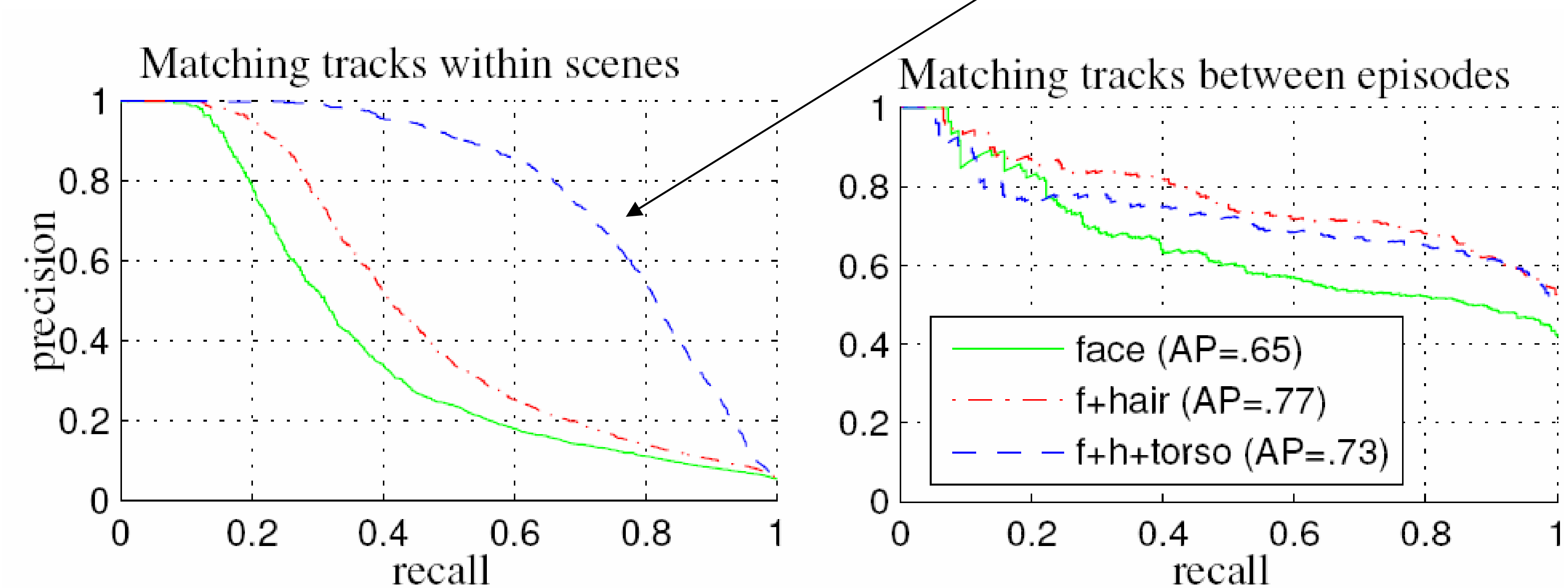
Ramanan, Baker, Kakade, "Leveraging archival video for building face datasets", 2007



Ramanan, Baker, Kakade, "Leveraging archival video for building face datasets", 2007

Does it work?

Significantly better with torso



- Precision – Of images labeled ‘Joey’ what percentage are really ‘Joey’
- Recall – Given query for ‘Joey’ what percentage of all ‘Joey’ shots are returned
- AP – Average Precision

Ramanan, Baker, Kakade, "Leveraging archival video for building face datasets", 2007

