# Face Recognition by Humans and Machines

# A Tutorial Survey

CVPR'01 Short Course

Instructor:

Baback Moghaddam

Mitsubishi Electric Research Laboratories

# Note to Registrants

At press time these course notes were still in preparation and contributed material from outside was still trickling in. As a consequence this collection of slides is about 80% complete.

Final version of the course slides can be found at:

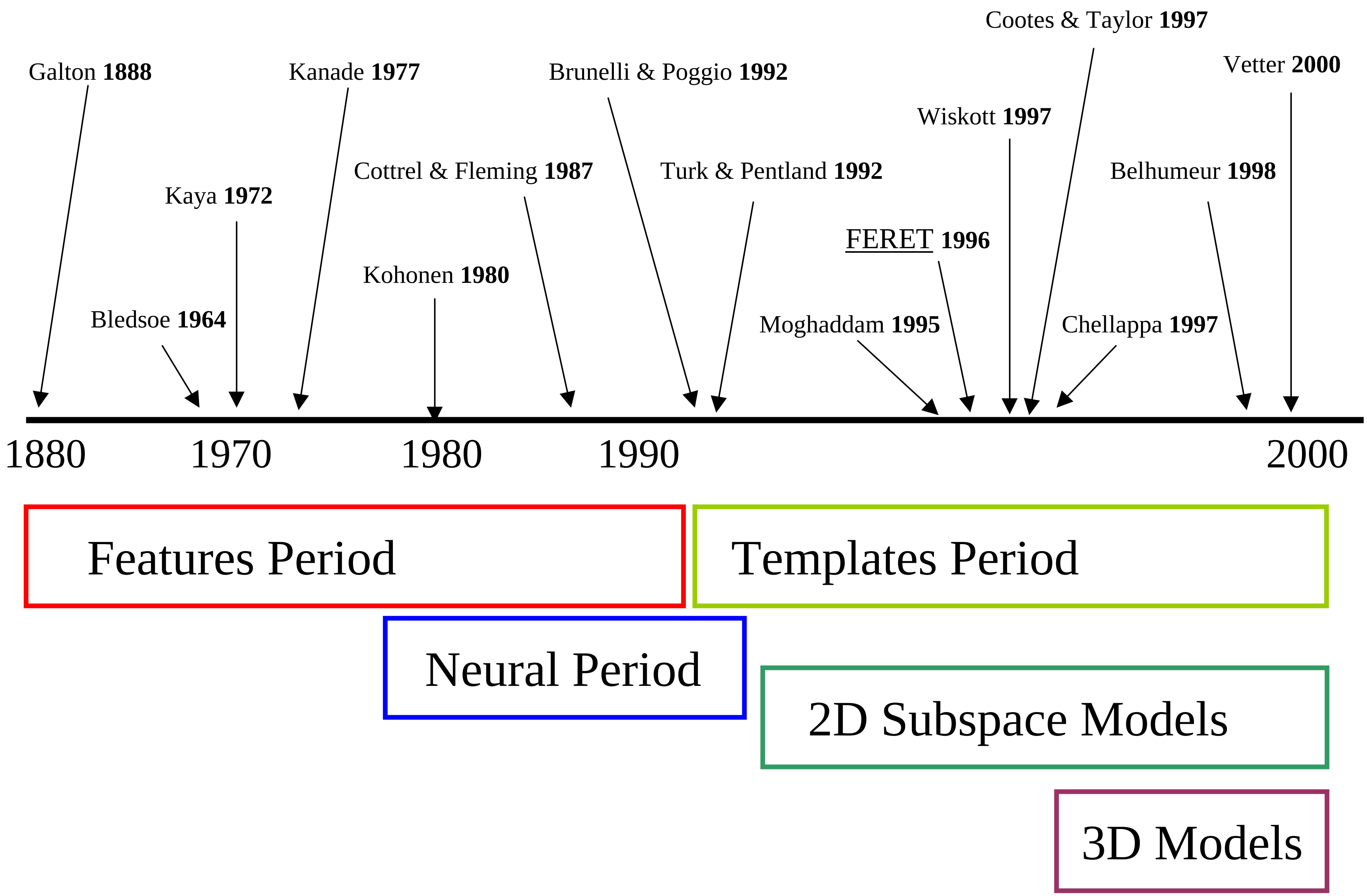
http://www.merl.com/people/moghaddam/cvpr01.html

If you have any questions contact me at baback@merl.com

# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# A Brief History (1900-2000)



# Sir Francis Galton (1822-1911)

- Face Research
  - “Personal identification and description,” Nature, 1888
  - “Numeralized profiles for classification and recognition,” Nature, 1910
- Eugenics Research
  - "Hereditary talent and character." (Macmillan's 1865)
  - *Hereditary Genius* (1869)
  - "The possible improvement of the human breed under the existing conditions of law and sentiment." (Annual Report of the Smithsonian Institution, 1902)

# Face Recognition Surveys

- **Samal & Iyengar**, “Automatic Recognition and Analysis of Human Faces and Facial Expressions,” *Pattern Recognition*, vol. 25, 1992

- **Valentin, Abdi, O’Toole & Cottrell,** “Connectionist Models of Face Processing: A Survey,” *Pattern Recognition*, vol. 27, 1994

- **Chellappa, Wilson & Sirohey**, “Human and Machine Recognition of Faces: A Survey,” *Proc. IEEE*, vol. 83, 1995.

- **Grudin,** “On Internal Representations in Face Recognition Systems,” *Pattern Recognition*, vol. 33, 2000

- **Zhao, Chellappa, Rosenfeld & Phillips**, “Face Recognition: A Literature Survey”, UMD CS-TR-4167, 2000

# Aspects of Face Processing

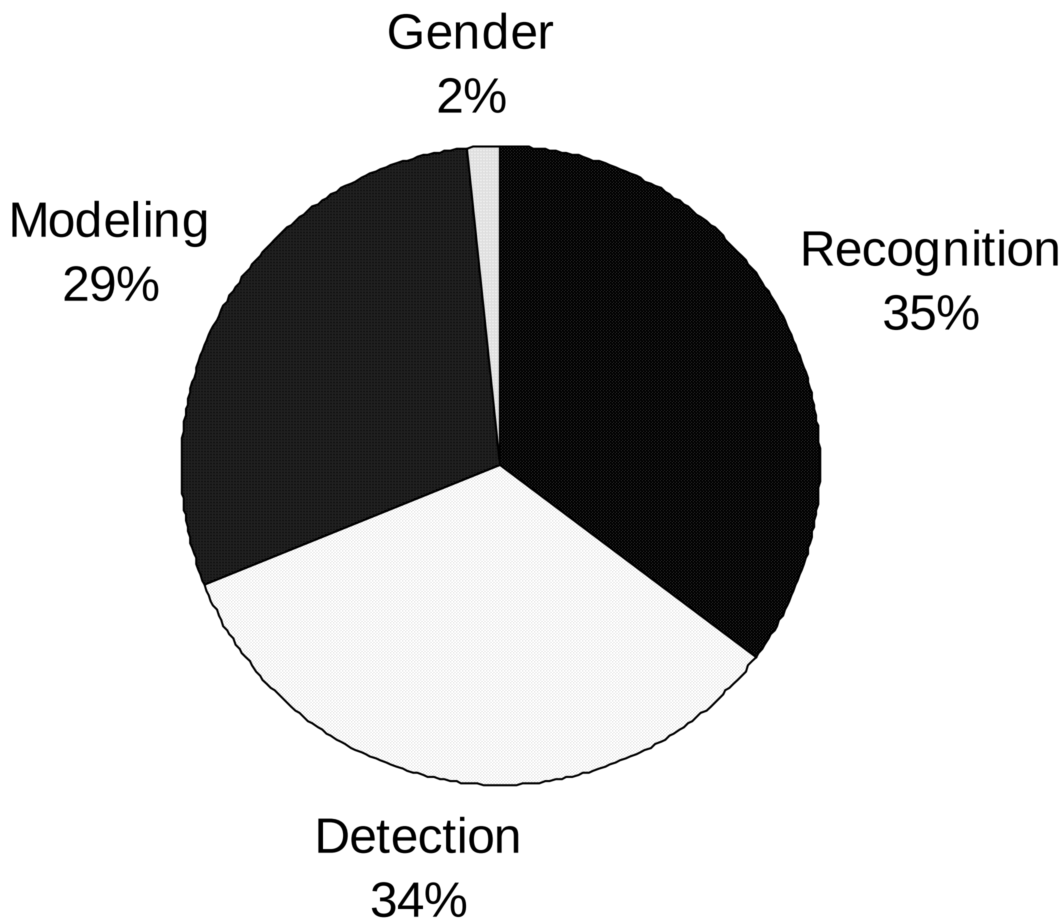
- Recognition
  - familiarity (membership)
- Identification
  - who is it? (assign identity label)
- Verification
- Classification
  - expression, gender, race, age, etc

# Applications of Face Biometrics

- financial transactions
- check-in or boarding planes
- crossing borders
- casting votes
- security or surveillance
- identity fraud
- criminal justice & law enforcement
- access to facilities, databases or privileged information, etc

# Face Publications by Category

(from F&G'95/96/98/2000)

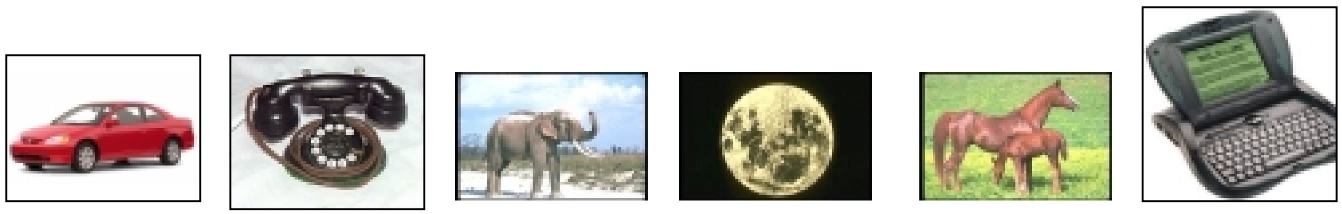


# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# Why Face Recognition is Easy!

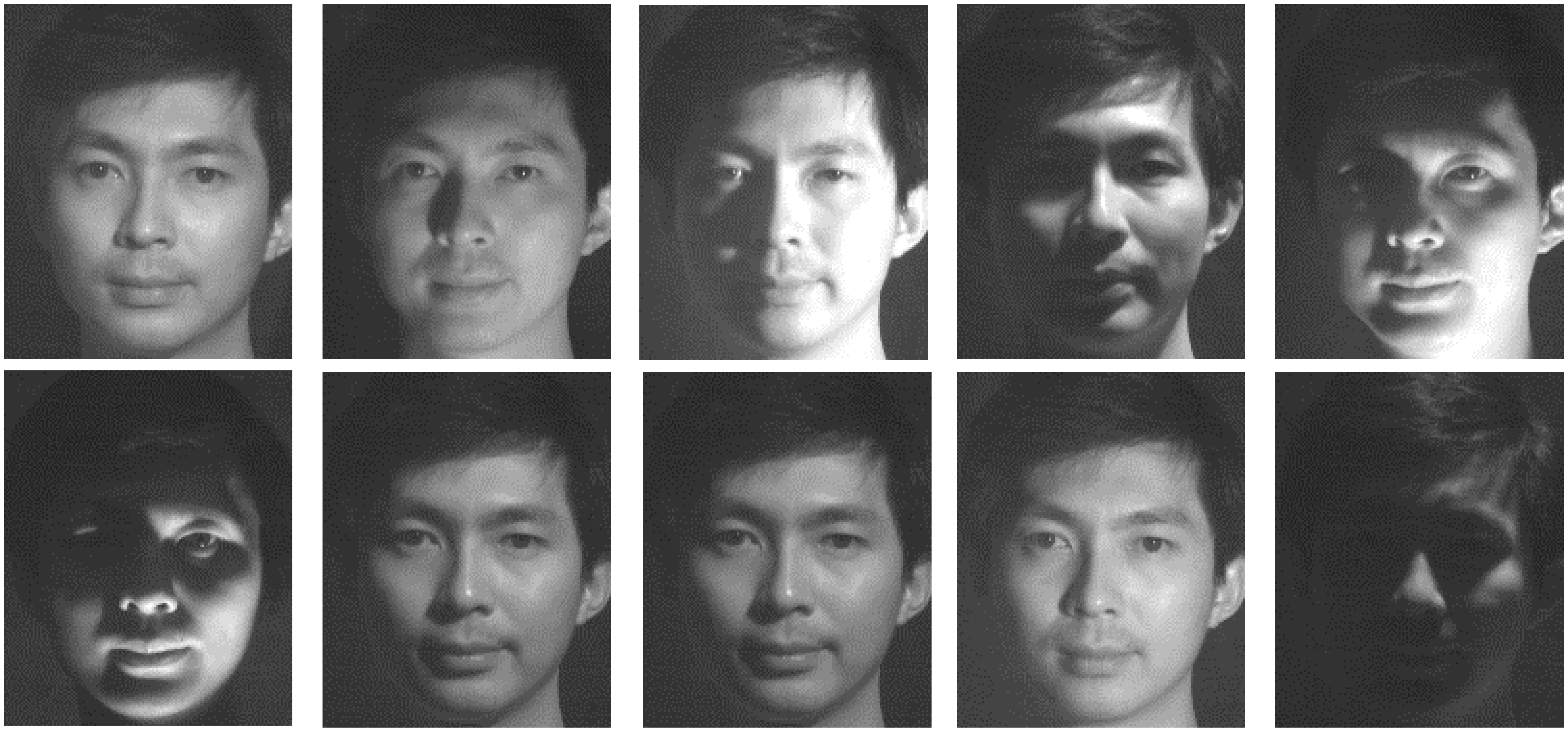
- It is **<u>not</u>** general object recognition!



- It is a single-class object recognition task

  – representation & matching can be optimized

# Why Face Recognition is Hard!



"The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity."

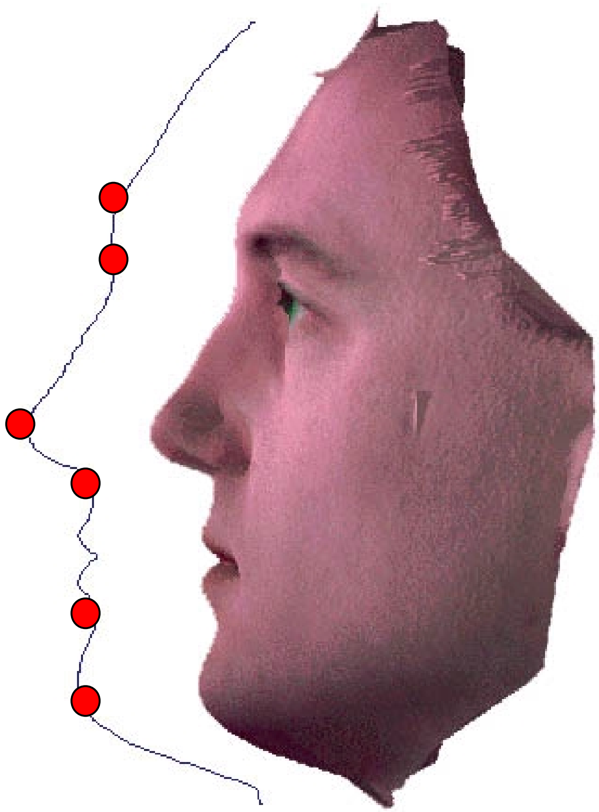
-- Moses, Adini, Ullman, ECCV '94

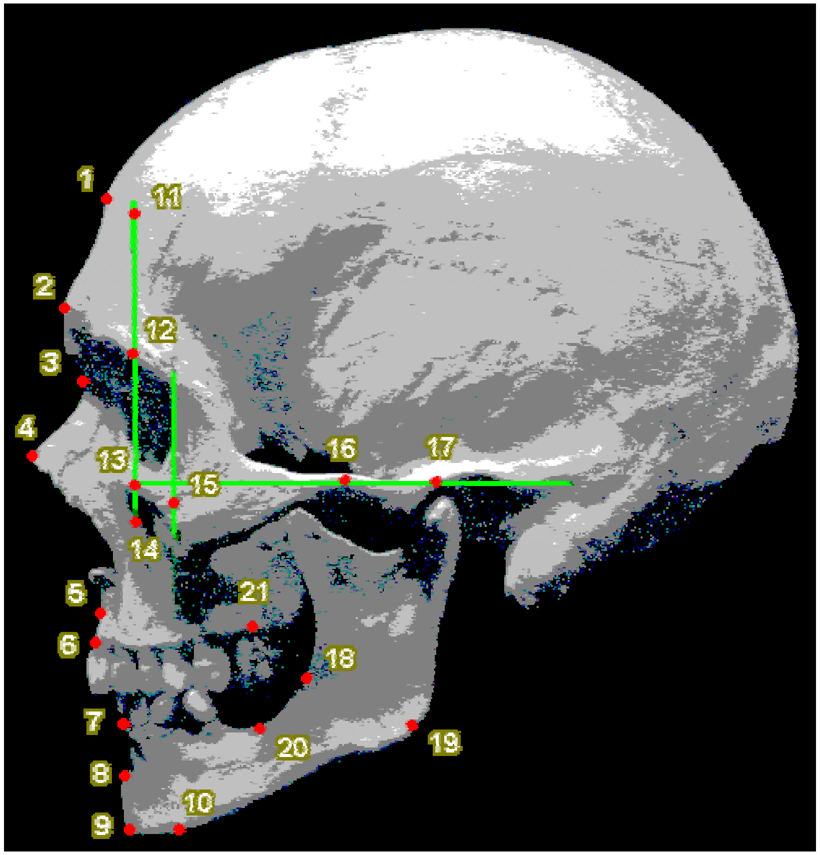
# Computational Face Models

- Feature-based
  - fiducial points
  - distances, angles, areas, *etc*
  - geometrical

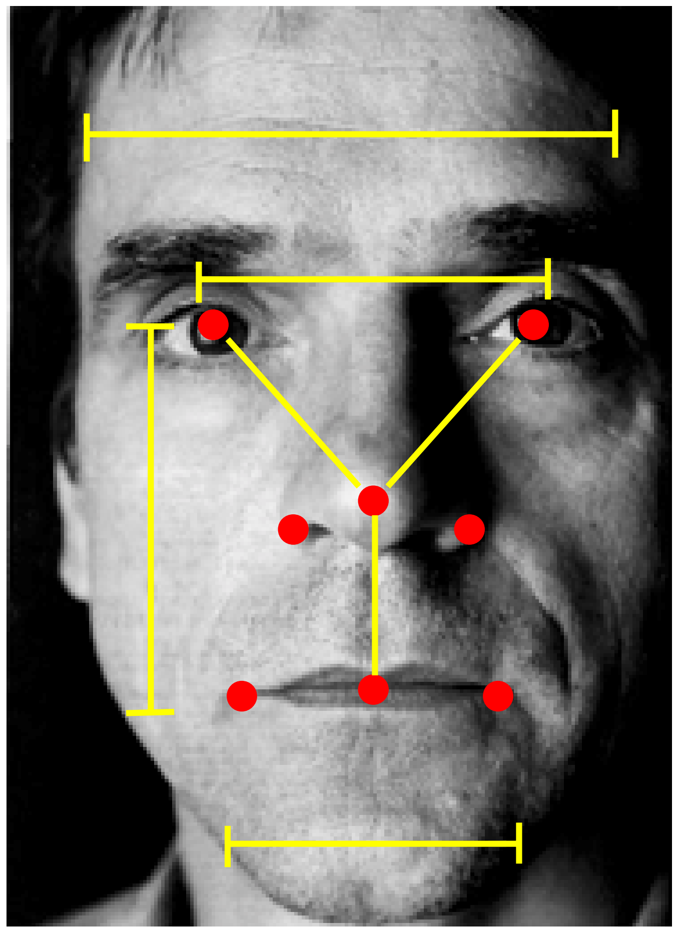
- Template-based
  - holistic
  - appearance based, images
  - statistical