Discussion Points

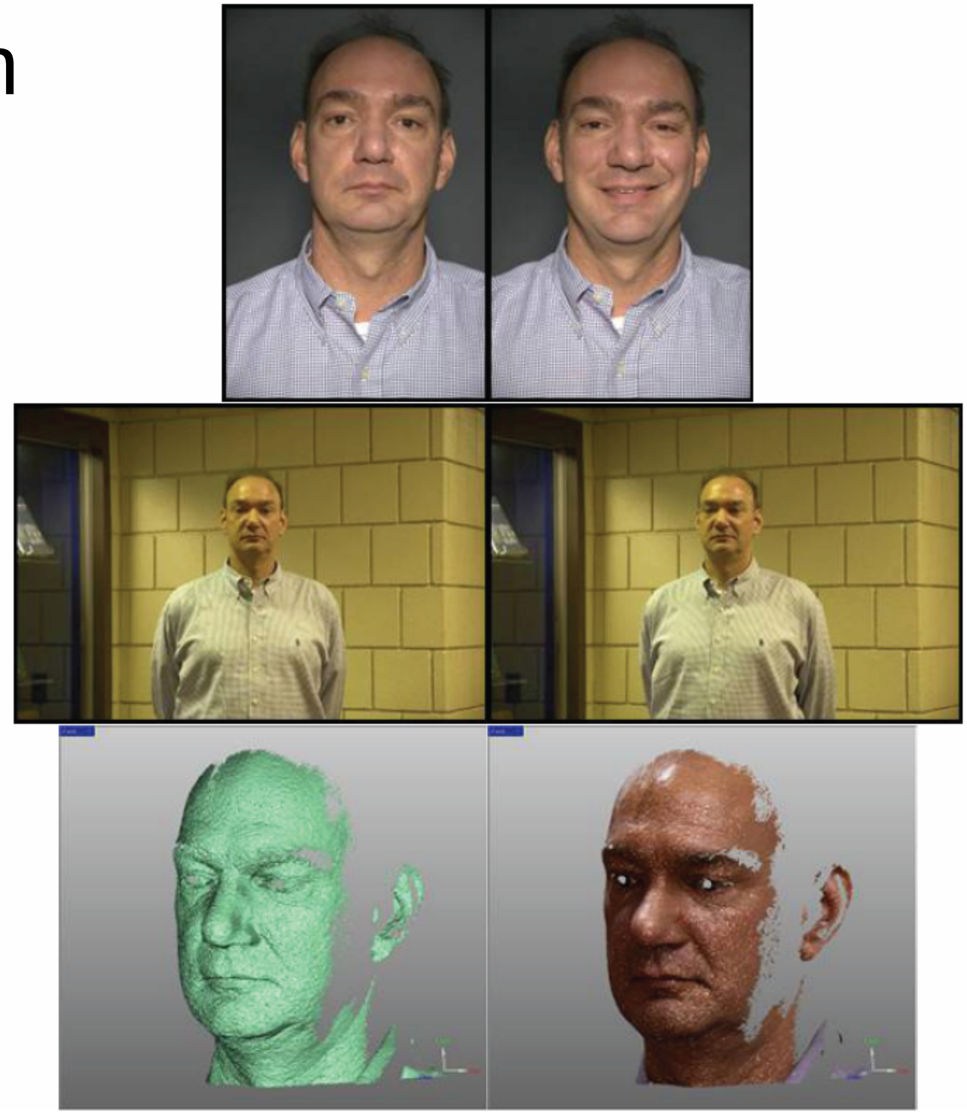
- What other cues could be used besides clothing and hair to link different poses of same character?
- Could low resolution you tube style videos be turned into higher resolution by using information/models gathered from different frames?

Check Point

- Detection
  - Viola-Jones: cascade of rectangle features
- Identification
  - EigenFaces: Auto determined face space
  - Elastic Graph
  - Active Appearance Models: Manual starting point for determining vector describing shape and texture
  - Video/Multi-view: a lot more raw information exists
- Now we'll look at evaluation methods and results
- And what's bubbling up as new/improved approaches

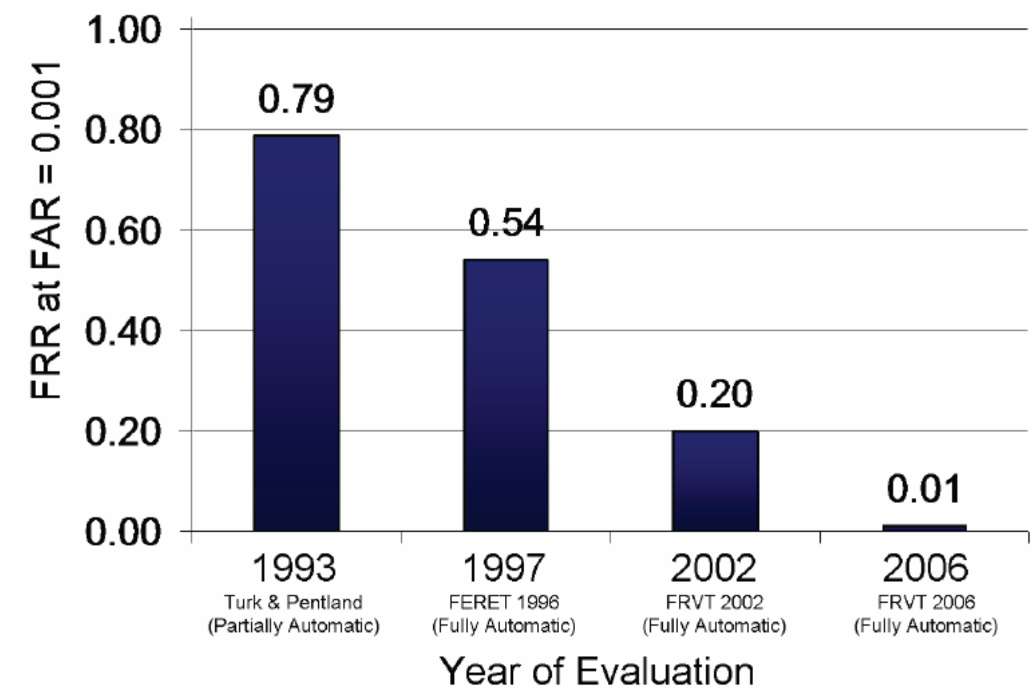
Means of Evaluation

- Face Recognition Vendor Test 2006
- Government effort to evaluate face recognition technology
- 2006 first time to examine 3D recognition
- 2006 first time performance compared to human performance