# Features: Profile





# Features: Frontal



# Templates

## Whole Face

ORL database -- pose/expression

## Regions

Brunelli & Poggio (1993)

# Human Face Representation

- Feature-based or “configural”
  - [Roberts & Bruce 1988]
- Prototypes or “schemas”
  - [Goldstein & Chance 1980]

# Human Face Representation

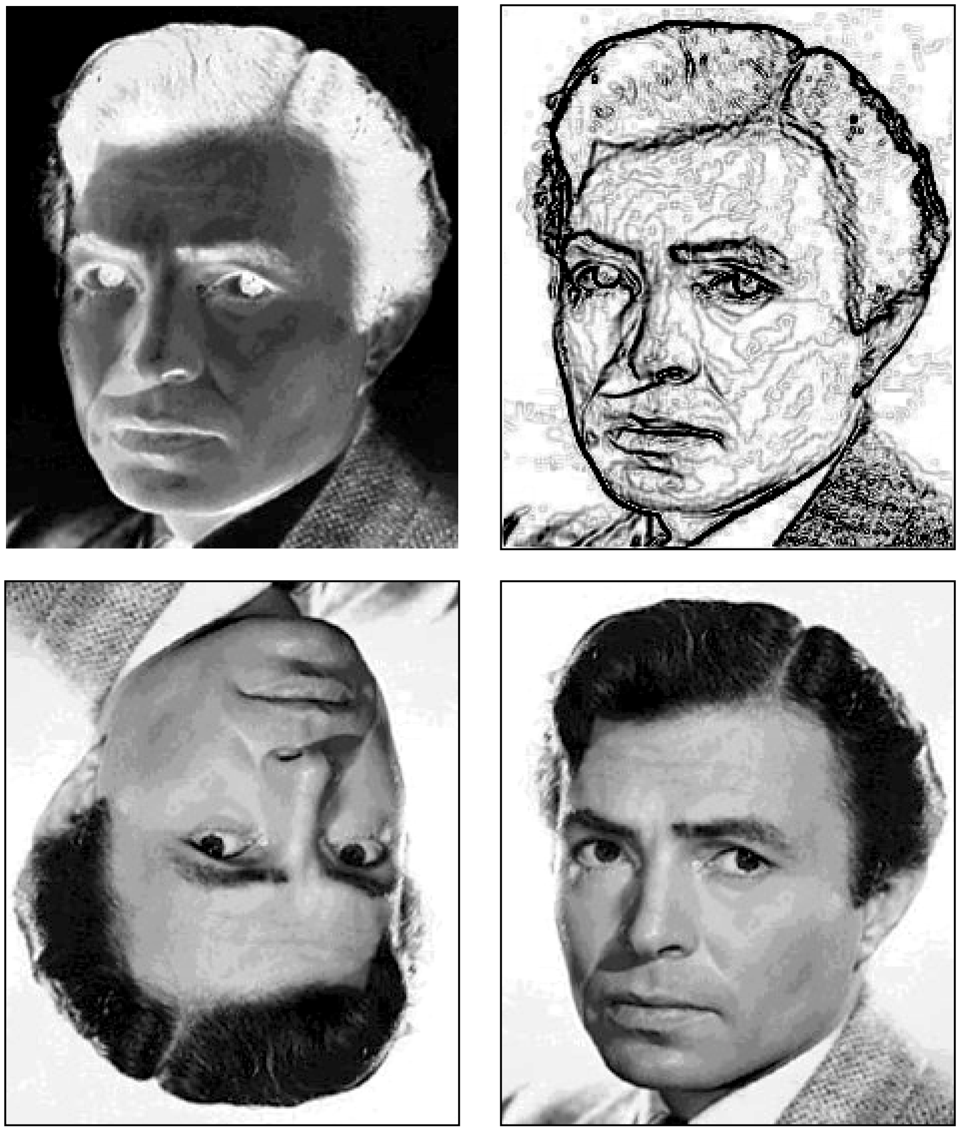
- PCA is a good model of human memory

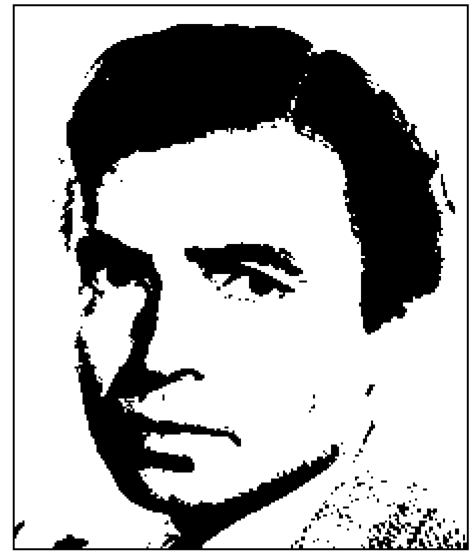
  [O'Toole et al 1994]

- Distinctiveness relates to recall ability

- Recognition is very hard with
  - line-drawings (with no shading)
  - luminance negatives
  - upside-down faces

# Impeded Face Perception

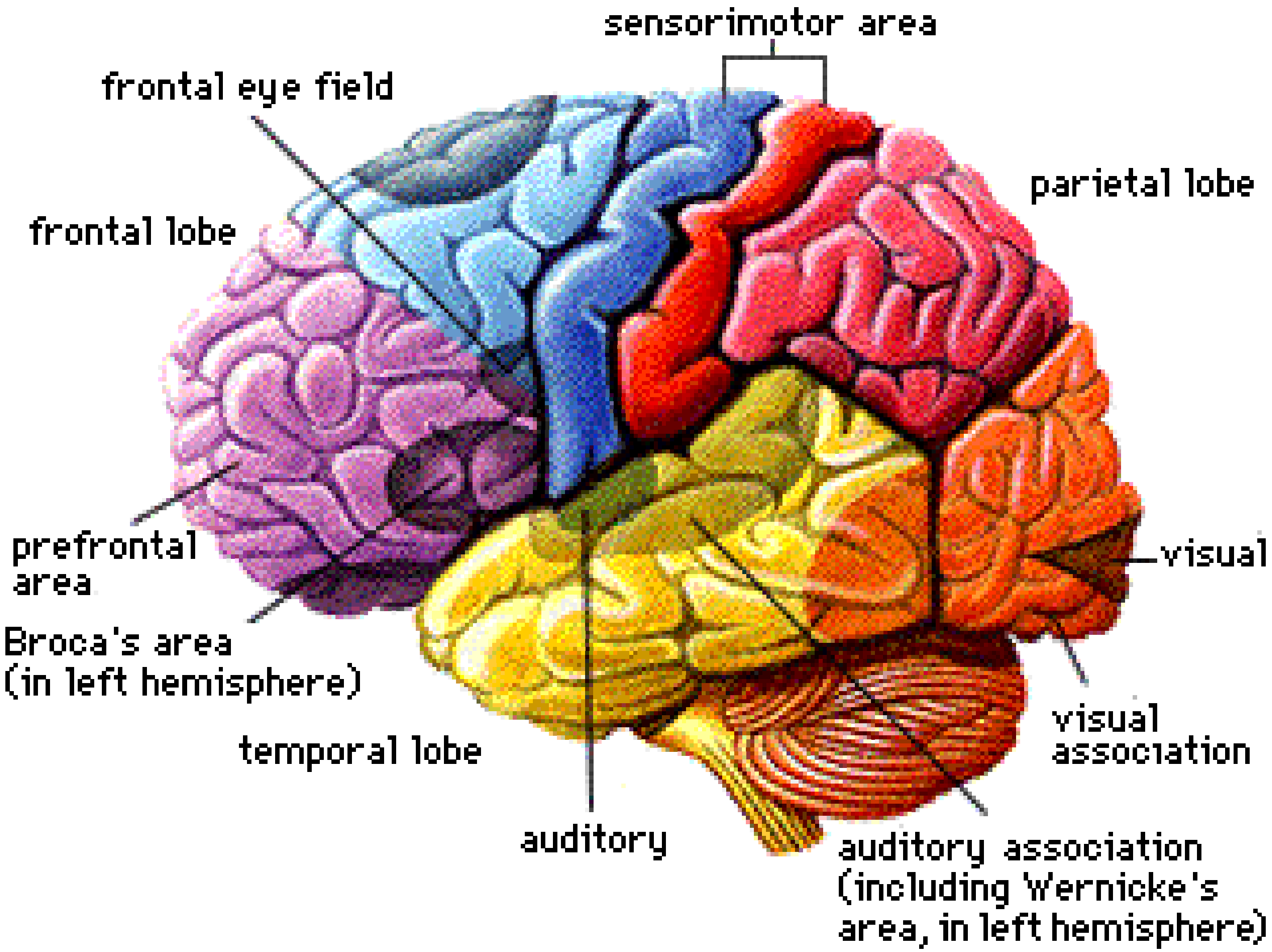




“Mooney” faces

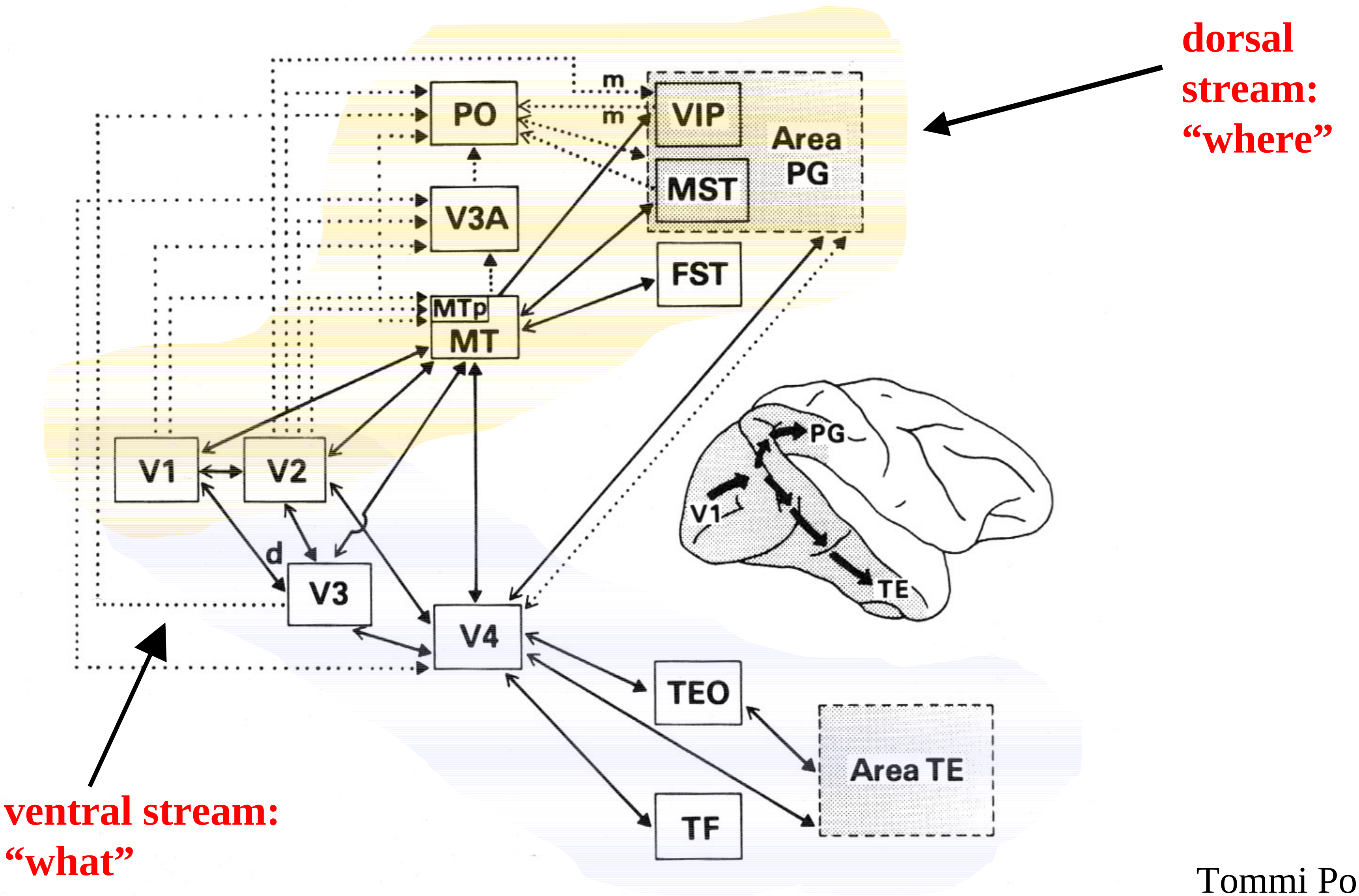
The “Face Inversion” effect (upside-down faces) has been used extensively as an experimental tool!

# The Human Brain



# The Visual System in Primates:

## Two Pathways: "What" and "Where"



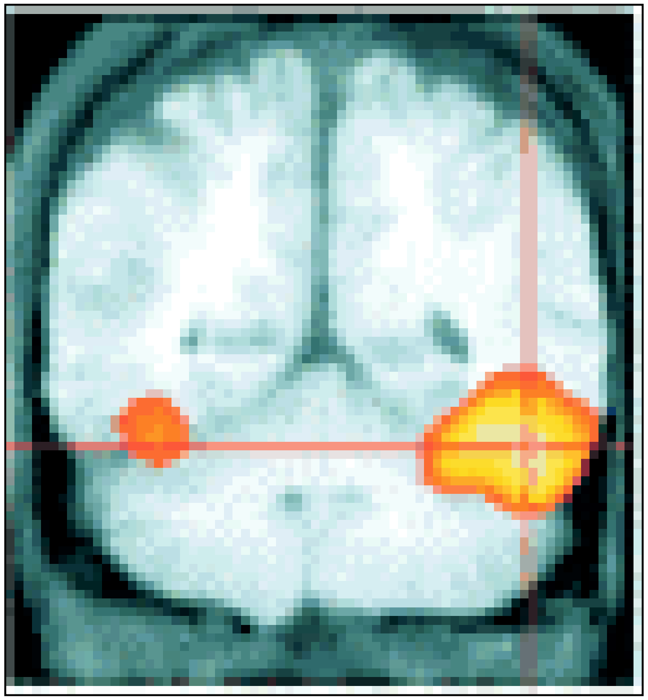

Tommi Poggio

# Hemispheric Specialization

• Right hemisphere is biased for face recognition

• Left hemisphere better at feature-based processing (less at holistic)

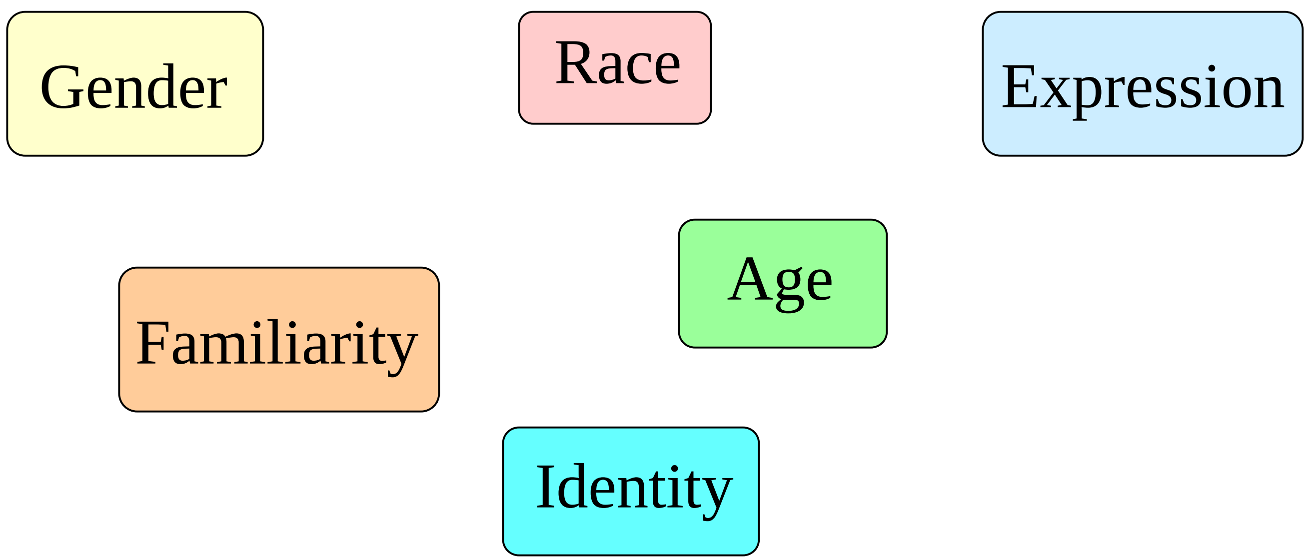
A particular brain wave (N200) occurs most strongly in fusiform regions of the right hemisphere when individuals view upright faces, but not when viewing inverted (or scrambled) faces (Allison et al., 1994).



**Activation of the right fusiform area (in the inferotemporal cortex) during face processing (Nakamura et al., 2000)**

# Face Perception in Humans

- Cortical localization in IT/STS [Desimone et al. 1984]
- Independent face modules [Bruce et al. 1986]

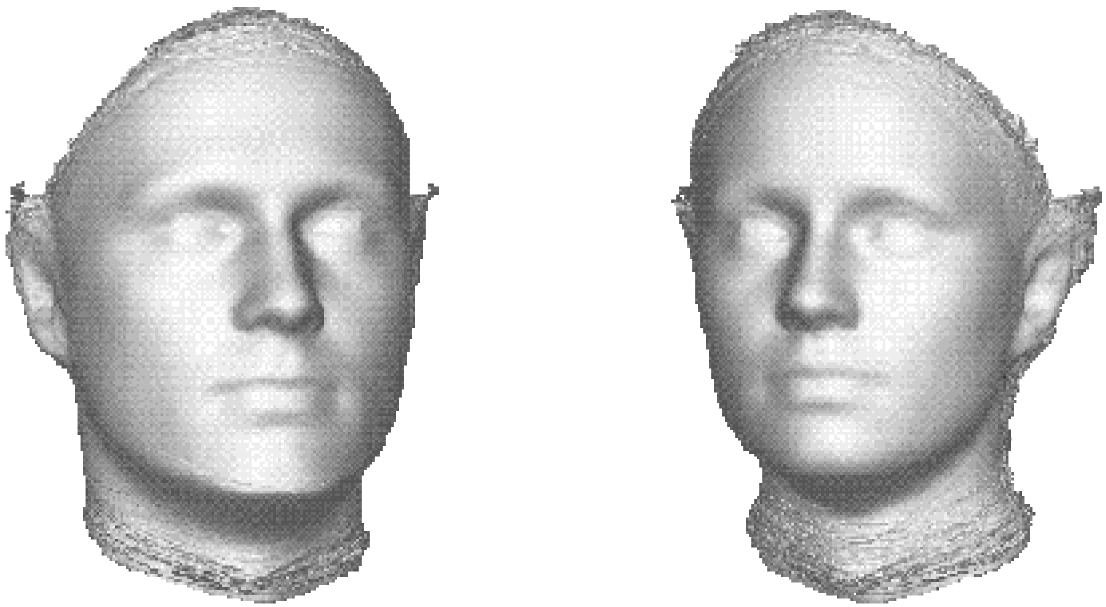


# Gender Prototypes

Images courtesy of University of St. Andrews Perception Laboratory

# Gender Shape Prototypes

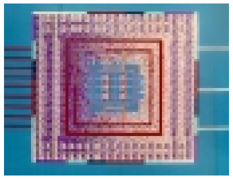
## O’Toole et al (1998)



*The average head plus versus minus the first eigenvector for the head surface data is shown. The analysis was performed on 65 female and 65 male heads. Individual face projections onto this eigenvector were highly correlated to the gender of the face.*

“Sex classification is better with 3D head structure than with texture. “
A.J. O'Toole, T. Vetter, N.F. Toje and H.H. Bülthoff, H. H. **Perception**, 26:75-84.

# Humans *vs.* Machines



| | | |
|---|---|---|
| Number of Elements | $10^{14}$ synapses | $10^{8}$ gates |
| Size of Elements | $10^{-6}$ meters | $10^{-6}$ meters |
| Power Consumption | 30 W | 30 W (CPU) |
| Processing Speed | 100 Hz | 1 GHz |
| Computational Style | parallel / distributed | serial / centralized |
| Fault Tolerance | yes | no |
| Learning Potential | yes | no |
| Intelligence/Consciousness | usually | not (yet) |

# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# Human Face Perception

*courtesy of*

Bradley C. Duchaine

Vision Sciences Laboratory
Department of Psychology
Harvard University

# Face Recognition

# Importance of Faces

- Identification—an ancient computational problem
- Emotion recognition—others provide information regarding their situations on the face.
- Other information available:
  - --locus of attention.
  - --facial attractiveness.
  - --age.
  - --sex.
  - --lipreading.
  - --personality types?

# How is Face Recognition performed?

There are a number of different possibilities:

- Domain-general object recognition procedures that operate on all objects.

- A face-specific procedure or procedures that operate solely on faces.
  - --specialized either phylogenetically or ontogenetically.
  - --If specialized ontogenetically, it must be the product of a domain-general procedure that develops expertise.

- A mixture of domain-general & face-specific procedures.

# Face Recognition Debate

For over 30 years, researchers have debated the scope of the mechanisms of face recognition—excellent example of the scientific process in psychology.

•Evidence from many sources.

•Debate has focused primarily on the question – "**Are faces special**?"

•In other words, do faces receive processing that differs from the processing applied to other objects and does this processing have a different neural substrate.

# Holistic and/or configural processing

Many have speculated that the special process is holistic and/or configural.

Holistic—integration of information from the whole face.

Configural—representation and recognition of the spatial configuration of the parts of the face.

# Processing of faces

- Generally, object recognition is considered to operate via feature recognition—in other words, recognition of objects based on the shape of its parts.

- This is difficult with faces, because they all share the same features.

- One solution could involve recognizing the relations between the parts of the face rather than the parts.

- Called configurational—computation of spatial config.
- Called holistic—integration over whole face.

# Sources of evidence in debate

- Cognitive experiments/Psychophysics
- Neuroimaging (fMRI)
- Evoked response potentials
- Single-cell studies
- Studies of neuropsychological cases.

# Cognitive & Psychophysical studies

- Many studies have compared whether faces appear to be processed differently than other objects by normal subjects.

- If there are different procedures used on faces this is powerful evidence that faces are special.

- However, stimulus characteristics could activate special procedures, so one must choose stimuli carefully.

- Inverted faces provide an ideal test case, because they are identical in terms within class similarity, complexity, and configuration.

# Yin’s studies (1969)

Yin initiated the face recognition debate with his studies of the effect of inversion on the recognition of faces versus other types of objects.

•Stimuli were unfamiliar faces, airplanes, houses, and stick figures—all typically seen only in upright orientation.

•Presented in an old/new discrimination paradigm
--Present a set of target stimuli.
--Test with target stimuli and distracters.

# Outcome of Yin's studies

Upright—Subjects recognized the faces best.
Inverted—Subjects recognized the faces worst.

Inversion dramatically decreased performance for faces, but only modestly decreased performance for other objects.

Yin speculated that holistic processing of faces led to their advantage in upright condition.  Within-class similarity drove poor inverted performance.

Similar results with dogs, houses, familiar faces, & unfamiliar faces.

# Diamond & Carey study (1986)

Questioned uniqueness of faces. Believed that it was not faces, per se, that evoked special processing.

Special properties of faces:

- Faces have the same parts & configuration—they are superimposable.
- People have great exposure to faces, and so may have developed expertise with faces.

Dogs are superimposable, and dog show judges have great exposure to dogs. Thus, they decided to compare the inversion effects for dogs for normal subjects and dog show judges.

# Diamond & Carey outcome

Predictions:

If superimposability & expertise drove the great face inversion effect, then dog show judges should show a large inversion effect while normal subjects will not.

Outcome:

- Both normals & judges showed a face inversion effect.
- Only judges showed a significant dog inversion effect.
- However, judges were no better than normals in the upright dog condition.  Normals showed no difference between upright & inverted.

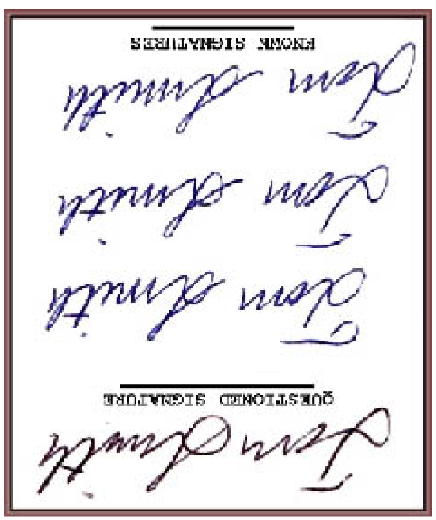
Diamond & Carey concluded that faces are not special, but such strong inferences do not seem warranted by the results.

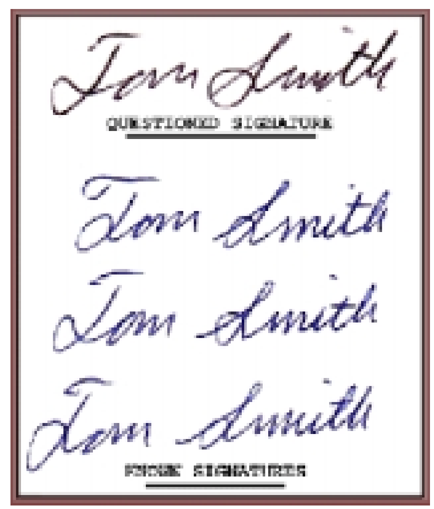
# Bruyer & Crispeels (1992)

- Ran a study similar to Diamond & Carey, but used handwriting as stimulus class.

- Compared normals & handwriting experts.

- Found large inversion effect only for experts.





# Interpretation of both studies

• It appears that large inversion effects can be found for non-face classes.  Thus, it may be that faces <u>and</u> other classes can be recognized with specialized procedures.

•The developmental origin, however, is unclear.  It seems certain that the expertise for dogs & writing develop ontogenetically.  However, the face expertise could develop ontogenetically or phylogenetically.

•The studies of the effect of inversion are equivocal.

# Chimeric Faces

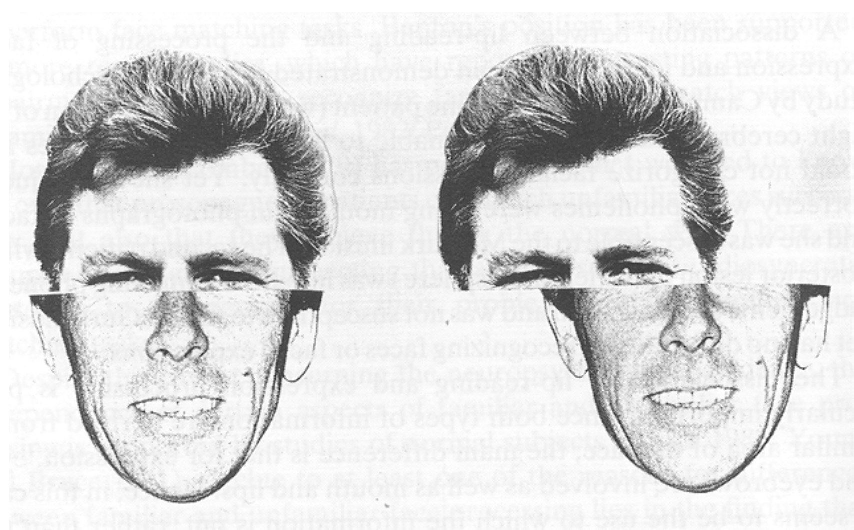
(Young et al. 1987)

A study using chimeric faces indicates the operation of holistic processing.

• Chimeric faces created by combining the top half of one individual's face with the bottom half of another.

• Halves could be aligned or misaligned.

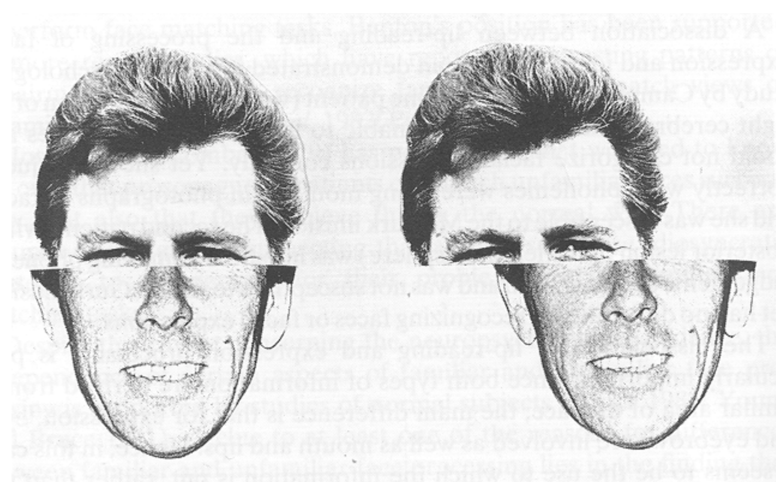
•Task: Recognize either only the top half or only the bottom half.



# Outcome with Chimeric Faces

- Upright performance significantly better for non-composites than for composites.
- Inverted performance unaffected by alignment.
- Inverted famous face recognition was faster than upright famous face recognition!

Interpretation: Interference from representation derived from whole face.
Holistic processing was restricted to upright faces.



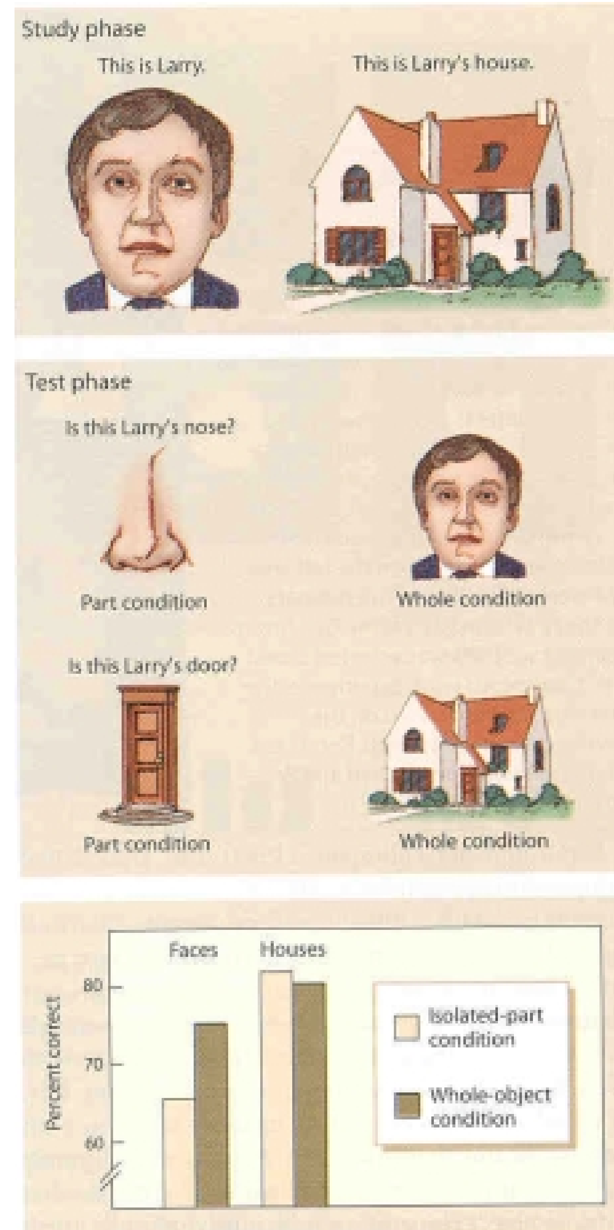
# Parts Recognition in Faces

Farah et al (1998) investigated the impact of the whole face on face part recognition. They compared this to house part, scrambled face and inverted face part recognition.

For non-face, recognition of parts was the same for the isolated and whole conditions.

In contrast, face part recognition was much worse in the isolated condition.

- Faces are represented more <u>holistically</u>.



# Further Part Studies

Tanaka & Sengco (1997) investigated whether it is the presence of any facial configuration or the study configuration that elevates performance.

| | Faces | | | | Houses | | | |
|---|---|---|---|---|---|---|---|---|
| | Features | | | | Features | | | |
| | Eyes | Nose | Mouth | *M* | Big Window | Small Window | Door | *M* |
| | | | | Experiment 3 | | | | |
| Isolated parts | 66 | 70 | 75 | 70 | 92 | 75 | 84 | 83 |
| New configuration | 79 | 71 | 76 | 75 | 92 | 76 | 85 | 84 |
| Old configuration | 83 | 76 | 85 | 81 | 91 | 72 | 85 | 83 |
| | | | | Experiment 4 | | | | |
| Isolated parts | 72 | 61 | 75 | 69 | 82 | 78 | 74 | 78 |
| New configuration | 75 | 70 | 78 | 74 | 81 | 81 | 74 | 79 |
| Old configuration | 87 | 78 | 88 | 85 | 81 | 81 | 76 | 79 |

Performance is best with the old configuration, so it appears to be the precise configuration seen rather than any configuration.

# Part Recognition Caveat

- Since this whole condition advantage was shown for faces, it shows that faces receive special processing.

- However, when subjects are trained to expert levels on Greeble recognition, they show a similar advantage for Greeble part recognition.

- Thus, an effect that was thought to be face-specific is not face-specific, but can be attributed to expertise.

Tarr et al (1998)

# Recognition via Configural or Feature Differences

- The different processes applied to upright & inverted faces was elegantly demonstrated by varying the distinguishing information available to subjects.

- Two types of faces were created:
  - --Configural faces.
  - --Featural faces.

Task: Decide whether two simultaneously presented faces were Same or Different.

# Outcome

Featural faces: Faces that differed in features but not configuration.

Upright = 91%

Inverted = 88%

There are no significant differences for these faces.

Configural faces: Faces that differed in configuration but not features.

Upright = 81%

Inverted = 55% (almost at chance)

# Interpretation

- Inverted faces do not appear to be coded configurally (or very coarsely).

- The feature recognition procedures that operate on upright and inverted faces appear to be the same.

# Evidence from Single-Cell Recordings

- Evidence from single-cell recording was critical in understanding the organization of lower-level visual areas.

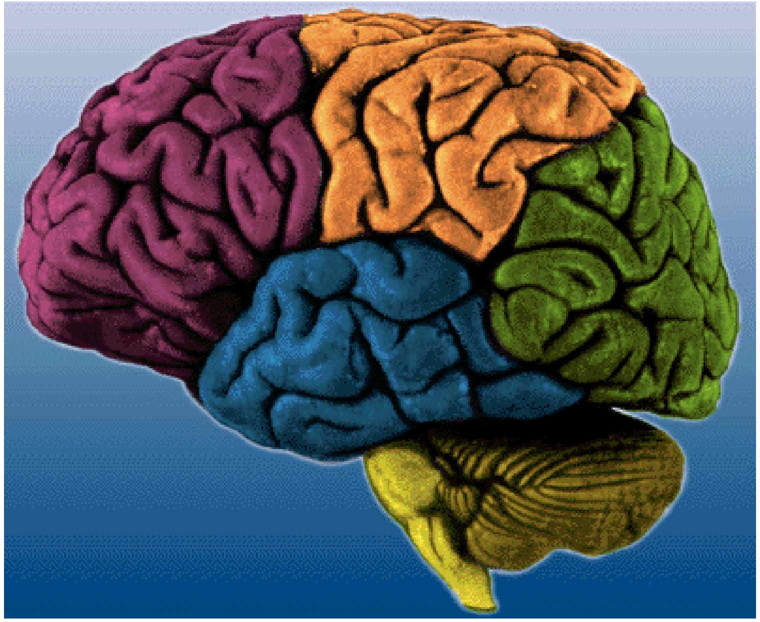
- Researchers have been exploring higher visual areas for the last 30 years in search of the neurons involved with object recognition.

# Many types of face cells

Face-specific cells in temporal lobe, frontal lobe, and the amygdala.