Phillips, Scruggs, et al., "FRVT 2006 and ICE 2006 Large-Scale Results", 2007

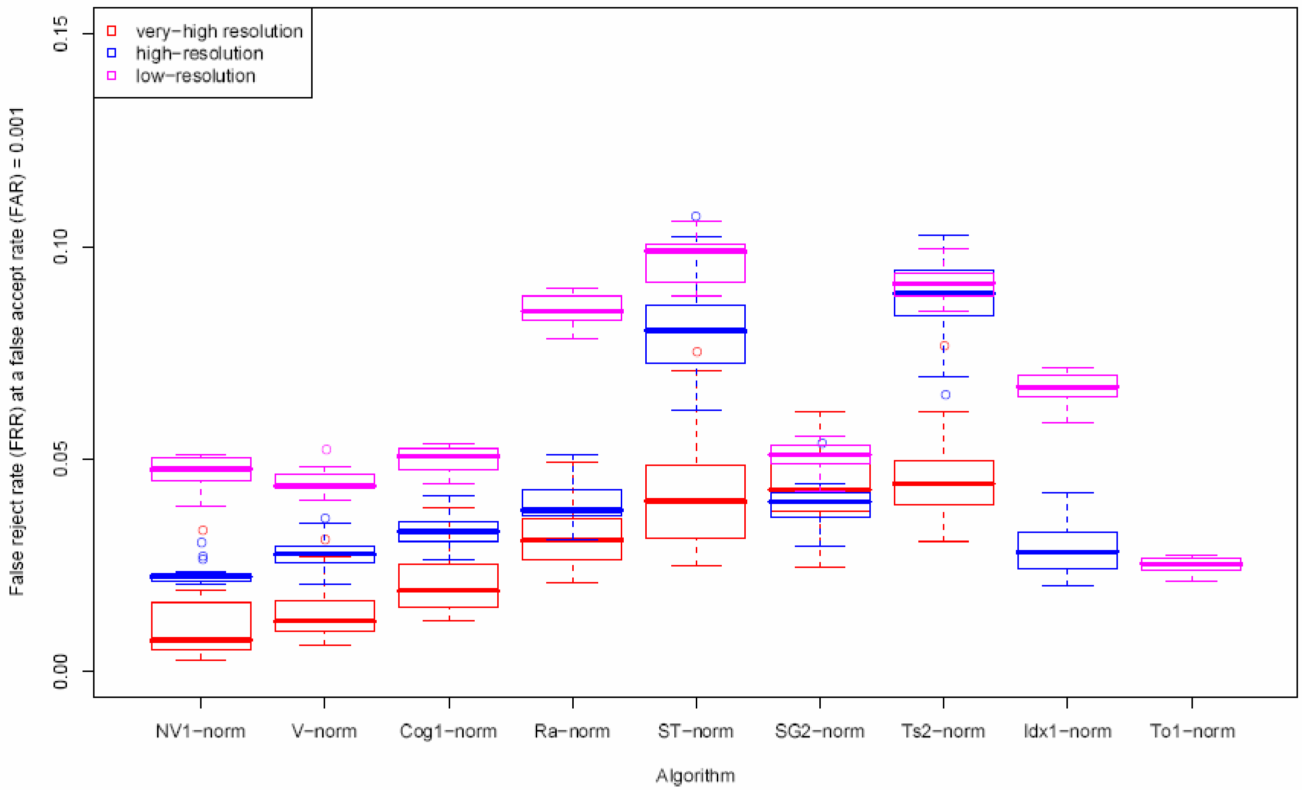
Face Recognition Vendor Test 2006



- High resolution
- Controlled illumination
- 10 million image database

Phillips, Scruggs, et al., "FRVT 2006 and ICE 2006 Large-Scale Results", 2007

What appears to work well today

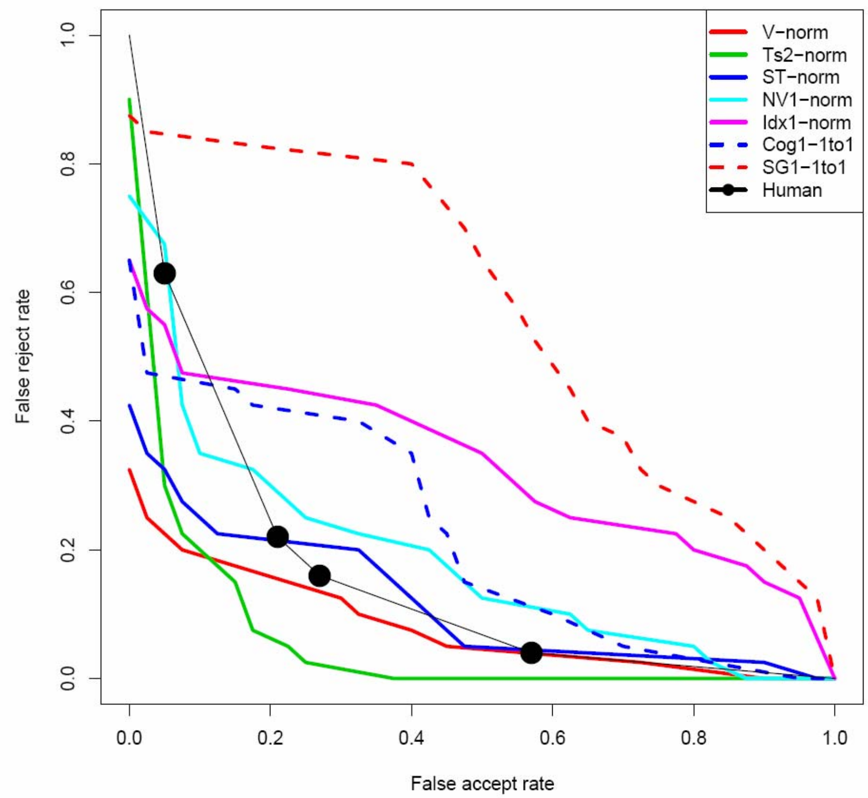


Group	Still norm
U. of Cambridge	
Cognitec	COG1-NORM
Geometrix	
U. of Houston	
Identix	IDX1-NORM
Iritech	
Neven Vision	NV1-NORM