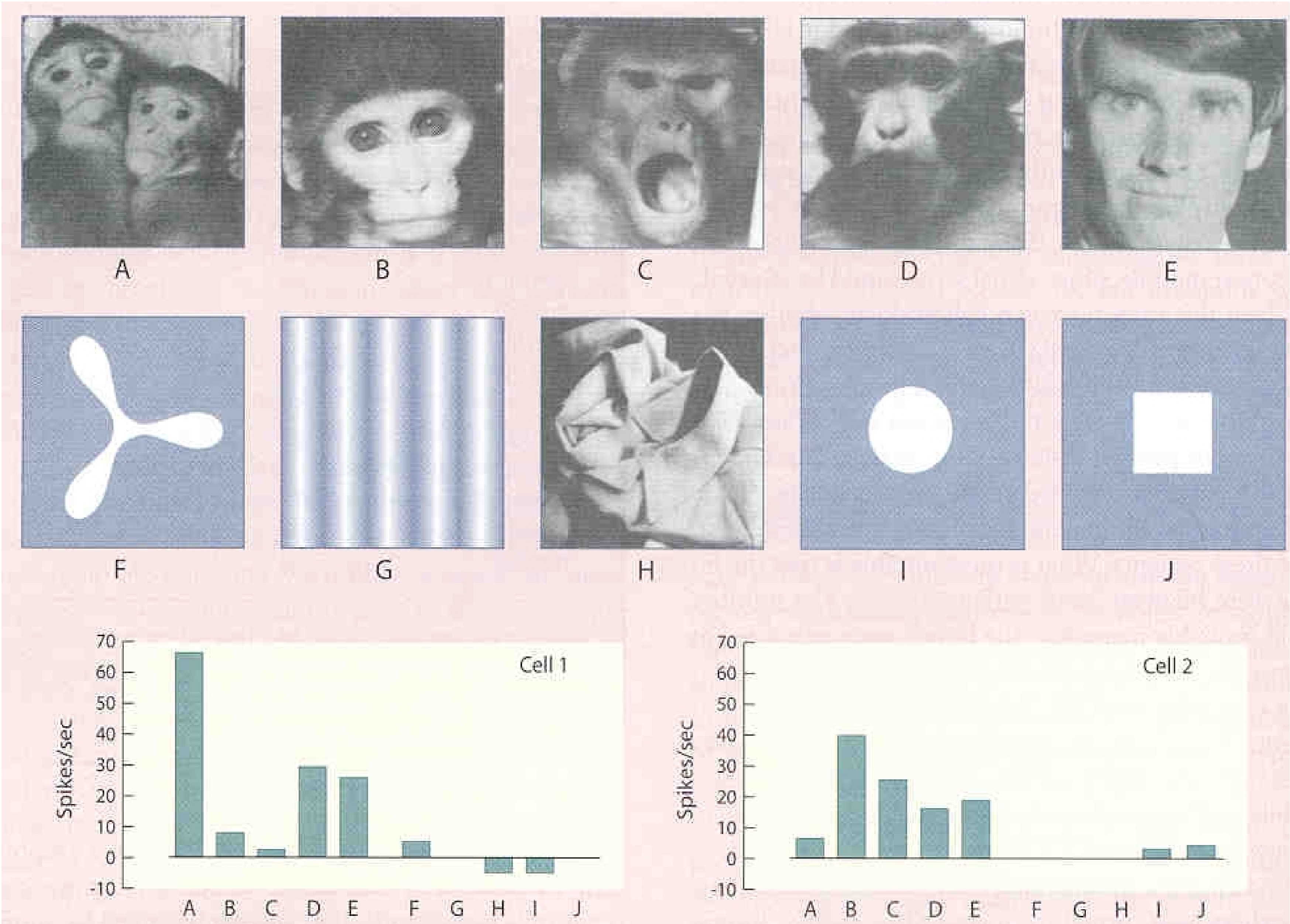
- Different cells selectively responsive to identity, pose, or emotional expression.

- Response properties appear well-suited for face recognition. Invariant over size, color, expression, lighting, etc.



- Large receptive field.

# Response of two face cells



# Cells in Infant Macaques

Rodman (1994)

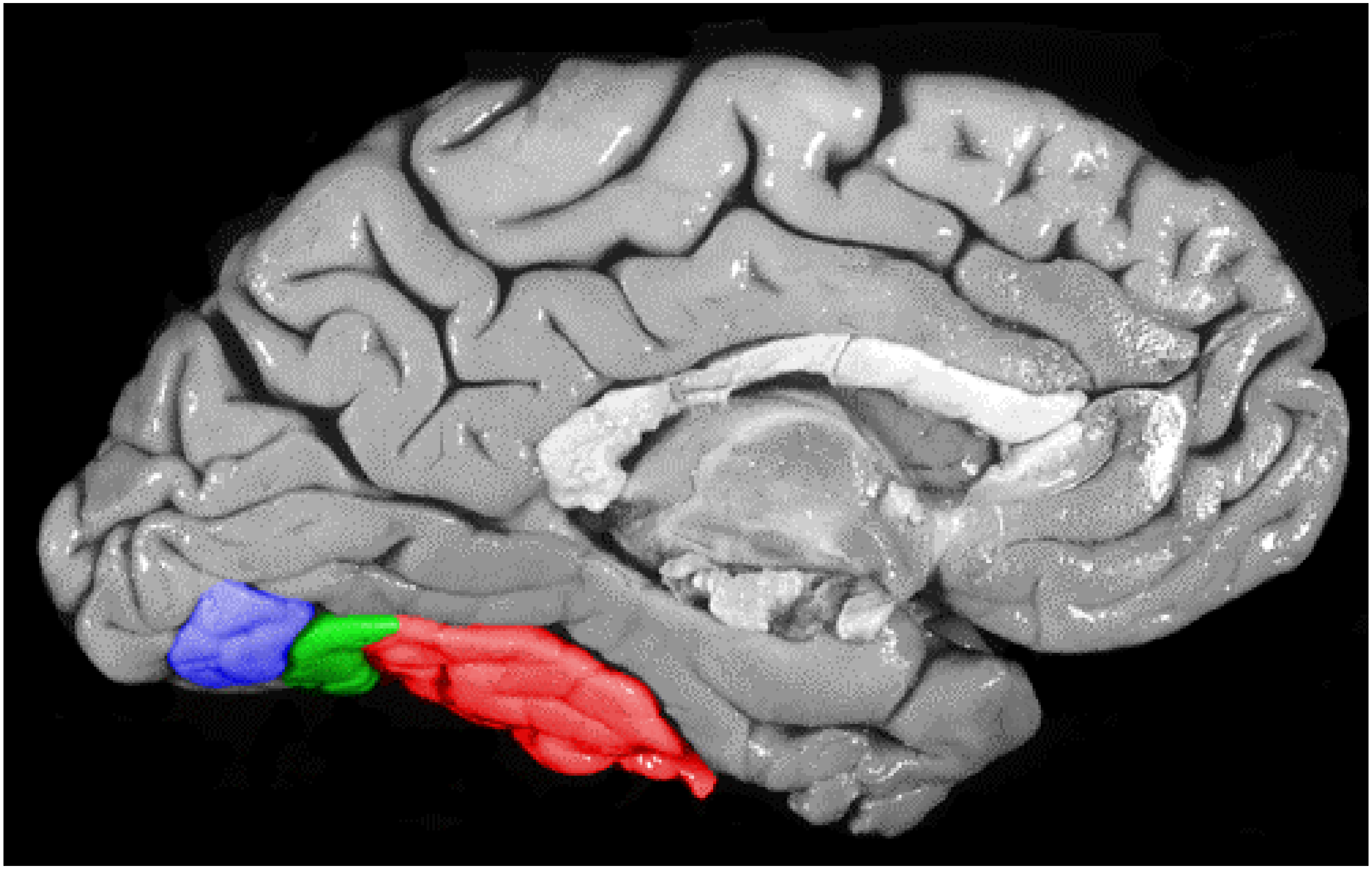
- Response properties in infant macaques are quite similar to adult neurons.
- Indications that some prewiring of these areas for face representation.
- However, these macaques were not isolated and so had seen monkey & human faces.

# Single-cell studies in Humans

Areas in the temporal lobe and prefrontal cortex have face specific cells. Vignal, Halgren (2000)

- Electrical stimulation of prefrontal areas produce face hallucinations. However, this region may be involved with fear recognition rather than identity recognition.

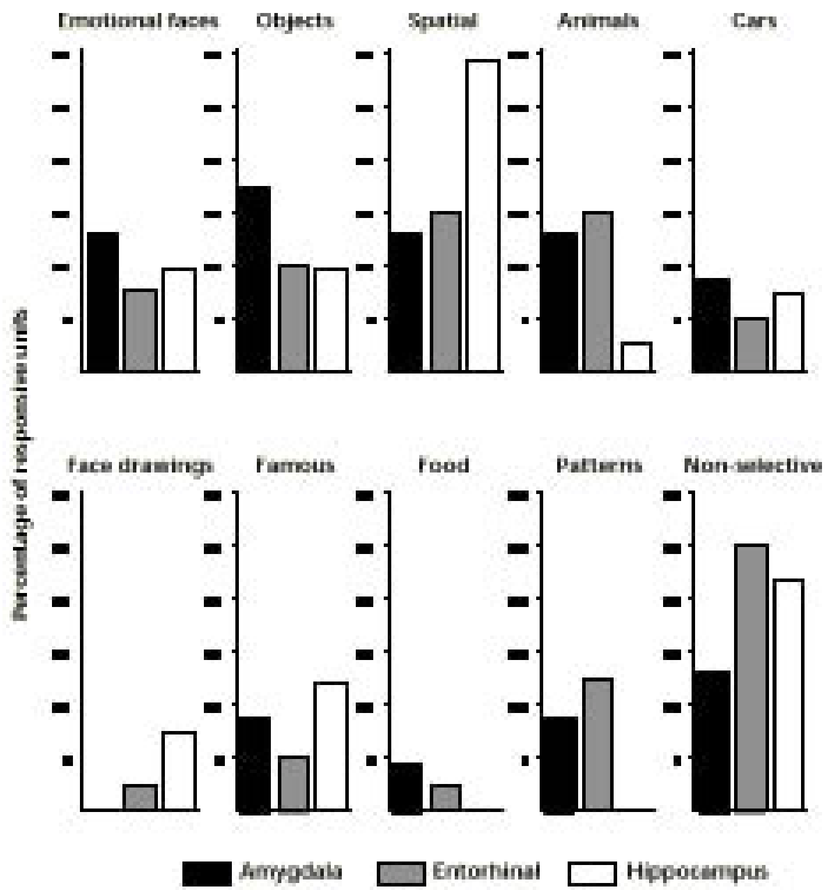
- Recordings from the fusiform gyrus and inferotemporal gyrus have shown regions that respond to faces but not cars, butterflies, or other control stimuli.
  - --Stimulation of this region produces temporary inability to name famous faces.



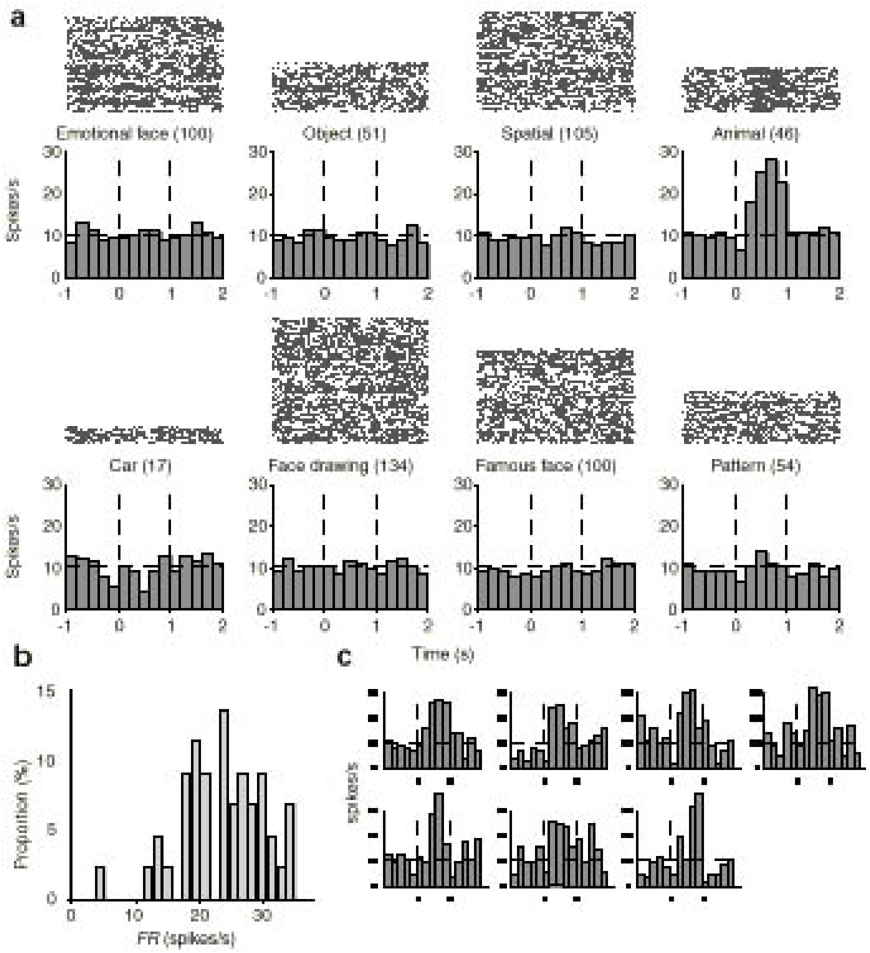
Fusiform gyrus can be seen in green.

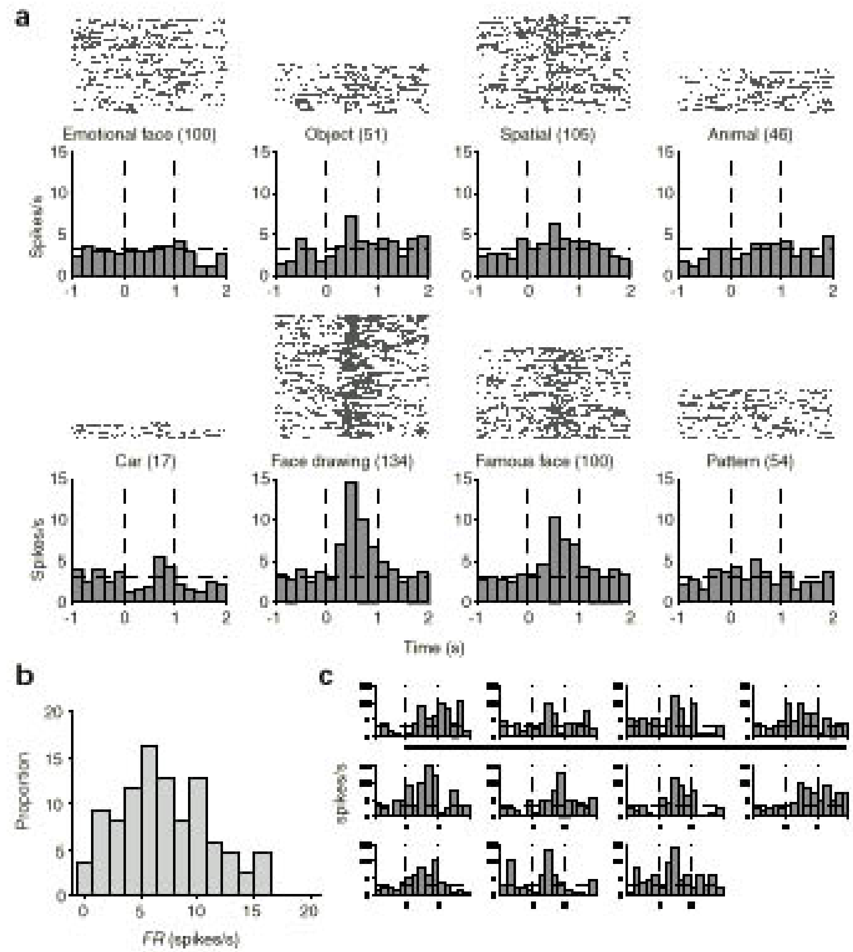
# Koch single-cell study (2001)

Recordings in the hippocampus, entorhinal cortex, and the amygdala have found many category-specific neurons.



# Category-specific cells




**Animal specific cell** **Familiar face specific cell**

# Implications of face cells

- Face-specific cells are expected if there are face-specific procedures, and as a result, they have been used to argue for face-specific processes.

- However, any recognized object must be recognized somewhere, so there must be cells responsive to every class.

- At present, we have little idea about specificity for other classes.

- Logothetis have found cells specifically responsive to an artificial stimulus class, and so specificity may be widespread.

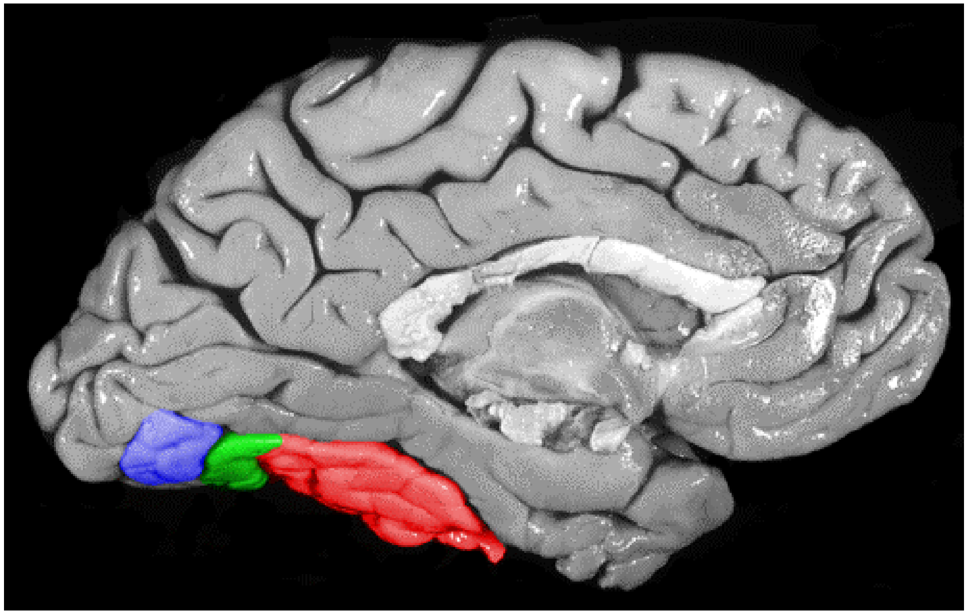
- As a result, single-cell recordings are unequivocal.

# Neuroimaging Evidence

In the last decade, many neuroimaging studies have compared activity in response to faces and other objects.

Main finding has been that an area of extrastriate cortex, a portion of the fusiform gyrus, is active in response to faces.

Green area in brain.



# Fusiform Gyrus

- Often referred to as Fusiform Face Area (FFA).
- Post-mortems of prosopagnosics have found lesions in the FFA.
- Same location as face-specific cell responses.
- Near areas involved with color vision & achromatopsia cooccurs with prosopagnosia.
- A prosopagnosic showed no activity in FFA in response to faces.

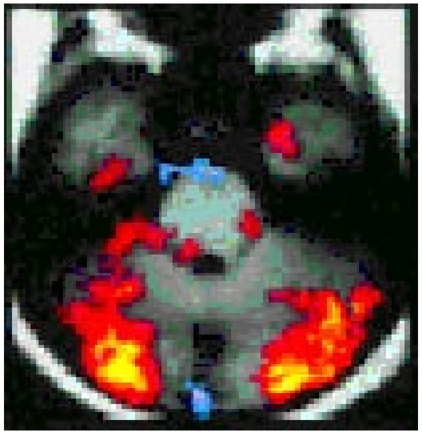
# What activates the FFA?

Kanwisher & her colleagues have probed the type of stimuli that activate this region.

- Strong response: Frontal shots, profiles, cartoon faces, inverted faces?!, inverted cartoon faces, cat faces, faces with no eyes, & eyes alone.

- Weak response: Schematic faces, animal bodies, houses, back of head.

- FFA appears broadly tuned.



# Neuroimaging, Binocular Rivalry, Faces, & Places

Basics of binocular rivalry: Tong & Nakayama (1998)

•In a display inducing binocular rivalry, a different image is provided to each eye.

•Rather than fusing the images, the visual system often cycles between images so that the observer is only aware of one at a time.

•Thus, by presenting two images in a rivalrous display, one can hold the stimulus constant while the contents of awareness vary.

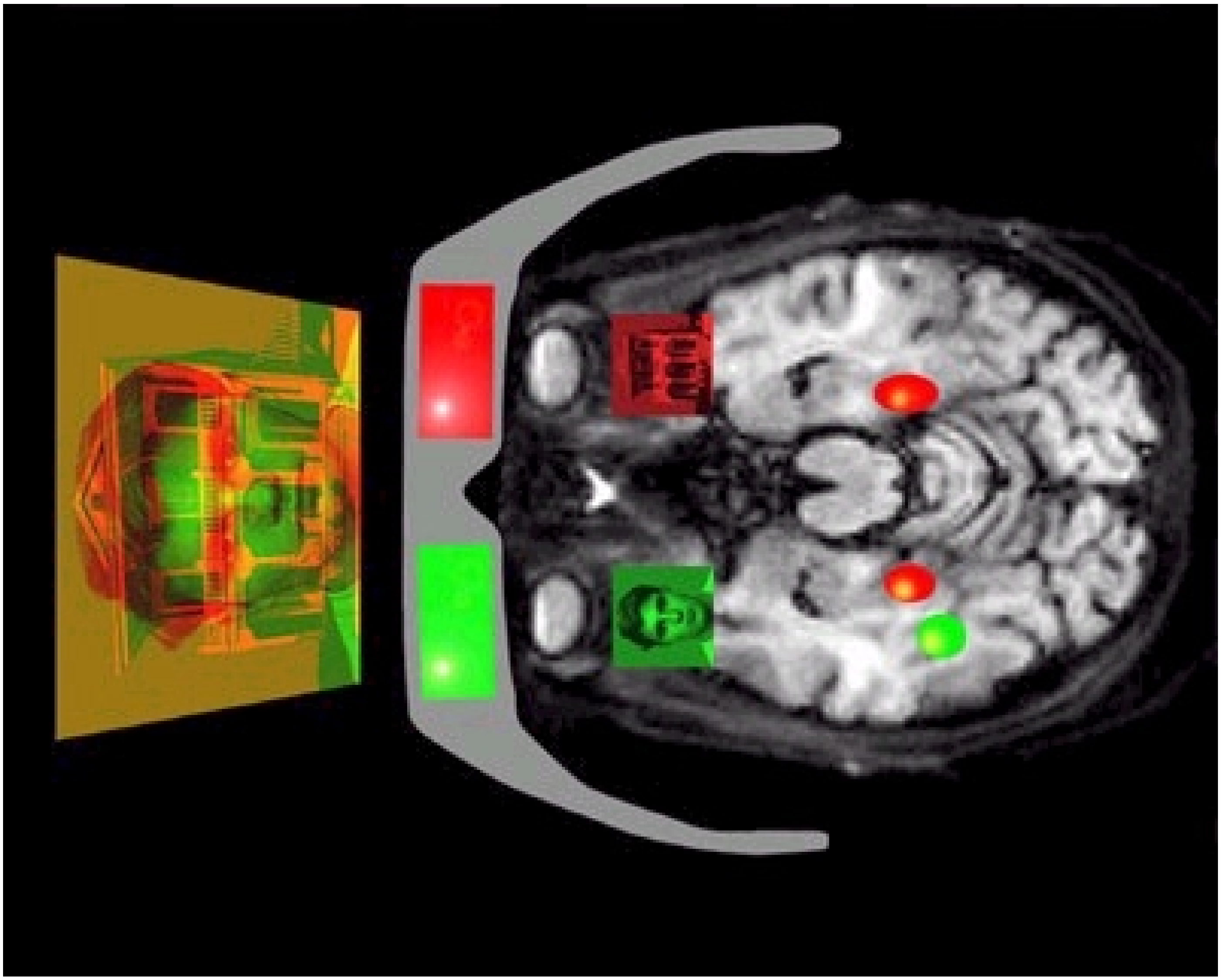
# Impact of Awareness

Rivalrous display presented houses to one eye and faces to the other eye.

Faces chosen because of consistent FFA response.

Places chosen because of consistent response in Parahippocampal Place Area (PPA).

Subjects were scanned and reported whether they saw a house or a face.



# Perception strongly correlated with Awareness

Perception of houses positively correlated with PPA activity while perception of faces positively correlated with FFA activity.

- Thus, more support for face-specificity.
- More interestingly, awareness appears to be necessary for activity in the FFA or the PPA. Presumably, there is some process occurring prior to processing in these areas that determines what reaches awareness.
- Doesn’t demonstrate that FFA or PPA activity is necessary or sufficient for awareness.

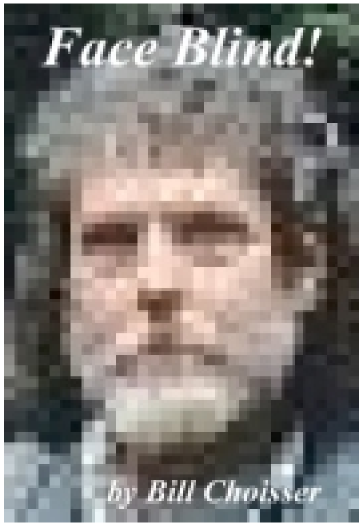
# Neuropsychological evidence

The most persuasive evidence regarding whether there are face-specific procedures has come from studies of individuals with neuropsychological conditions that selectively impair certain aspects of visual recognition.

- Cases of acquired damage are most common (stroke, head trauma, etc.)

- More cases of congenital impairments have documented recently.

- Many varieties of neuropsychological impairments affecting face recognition.



# Face Perception Disorders

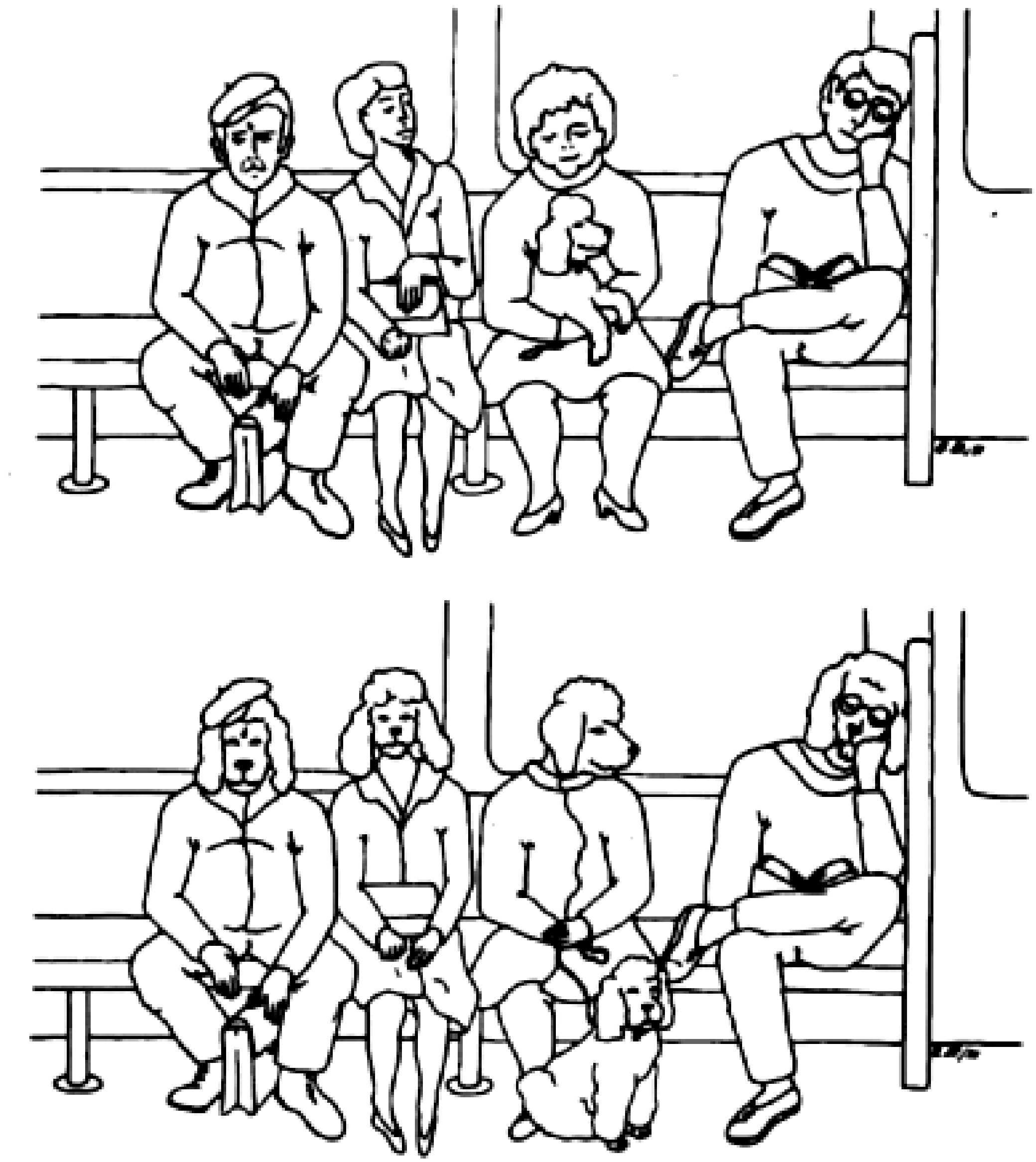
•Prosopagnosia, the inability to recognize faces, will be the main focus of our discussion.

•Metamorphosia—perceptual distortions affecting the face. ex—One woman noticed that the parts of faces did not appear to move together when the face moved.

•Palinopsia— “pasting” of a face onto the wrong head.

ex—one woman pastes faces associated with the face she is perceiving and gets faces “stuck” and pastes them on everyone.

# Poodle Face Palinopsia

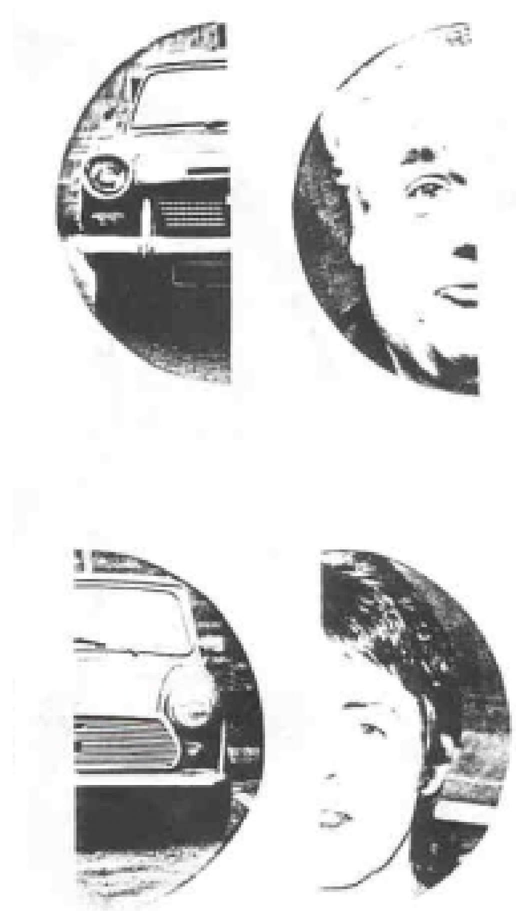


# Facial Neglect

Young, Andrew (1990)

•Neglect is most commonly seen for half of the visual field but sometimes it is restricted to objects.

•In facial neglect, one case reported neglected the left-side of faces (perceiver’s left) regardless of the orientation;
unimpaired on line bisection tasks, word reading, or car front recognition.



s, and half-objects used in identification task. Le
lf-stimuli in the bottom row. From left to right a
objects, asymmetric view half-objects, half-car-f
faces and inverted half-faces.

# Synthaesthesia & Delusional Misidentifications

- Synaesthesia typically refers to a condition in which percepts in one modality evoke percepts in another. ex.—The man who tasted shapes.

- Recent report of a patient who sees faces when seeing other objects.

---

## Delusional Misidentification Syndromes

- Capgras—believe that acquaintances have been replaced by replicas.
- Fregoli—believe that famous people are disguised as others.

# Characteristics of Prosopagnosia

• Often unable to recognize close friends, relatives, and self from facial information.

• Often, though not necessarily, accompanied by impairments in the recognition of other info from the faces such as emotion, gender, attractiveness, age, race, gaze direction, etc.

• Perception of the face is often compromised, but this may not always be the case.

• Acquired cases usually have bilateral damage, but it is sometimes restricted to the right hemisphere.

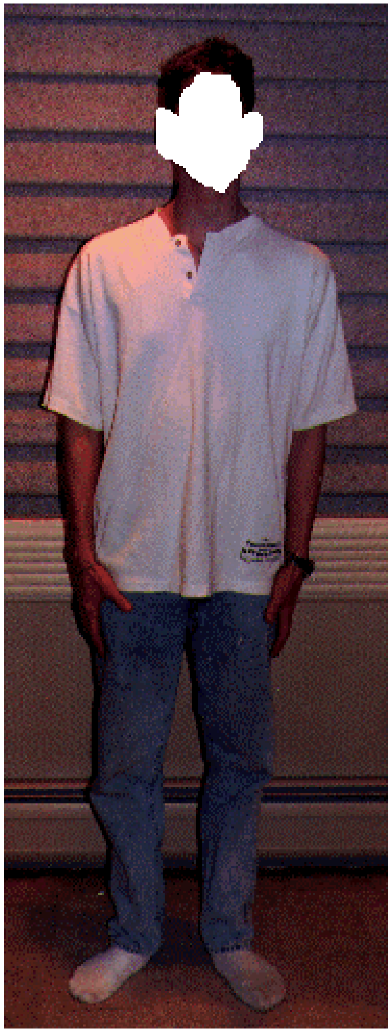
• Sometimes cooccurs with achromatopsia, autism/ Asperger syndrome, and topographagnosia.

• Prosopagnosics are usually able to discern identity via other channels such as voice, gait, hair, or clothing.

• Often develop expert recognition for other features.

  --BC uses hair, facial hair, and jeans.

  --Another uses sunglasses.



• Not surprisingly, face blindness often results in severe social problems.

# Explanations of prosopagnosia

There have been four major hypotheses to explain prosopagnosia. These hypotheses posit problems with hypothetical perceptual processes. Such processes are assumed to exist in normal subjects, but not operate normally in prosopagnosics.

As a result, evidence against a particular hypothesis is also evidence against a particular design for human object recognition.

# Four explanations

Each explanation proposes that prosopagnosia is a manifestation of damage or improperly development of a particular process. Here are the four proposals:

- Evolved face-specific procedures.
- Domain-general individuation recognition procedures.
- Domain-general configural processing procedures.
- Domain-general expertise procedures.

* Hypotheses 3 & 4 are not mutually exclusive.

# Evolved face-specific hypothesis

Claims that there are procedures that are activated by stimuli that have the geometric properties (or whatever other qualifying criteria may exist) of upright faces.

- For the system to be face-specific, and not domain-general, these procedures must be activated only by faces.
- Prediction: Prosopagnosia should sometimes be found without any other accompanying recognition deficit (pure prosopagnosia)
- There are no cases that can falsify this hypothesis, so testing must focus on falsification of the other hypotheses.

# Individuation hypothesis

• Proposes that there are procedures specialized for the recognition of individual items from within a category (Bob's face, Bob's car, Bob's wallet).

• Prosopagnosia is simply the most obvious manifestation of this impairment, because of the ubiquity and difficulty of face recognition.

• Many prosopagnosics do have agnosias for other types of objects.

• Prediction: prosopagnosics should show impairments with any recognition task that requires individual item recognition.

# Evidence & the individuation hypothesis

- There are a few cases that appear to show normal individual item recognition with impaired face recognition.

- There are a number of cases showing a dissociation between face recognition & place recognition, but we will not discuss these because place recognition may be performed by specialized procedures.  We are interested in face & objects at this point.

# De Renzi's patient

- 72-year-old lawyer who had a stroke.
- Unable to recognize individuals via their faces.
- Able to identify personal belongings from among 6-10 similar items.
- Could identify his handwriting from others.
- According to the patient & his wife, he only had agnosia for faces.

- <u>Problem</u>:  Difficult to compare to others; few categories tested.

# WJ:
# Human faces *vs* sheep faces

WJ suffered a stroke and became a sheep farmer due to his prosopagnosia.

He was severely impaired with human faces, but was better than controls with sheep faces.

Thus, his impairment is not an impairment with individuation.

# LH: Inverted Inverted Faces Effect

Other evidence indicates that upright & inverted faces are processed differently. This has been powerfully demonstrated in experiments with LH (damage due to a car accident).

- Normals: Upright Faces 94%; Inverted Faces 82%
- LH: Upright Faces 58%; Inverted Faces 72%

Although the stimuli are identical (aside from orientation), LH was much better with inverted faces! It appears that upright faces cannot be rerouted to the procedures that are performing recognition of the inverted faces.

# Inverted Inverted Shoe Effect

- LH’s inverted inverted face effect was considered very powerful evidence for the existence of face-specific procedures.

- However, this interpretation was undercut by an experiment showing that LH also showed an inverted inverted shoe effect.  He was better with inverted shoes than upright shoes whereas normal subjects showed a normal inversion effect.

- Are there specialized procedures for shoe recognition?

# Henke's subject

A German group tested two prosopagnosics on their identification of cars & fruit/vegetables.

•One of the patients showed normal performance on both tasks.

•<u>Problem</u>:  Not clear that this is individuation in the same sense that faces are.  Toyota or Apple versus OJ Simpson.

*Scores of W.F., M.T., and of Control Participants in the Exemplar Recognition Tests*

| | Case 1 (W.F.) | Case 2 (M.T.) | Control Mean/S.D. |
|---|---|---|---|
| Cars: Brand (max = 31) | 16 | 28 | 27.7/3.7 |
| Model (max = 31) | 6 | 21 | 20.6/6.4 |
| Fruit and Vegetables (max = 34) | 14 | 29 | 31.1/2.0 |

# Two problems

There are two problems with all of the reports of spared object recognition with prosopagnosia.

•Response time was not measured for any of these tests.

--Possible that prosopagnosics were able to achieve normal accuracy performance due to especially long latencies.

--Measurement of RTs can rule out speed/accuracy trade-offs.

•Measures of accuracy were not bias-free.

--Discrimination may have been normal, but varying criterions make such measures questionable.

# Double Dissociation

- Double dissociation:  report of at least two different neuropsychological cases with opposite patterns of normal and impaired abilities.

- Double dissociations are powerful evidence for the existence of two independent mechanisms for the two tasks.

- For example:

  --Bill Choisser—prosopagnosic with normal object recognition (so far)

  --CK—object agnosic with normal face recognition.

- Difficult for the individuation hypothesis to explain.

# Bill Choisser

- •Born in 1946.
- •No history of head trauma.
- •Recognized his prosopagnosia in his late 40s.
- •Produced a very detailed web site about prosopagnosia (www.choisser.com/faceblind)
- •Recognizes people using hair, facial hair, & jeans.
- •Very intelligent (IQ 131), MIT student, lawyer, engineer.
- •Reports no other visual difficulties—no topographagnosia.
- •Other neurological problems: CAPD & slight motor problems.
- •Genetic root of CAPD & motor issues.
- •Family stories indicate presence in males on paternal side.

# Bill's Low Level Abilities

| BORB tasks | Controls | BC |
|---|---|---|
| Length Match | 26.9 (1.6) | 29 |
| Size Match | 27.3 (2.4) | 29 |
| Orientation Match | 24.8 (2.6) | 27 |
| Position of Gap Match | 35.1 (4.0) | 38 |
| Overlapping Figures | | Normal |

| Kit Tasks | | |
|---|---|---|
| Hidden Figures | 14.0 (6.0) | 23 |
| Hidden Patterns | 148 (38) | 193 |
| Copying | 25.9 (9.3) | 52 |

Thus, he does not appear to have any low-level deficits that contribute to his recognition difficulties.