Rafael	RA-NORM
Sagem	SG2-NORM
Sagem-Iridian	
SAIT	ST-NORM
Toshiba	TO1-NORM
Tsinghua U.	TS2-NORM
Viisage	V-NORM

- Neven Vision appears to get Gabor Wavelet ‘face template’ from local features

Phillips, Scruggs, et al., "FRVT 2006 and ICE 2006 Large-Scale Results", 2007

Man vs. Machine



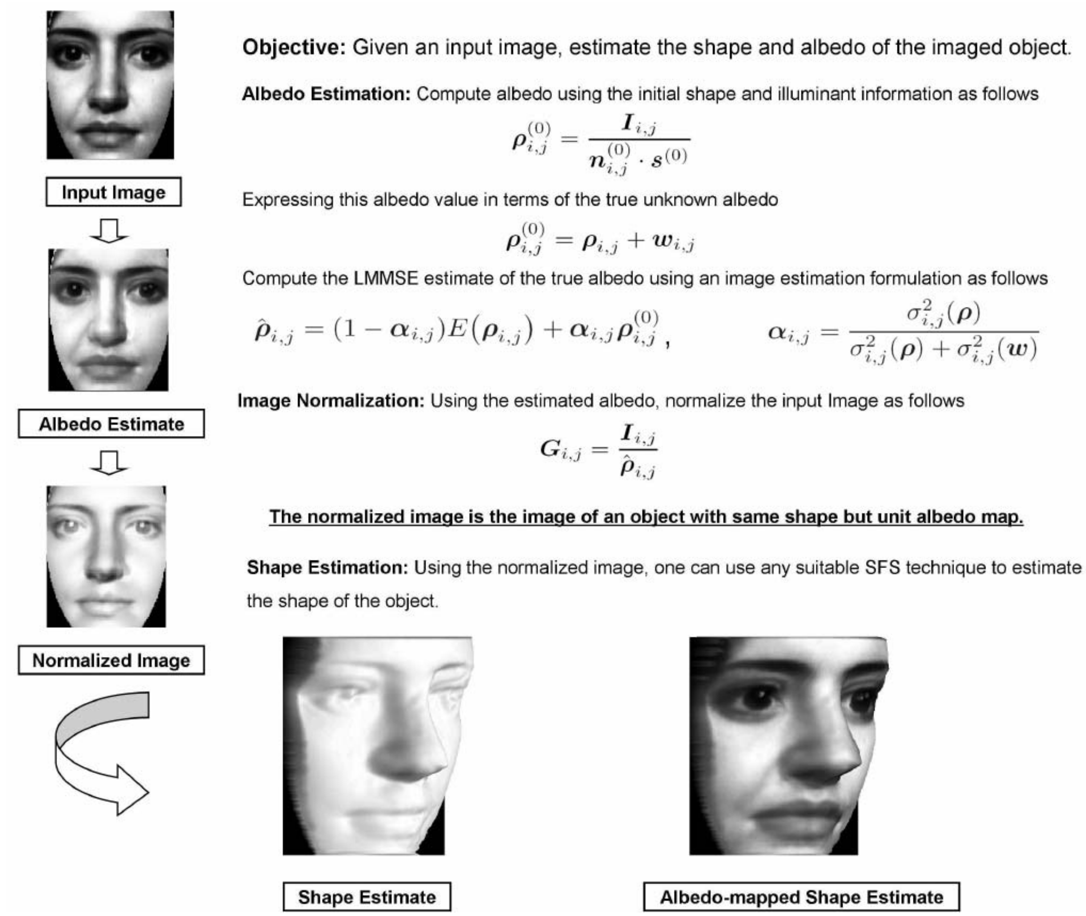
- Humans and machines asked to judge similarity of 80 pairs of faces (sureness of similarity ranked 1-5)
- 40 male, 40 female
- Faces deemed ‘moderately difficult’ – uncontrolled illumination