# Bill’s Face Recognition

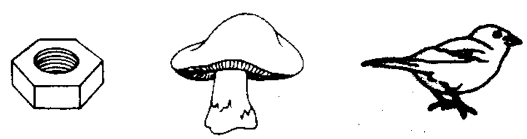
| Test | Controls | BC |
|---|---|---|
| Profiles | 27.1 (1.3) | 24(z = -2.53) |

| One in Ten | | |
|---|---|---|
| d’ | 3.61 (.49) | 2.15 (z = -3.0) |
| Yes RT (msec) | 774 (121) | 1399 (z = 5.17) |

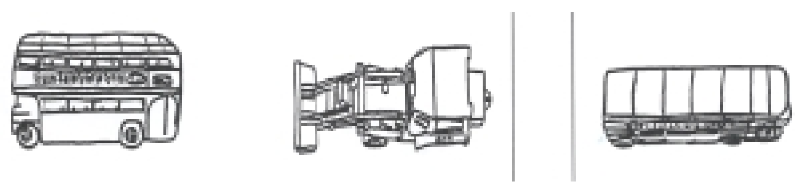
| Famous People | | |
|---|---|---|
| Famous Faces | 23.6 (1.4) | 6 |
| Famous Names | | 25/25 |

# Bill's Object Recognition

| | Controls | BC |
|---|---|---|
| S & V drawings | | 256/259 |



| | Controls | BC |
|---|---|---|
| Minimal Feature Match | 23.3 (2.2) | 25 |

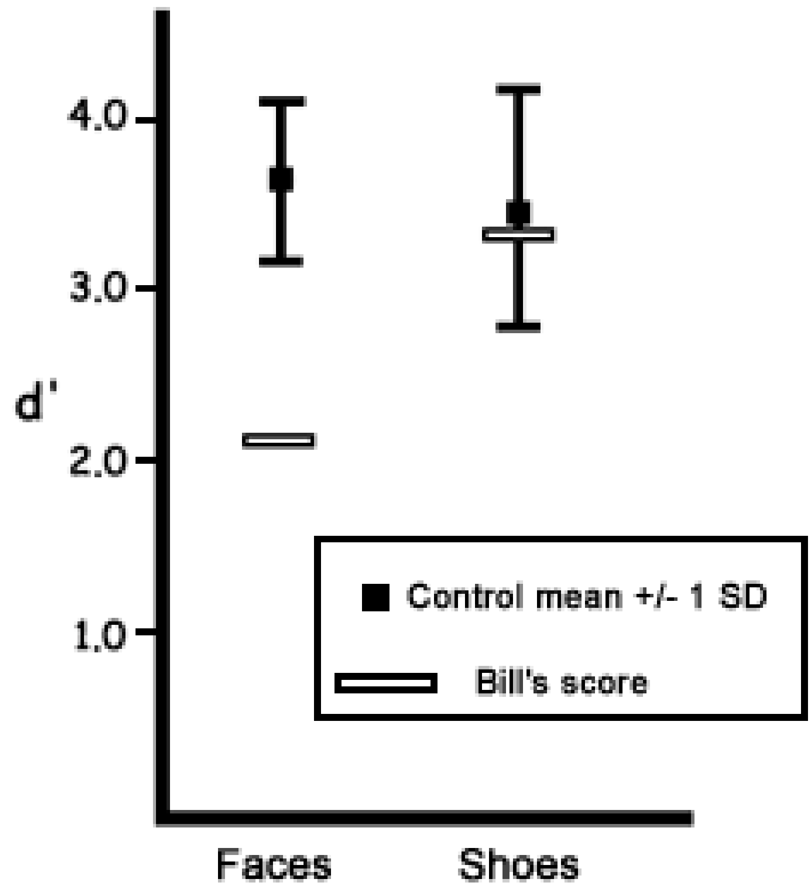


| | Controls | BC |
|---|---|---|
| Foreshortened Match | 21.6 (2.6) | 25 |

| | Controls | BC |
|---|---|---|
| Object Decision Task | 27.0 (2.2) | 26 |

# Face OIT vs. Shoe OIT

Bill manifested an impairment on the Face One in Ten task.  The individuation hypothesis predicts that he should show similar impairments with shoe recognition.



# Face OIT vs. Shoe OIT

| | Controls | Bill | Controls | Bill | Controls | Bill |
|---|---|---|---|---|---|---|
| | d' [-1 s.d., +1 s.d.] | d' | RT--Y (msec) [-1 s.d., +1 s.d.] | RT-Yes | RT--No [-1 s.d., +1 s.d.] | RT--No |
| Face OIT | 3.61 [3.13, 4.10] | 2.15 p = .0067 | 774 [653, 895] | 1399 p = .0002 | 530 [443, 617] | 1012 p = .0001 |
| Shoe OIT | 3.46 [2.77, 4.15] | 3.32 | 1583 [939, 2227] | 1280.789 | 812 [601, 1024] | 733.4724 |

Bill shows no impairment in discrimination or RT with shoes.

# Face Old/New Discrimination

Does Bill show impairments on old/new discrimination tasks with faces?

| | Controls | Bill | Controls | Bill | Controls | Bill |
|---|---|---|---|---|---|---|
| | d' [-1 s.d., +1 s.d.] | d' | RT--Hits (msec) [-1 s.d., +1 s.d.] | RT-Hits | RT--CR [-1 s.d., +1 s.d.] | RT--CR |
| **Faces #1** | 3.06 [2.51, 3.61] | 1.79 p = .021 | 962 [735, 1189] | 1278 p = .099 | 948 [734, 1163] | 1022 |
| **Faces #2** | 2.91 [2.51, 3.31] | 2.25 p = .065 | 977 [775, 1178] | 1345 p = .046 | 975 [778, 1172] | 1141 |

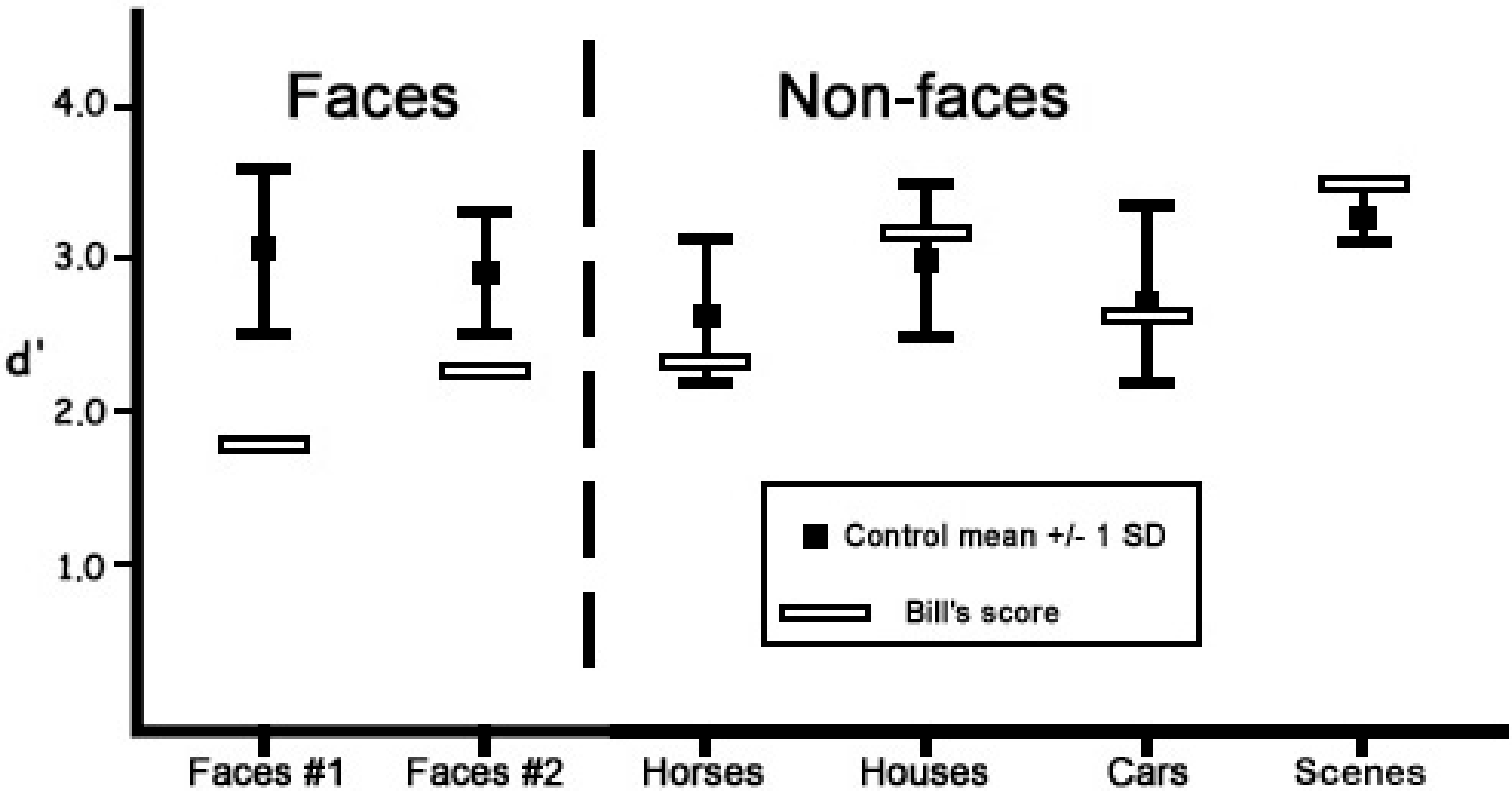
Bill shows consistent impairments on face old/new discrimination tests. Next we will compare his object recognition using this paradigm.

# Object Old/New Discrimination

| | Controls d' [-1 s.d., +1 s.d.] | Bill d' | Controls RT--Hits (msec) [-1 s.d., +1 s.d.] | Bill RT-Hits | Controls RT--CR [-1 s.d., +1 s.d.] | Bill RT--CR |
|---|---|---|---|---|---|---|
| Horses | 2.66 [2.16, 3.16] | 2.36 | 1198 [870, 1525] | 961 | 1151 [922, 1381] | 1088 |
| Cars | 2.78 [2.18, 3.37] | 2.67 | 1196 [790, 1602] | 1328 | 1143 [749, 1537] | 1230 |
| Houses | 2.99 [2.45, 3.54] | 3.14 | 1099 [829, 1368] | 793 | 1086 [874, 1297] | 1086 |
| Scenes | 3.27 [3.08, 3.46] | 3.47 | 1037 [799, 1275] | 801 | 1006 [768, 1240] | 846 |

Bill shows no impairments with object recognition or place recognition. His d’ scores and his RTs are all in the normal range.

# Faces vs. Non-Faces




Bill’s data is inconsistent with the individuation hypothesis.

# CK: Object agnosia without prosopagnosia

- Mr. CK sustained brain damage when he was struck by a car while jogging.
- Born in 1961.
- Has adapted well since his accident.
- Completed MA degree.
- Great difficult with basic level object recognition.
- Perception of objects is very piecemeal.
- Deficit appears to extend to body parts.
- Can copy objects, albeit in a piecemeal fashion.
- Integretative agnosia but can read normally and recognize faces.

# CK: Upright vs. Inverted

CK's face recognition has been tested with a large variety of face types.

•Upright famous face recognition: <u>Normal</u>.
•Inverted famous face recognition: <u>Severely impaired</u>.

This data demonstrates two things:
•CK's difficulty with individuation does not extend to upright faces, and so is inconsistent with the individuation hypothesis.
•Upright & inverted faces are processed by different procedures.

# Individuation hypothesis

- The individuation hypothesis is not capable of accounting for the double dissociation between face & object recognition seen in Bill Choisser & CK.

- Next we will discuss the evidence regarding the other two hypotheses:  the Configural Processing hypothesis & the Expertise hypothesis.

# Configural Processing Hypothesis

CPH proposes that prosopagnosia is a manifestation of damage to procedures specialized for the recognition of objects based on the configuration of their parts.

It predicts that all prosopagnosics should show impairments with tasks tapping configural processing.

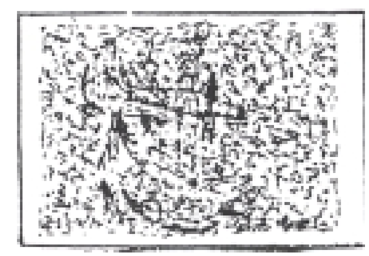
The originators of the CPH claims that tests of visual closure depend on configural processing.

- Require structuring of a percept based on individually meaningless parts.
- Recent review showed no reports of prosopagnosics who have performed normally on these tasks.

# Testing the CPH

CPH predicts that Bill should be unable to perform normally on tests of visual closure.

| | | |
|---|---|---|
| Gestalt Completion | 15.2 (3.6) | 18 |
| Concealed Words | 23.6 (6.4) | 23 |
| Snowy Pictures | 5.7 (3.0) | 13 |



Bill’s has no difficulties with these tasks.  This result is inconsistent with the CPH.  Thus, if these tasks require configural processing, a deficit in CP does not explain Bill’s prosopagnosia.

# Expertise Hypothesis

People are capable of developing expert abilities for classes of stimuli that they are required to recognize.

Researchers have found that many of the putatively face-specific effects can also be found for stimulus classes that people have developed expertise with.

It is plausible that faces are simply one of many classes for which people develop expertise.

If an individual didn’t have the ability to develop expertise, they would not develop expert procedures for faces and so would manifest face recognition deficits.

# Expertise Hypothesis Prediction

•According to the EH, developmental prosopagnosics do not have the procedures necessary to develop expert face recognition.

•Prediction: Prosopagnosics should be unable to develop expertise for non-face stimulus classes.

Gauthier shown that individuals develop expertise after participating in a Greeble training procedure. Thus, prosopagnosics should not be capable of developing expert Greeble recognition.

# Face-specificity conclusions

- The evidence for specialized procedures for faces seems unequivocal.  Much of the strongest evidence comes from dissociations of upright & inverted faces.

- The open question seems to be the developmental process that results in this specialization.

- Because Bill is a developmental prosopagnosic, this indicates that the specialization has been designed by natural selection rather than through domain-general expertise development procedures.

- However, this debate will continue for some time.

# What is a face?

- What types of stimuli activate the face-specific procedures?

- This question can be approached by determining the stimuli that are processed by face-specific procedures.

- This data can come from any hallmarks of face-specific process, but presently the best evidence comes from neuropsychological cases and psychological experiments.

# Lessons from CK

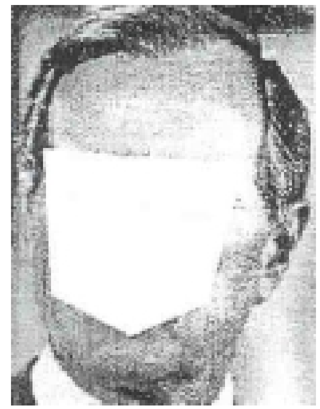
It appears that the only recognition procedures that are unimpaired in CK are face-specific procedures. Thus, if he can recognize a stimulus, it is probably being sent to his face recognition system.

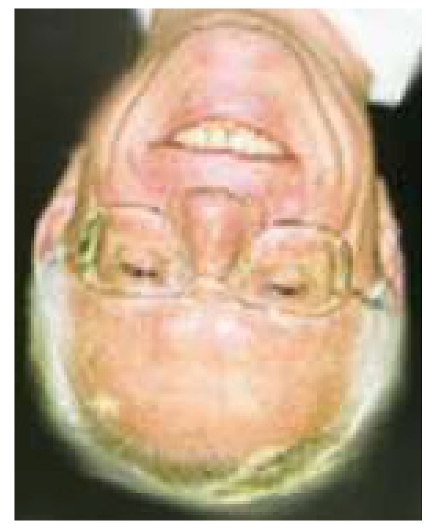
Normal recognition: Upright faces, cartoon faces, half faces cut vertically, caricatures, composite faces made of common objects, faces with just the internal facial features.

Impaired recognition: Inverted faces of any kind, half faces cut horizontally, external facial features (hair, jaw line, ears), fractured faces.

# Faces for CK

# Non-faces for CK





# Other experiments discussed?

- Appears that the sheep faces viewed by WJ didn’t satisfy requirements of his face recognition system.
- Imaging & neurophysiology results indicate stimuli that are not faces.
- However, not all of these results are consistent, and much more work needs to be done.

# Bruce & Young (1986)

A cognitive model of face processing has been proposed based on neuropsychological and psychological evidence.

At this point, the processes at each stage are not well understood, but the divisions of the model are well supported by the data.

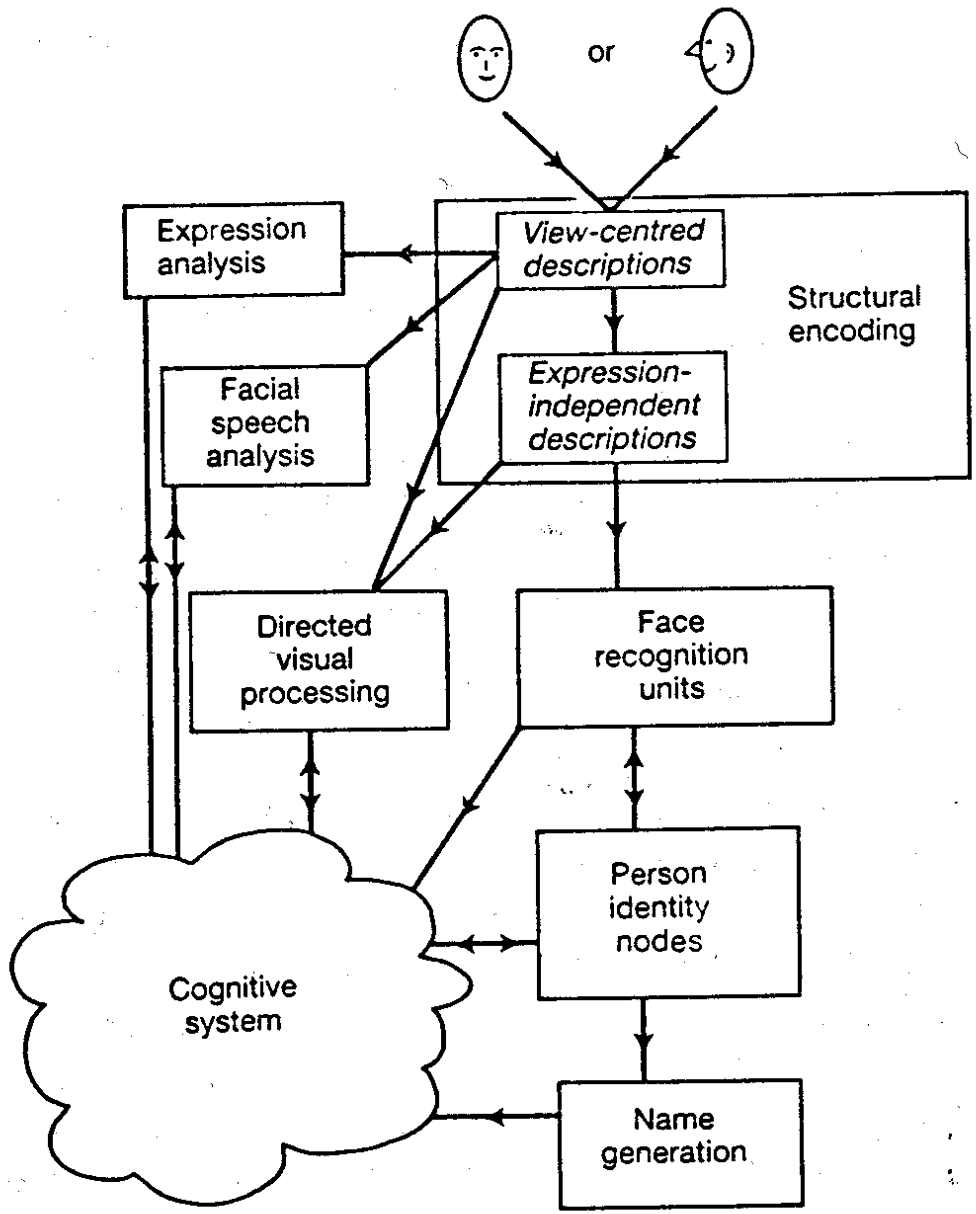



**Fig. 4.1** Bruce and Young's (1986) model of face recognition.

# Face Phenomena

There are a number of interesting phenomena involving faces that haven't been discussed.

- Cross-race recognition deficit.
- Robust representations of faces.
- and one more.

# Cross-Race Recognition Deficit

•Many experiments have found that people have more difficulty recognizing individuals from other races.

•<u>Contact hypothesis</u>: Because people view more same race (SR) faces, they develop expertise for a particular type of face. This expertise is not as effective with cross-race (CR) faces, and so there is a CR deficit.

<u>Prediction</u>: Exposure to faces of other races should lessen the CR deficit, but this is not always the case.

# Feature Selection Hypothesis

Levin, Daniel (2000)

•Feature selection hypothesis claims that the CR deficit results from observers' focus on different types of information in SR and CR faces.

--SR faces are processed for individuating info.
--CR faces are processed for race classifying info.

•These different emphases result in the selection of different facial features depending of the race of the face.

# Feature Selection Hypothesis

Predictions differ for white individuals with the CR deficit and those who do not show it.

CR deficit:

- Better at searching for black faces among white faces, but not vice versa.
- Better at discriminating the race of black faces.

No CR deficit:

- Similar performance on search task.
- Similar at discriminating race regardless of race.

# Visual Search for Race

Stimuli: Sets of 2-8 faces were presented. Target and distractor stimuli were from different races.

Task: Determine whether a face from a particular race is present (For example, is there a white face present?)

Results:
CR deficit—Faster detecting black targets, slower with white targets.

No CR deficit—Very small difference.

# Racial Discrimination

Stimuli: Morphs between an average black and white faces were created.

Task: Decide which of two faces is closer to an end of the continuum.

Results:
CR deficit—better at discriminating at black end!
No deficit—little difference at either end.

# Robust Representations

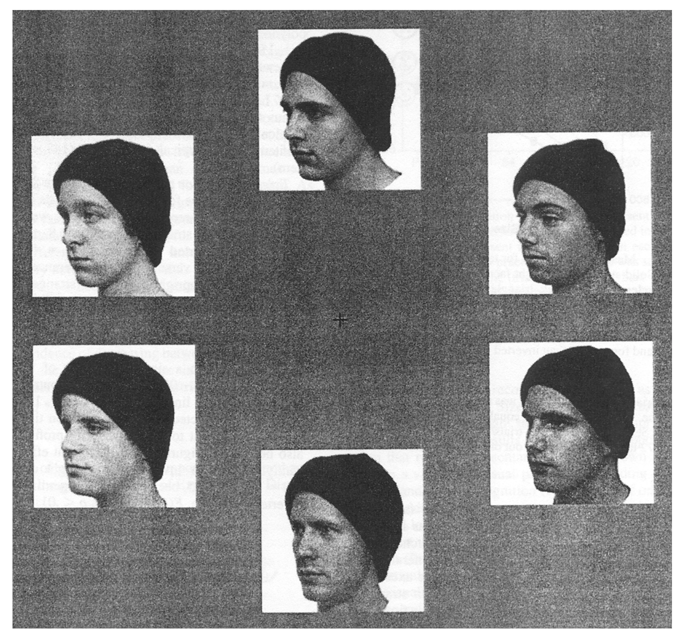
Tong & Nakayama (1999)

• Visual search tasks were used to compare representations of highly overlearned faces.

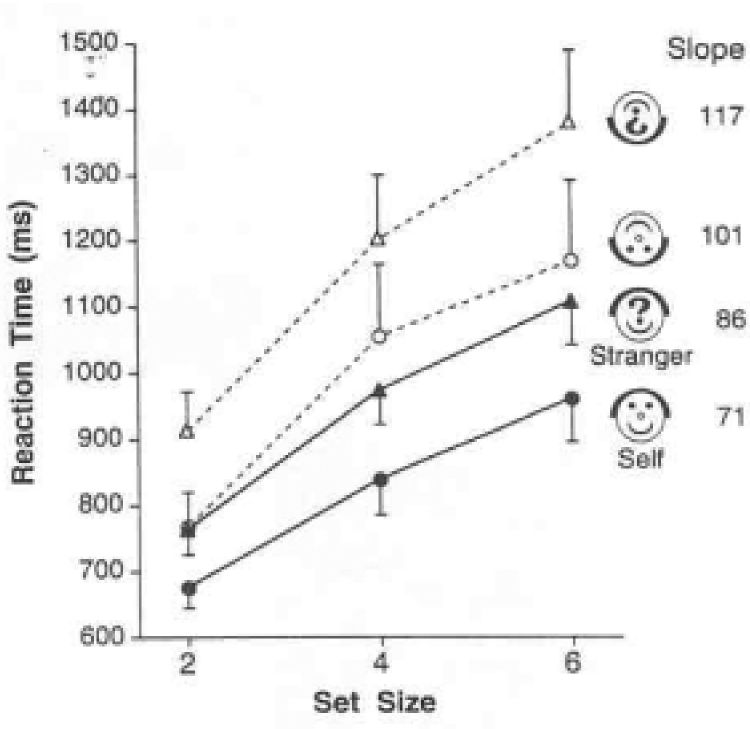
Task: Determine presence or absence of own face or an unfamiliar face.

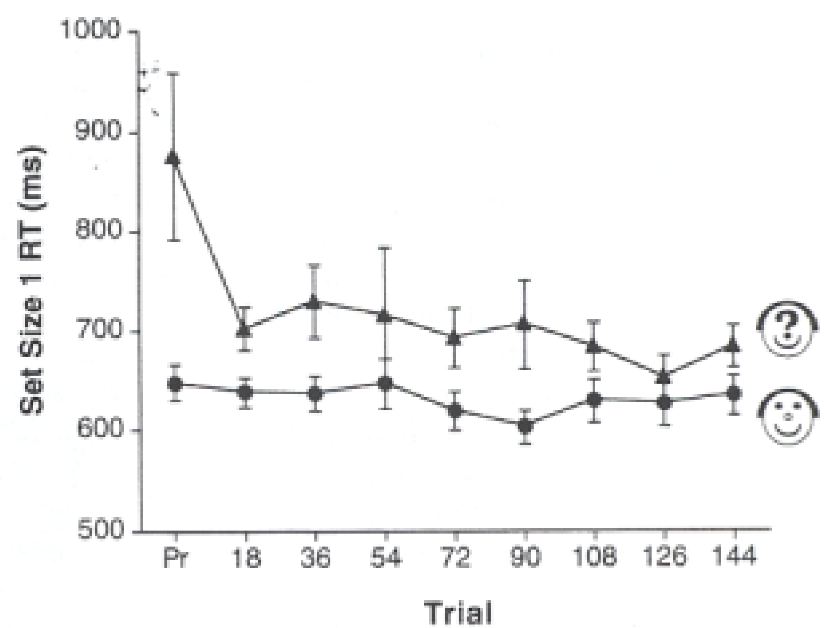
Stimuli: Shots from three different angles:

- front
- 3/4 profile
- full profile



# Results of Face Search




Subjects are faster detecting own face photos than photos of a stranger. Note that even after 144 trials at set size 1, there is still a difference.

There is a long-term component to face representations.


What's Wrong with
This Picture?

# What about hair!?

# Break!

# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# Face Detection Survey

*courtesy of*

Ming-Hsuan Yang
Honda Fundamental Research Labs

# Face Detection

- Goal: Identify and locate human faces in an image (usually gray scale) regardless of their position, scale, in plane rotation, orientation, pose and illumination
- The first step for any automatic face recognition system
- A very difficult problem!
- First aim to detect upright frontal faces with certain ability to detect faces with different pose, scale, and illumination
- One step towards Automatic Target Recognition or generic object recognition


Where are the faces, if any?

# Why Face Detection is Difficult?

- **Pose**: Variation due to the relative camera-face pose (frontal, 45 degree, profile, upside down), and some facial features such as an eye or the nose may become partially or wholly occluded.
- **Presence or absence of structural components**: Facial features such as beards, mustaches, and glasses may or may not be present, and there is a great deal of variability amongst these components including shape, color, and size.
- **Facial expression**: The appearance of faces are directly affected by a person's facial expression.
- **Occlusion**: Faces may be partially occluded by other objects. In an image with a group of people, some faces may partially occlude other faces.
- **Image orientation**: Face images directly vary for different rotations about the camera's optical axis.
- **Imaging conditions**: When the image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.

# Related Problems

- **Face localization**: aim to determine the image position of a single face; this is a simplified detection problem with the assumption that an input image contains only one face

- **Facial feature detection**: to detect the presence and location of features such as eyes, nose, nostrils, eyebrow, mouth, lips, ears, etc. with the assumption that there is only one face in an image

# Methods to Detect/Locate Faces

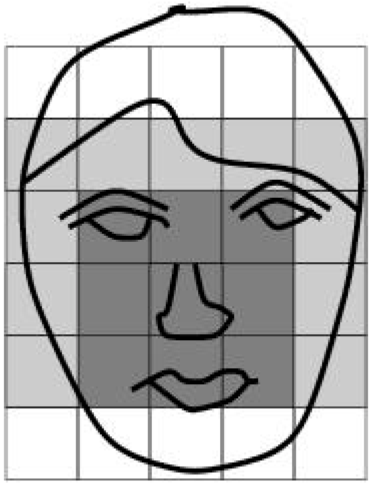
- Knowledge-based methods:
  - encode human knowledge of what constitutes a typical face (usually, the relationships between facial features).
- Feature invariant approaches:
  - aim to find structural features that exist even when the pose, viewpoint, or lighting conditions vary
- Template matching methods:
  - Several standard patterns stored to describe the face as a whole or the facial features separately.
  - The correlations between an input image and the stored patterns are computed for detection.
- Appearance-based methods:
  - the models (or templates) are learned from a set of training images which capture the representative variability of facial appearance.

# Knowledge-Based Methods

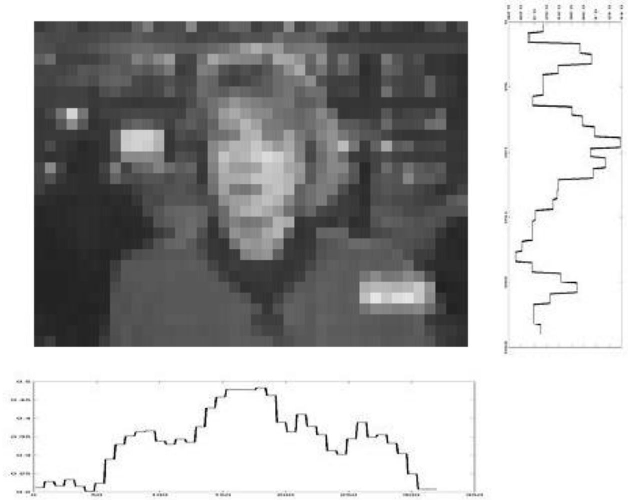
- Pros:
  - Easy to come up with simple rules to describe the features of a face and their relationships.
  - E.g.: a face often appears in an image with two eyes that are symmetric to each other, a nose and a mouth (in terms of relative locations and distance) .
  - Facial features in an input image are extracted first, and face candidates are identified based on the coded rules.
  - A verification process is usually applied to reduce false detections.
  - Work well for face localization (single face) in uncluttered background
- Cons:
  - Difficult to translate human knowledge into rules precisely: Detailed rules fail to detect faces and general rules may find many false positives.
  - Difficult to extend this approach to detect faces in different poses since it is challenging to enumerate all the possible cases.

# Knowledge-Based Method: Example

- [Yang and Huang 94]:
  3-level rules based on the intensity distribution and difference of multi-resolution images and edges detection



- [Kotropoulos and Pitas 94]:
  horizontal/vertical projection to locate features and then apply rules for localization

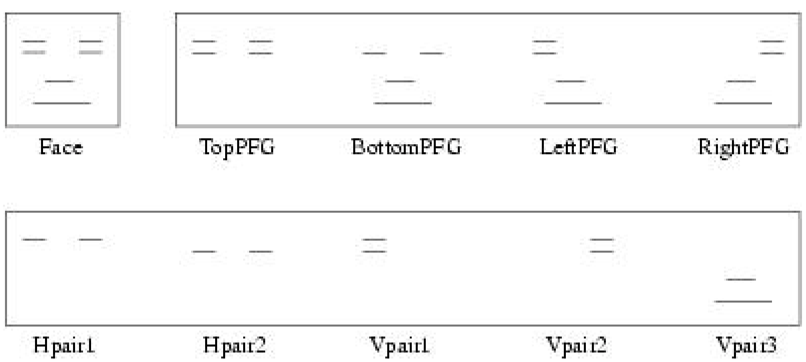


# Feature Invariant Approach

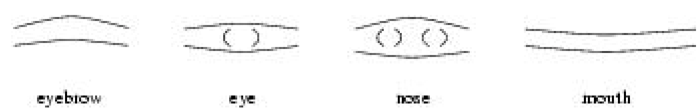
- Detect facial features (eyes, nose, mouth, etc)
- Group features into candidates and verify them
- Facial features: edge, intensity, shape, texture, color
- Pros:
  - Can locate faces in different pose and orientation
- Cons:
  - Difficult to locate facial features due to several corruption (illumination, noise, occlusion)
  - Difficult to detect features in complex background

# Feature Invariant Approach: Examples

- [Leung and Perona 95]: probabilistic method based on feature detectors and random graph matching
- [Yow and Cipolla 90]: Bayesian network of face models and components




Face model and component




Model facial feature as pair of edges

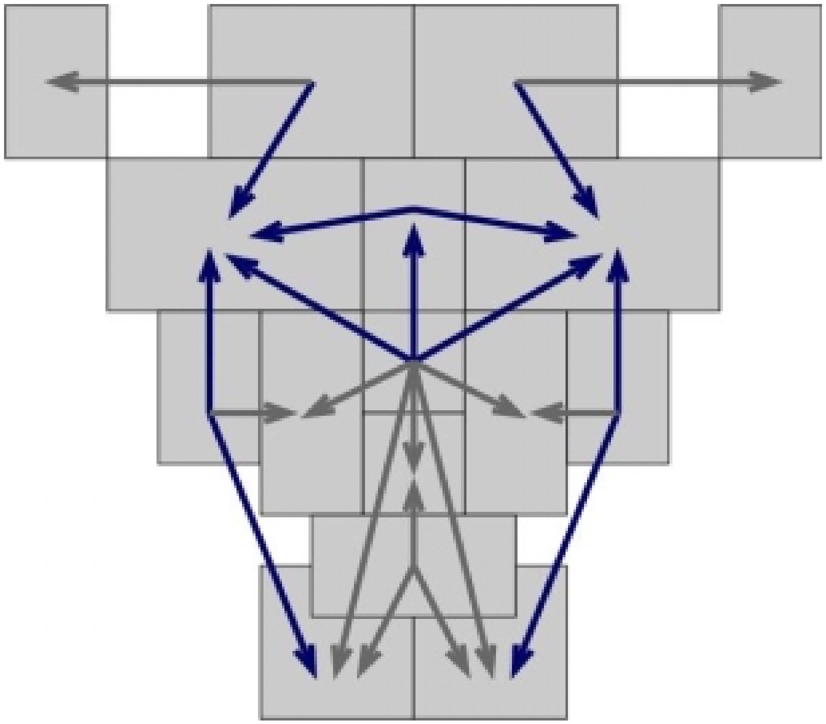

Apply interest point operator and edge detector to search for features

# Template Matching Methods

- Store a template
  - Predefined: based on edges or regions
  - Deformable: based on facial contours (e.g. snakes)
- Find correlation to locate faces
- Pros:
  - Simple
- Cons:
  - Expensive computation (due to correlation)
  - Difficult to enumerate good templates (similar to knowledge-based methods)

# Template Matching Method: Example

- [Sinha 94]:
  - Use relative pair-wise ratios of the brightness of facial regions (is one region darker than the other?)
  - Been applied to Kismet project

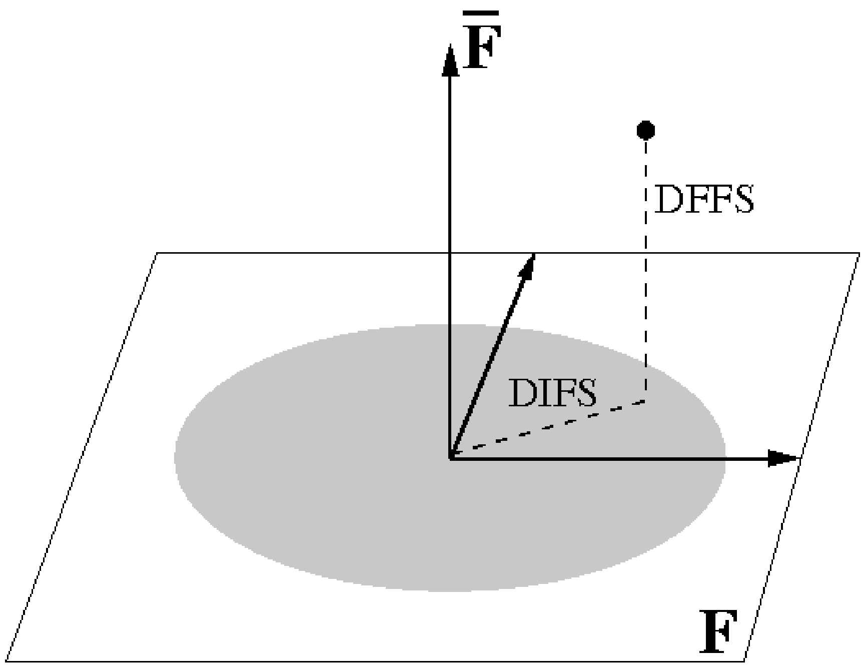


# Appearance-Based Methods

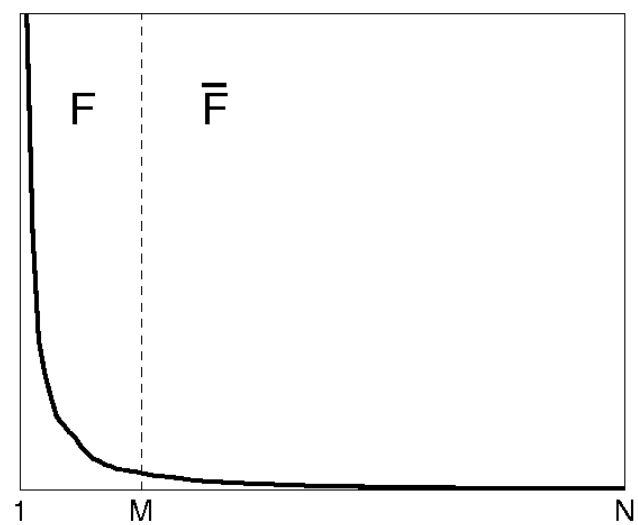
- Use positive (and possibly negative examples) of faces to train the classifier, or to estimate a probabilistic distribution using statistics or machine learning methods
  - Neural network: Multilayer Perceptrons
  - Eigenface or PCA, Factor Analysis
  - Support Vector Machine: Polynominal kernels
  - Mixture of PCA, mixture of factor analyzers
  - Distribution-based method
  - Naïve Bayes classifier
  - Hidden Markov Model
  - Sparse Network of Winnows
  - Information-Theoretical Approach: Kullback Relative Information
  - Inductive Learning: C4.5