Phillips, Scruggs, et al., "FRVT 2006 and ICE 2006 Large-Scale Results", 2007

Open Problems/Issues

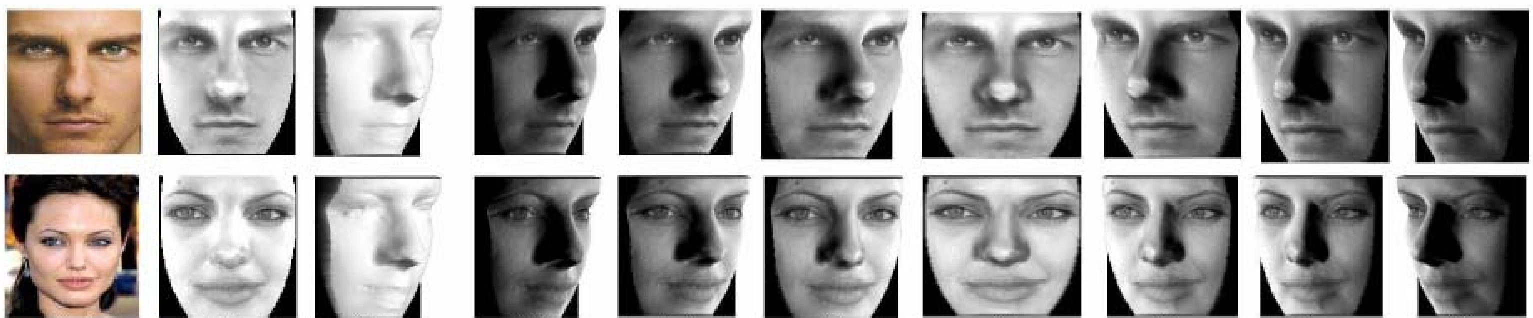
- Pose and Illumination
- Making best use of video
- Low resolution surveillance video
- Modeling what happens with age

Estimating Albedo and Shape



Biswas, Aggarwal, Chellappa, "Robust Estimation of Albedo for Illumination-invariant Matching and Shape Recovery", 2007

Example



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 10. Novel view synthesis in the presence of novel illumination conditions. (a) Input image, (b) Estimated albedo, (c) Recovered shape, (d)-(j) Synthesized views under novel illumination conditions.

Biswas, Aggarwal, Chellappa, "Robust Estimation of Albedo for Illumination-invariant Matching and Shape Recovery", 2007

Age progression challenges

- Wrinkles and baby fat are tough to model
- Approach today: use different models for different age ranges

Conclusion

- Viola-Jones appears to be more or less standard for detection
- Identification schemes are numerous
  - We seem to be moving closer to 3D shape/albedo/texture model
  - Video can help provide many poses/illumination/occlusion variances for same individual
- Computers as good as humans on high resolution frontal faces
- Challenges still exist

Discussion Points

- How much of face identification is a domain specific problem? Is it any wonder we see faces in the clouds, moon, etc?
- How about a cascade of features (Viola-Jones) for more tasks like identification?
- Seems likely different algorithms perform better for different applications and individual features, how about a generator to customize the algorithms for applications/individuals?