# Subspace Face Detector

- PCA-based Density Estimation  p(x)
- Maximum-likelihood face detection based on DIFS + DFFS



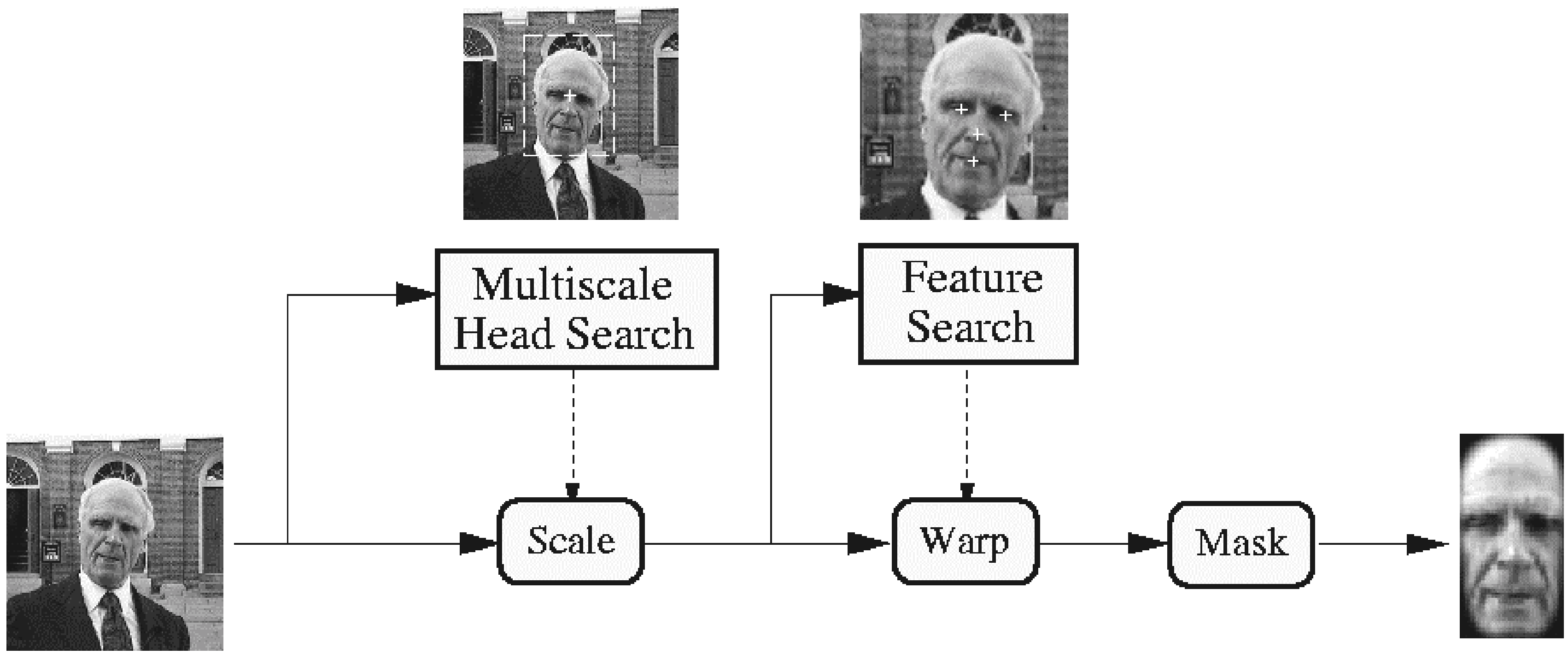

Eigenvalue spectrum




Moghaddam & Pentland, “Probabilistic Visual Learning for Object Detection,”  ICCV’95.

# Subspace Face Detector

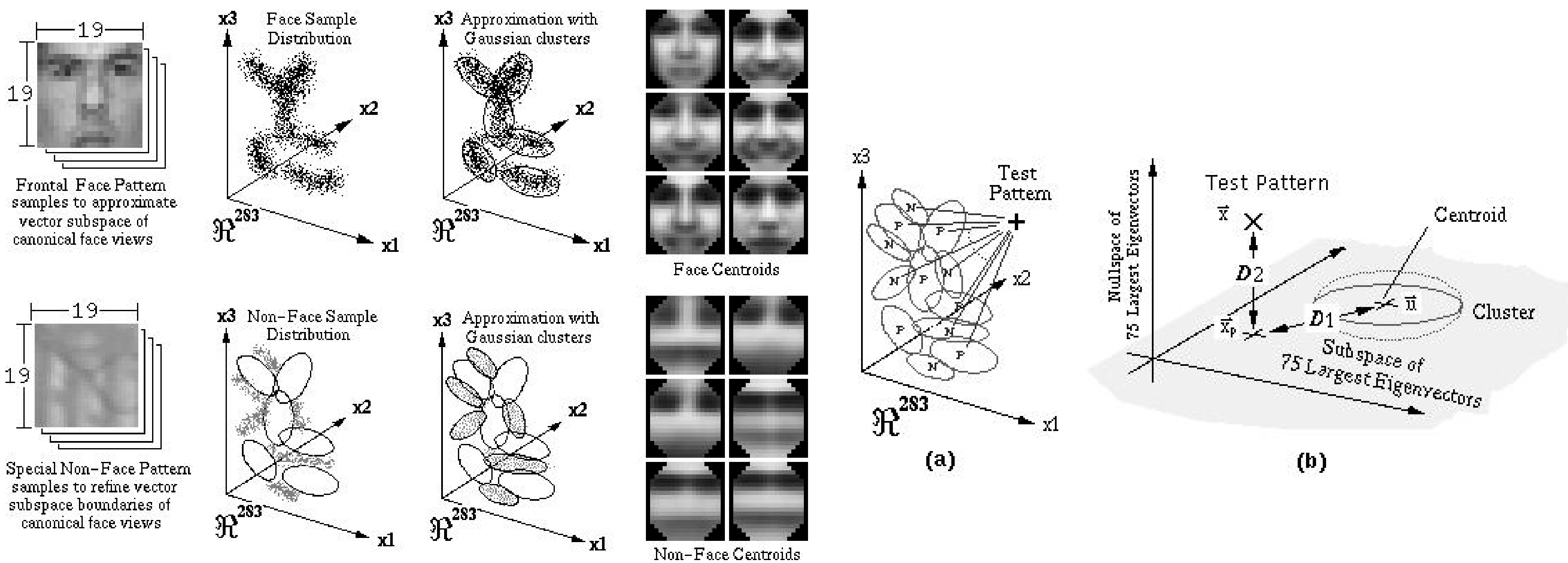
- Multiscale Face and Facial Feature Detection & Rectification




Moghaddam & Pentland, “Probabilistic Visual Learning for Object Detection,” ICCV’95.

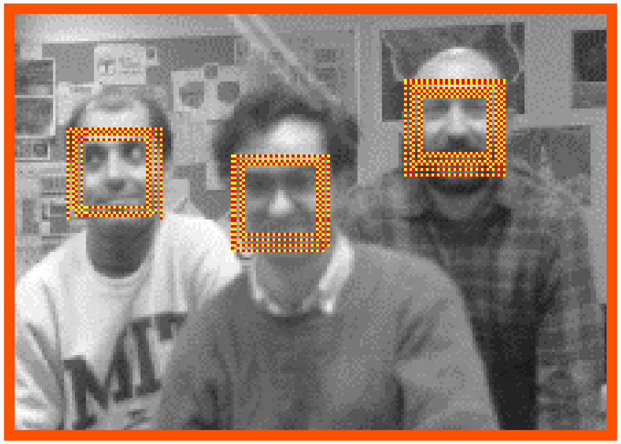
# Distribution-Based Face Detector

- Learn face and nonface models from examples [Sung and Poggio 95]
- Cluster and project the examples to a lower dimensional space using Gaussian distributions and PCA
- Detect faces using distance metric to face and nonface clusters



# Distribution-Based Face Detector

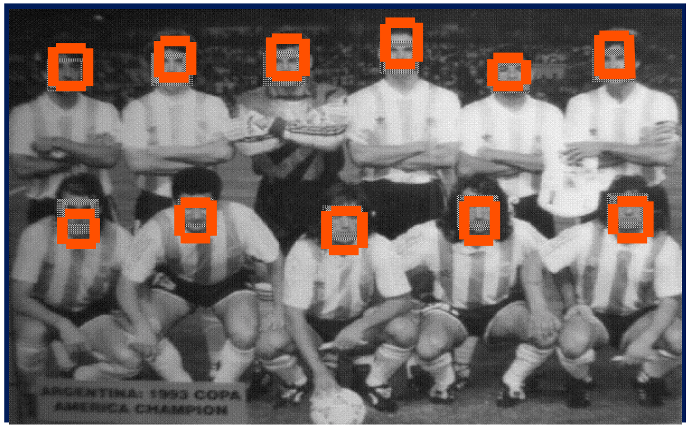
- Learn face and nonface models from examples [Sung and Poggio 95]




Training Database

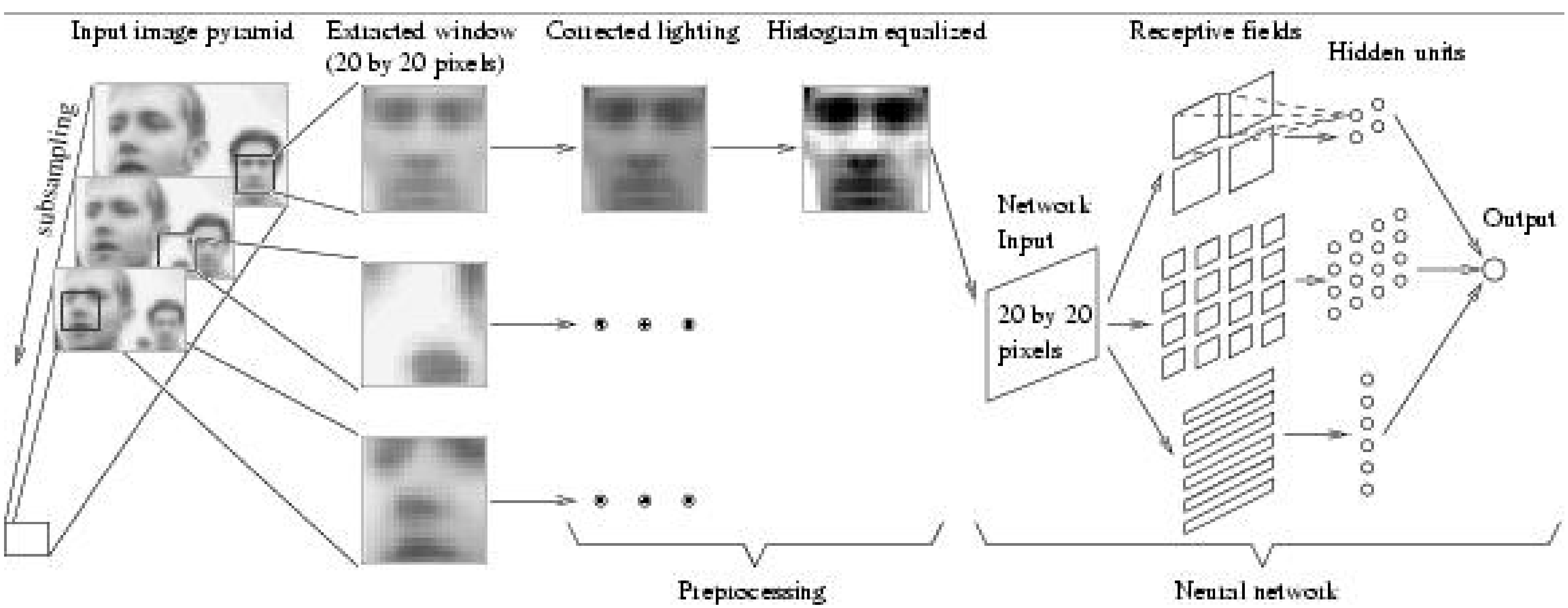
1000+ Real, 3000+ *VIRTUAL*

50,0000+ Non-Face Pattern



# Neural Network-Based Face Detector

- Train a set of multilayer perceptrons and arbitrate a decision among all outputs [Rowley et al. 98]



# Standard Test Sets

- MIT Test Set (http://www.cs.cmu.edu/~har): subsumed by CMU Test Set.
- CMU Test Set (http://www.cs.cmu.edu/~har) (de facto benchmark): 130 gray scale images with a total of 507 frontal faces.
- CMU Profile Face Test Set (http://eyes.ius.cs.cmu.edu/usr20/ftp/testing_face_images.tar.gz) : 208 gray scale images with faces in profile views
- Kodak Data Set (Eastman Kodak Corp): Faces of multiple size, pose and varying lighting conditions in color images.

# Research Issues

- Detect faces under varying pose, orientation, occlusion, expression, and lighting conditions
- Performance evaluation
- Standardized Testing
- Fast and real time face detectors
  - Viola and Jones CVPR'01

# References

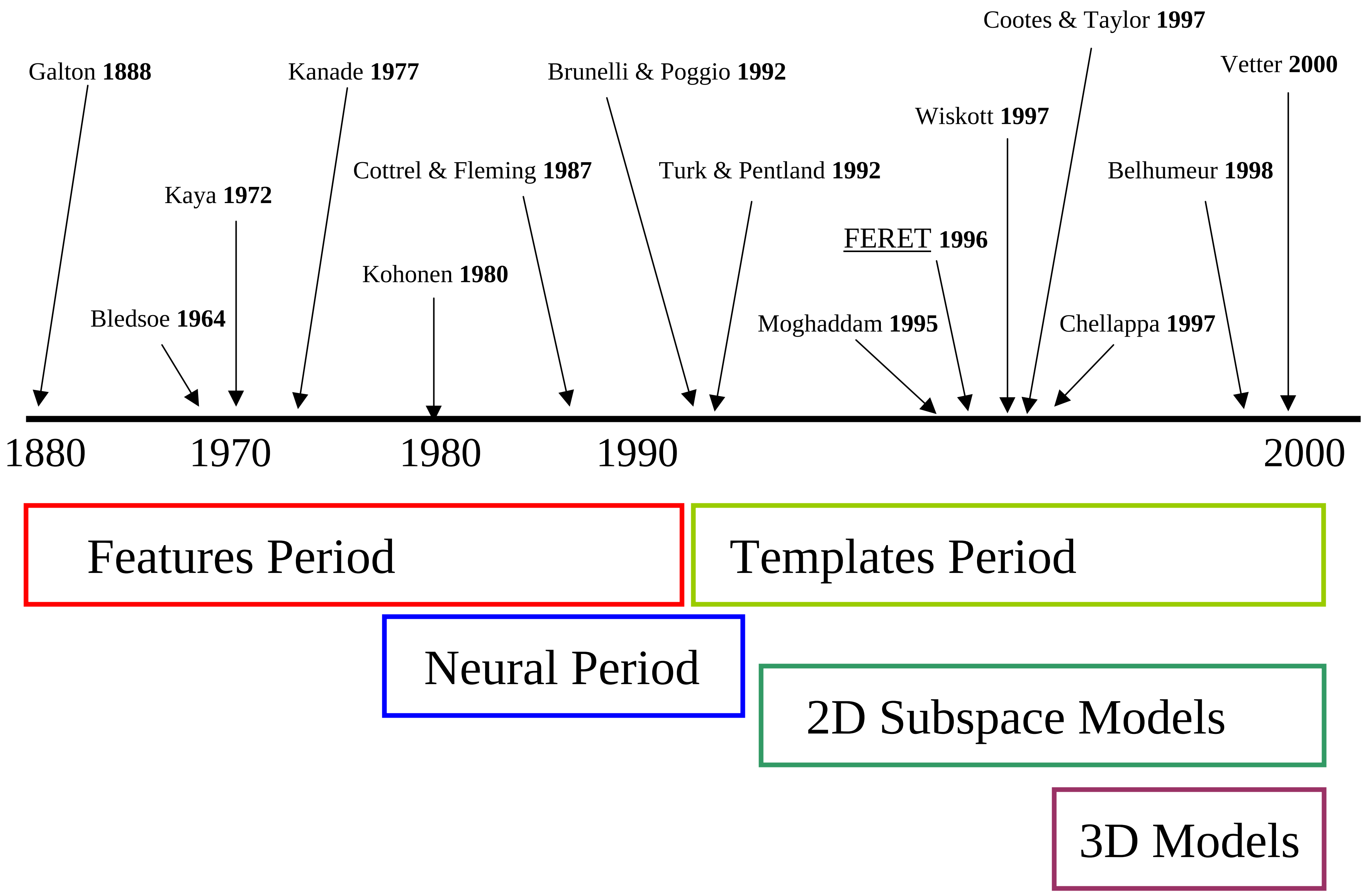

- M.-H. Yang, D. J. Kriegman, and N. Ahuja, “Detecting Faces in Images: A Survey”, to appear in IEEE PAMI.

- M. Pantic and L. J. M. Rothkrantz, “Automatic Analysis of Facial Expressions: The State of the Art”, IEEE PAMI 22 (12), pp. 1424-1445.

# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# A Brief History (1900-2000)

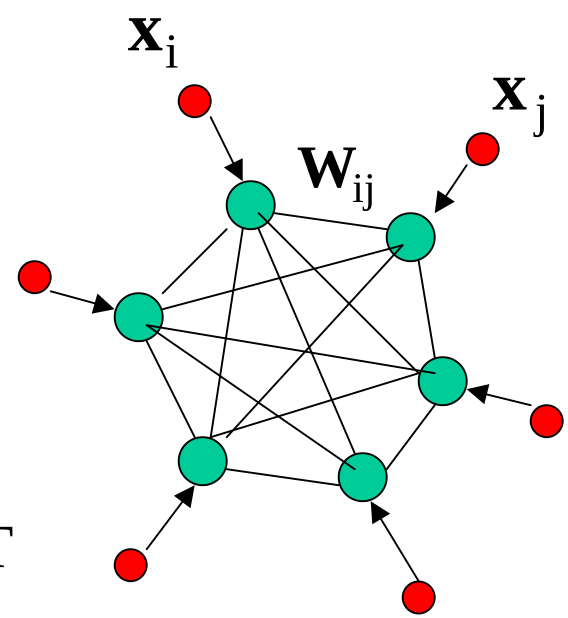


# Auto-associative Memory (CAM)

Anderson et al 1977   Kohonen 1979   Kohonen & Oja 1981

$$\hat{\mathbf{X}} = \mathbf{W}\,\mathbf{X}$$

$$\mathbf{X} \longrightarrow \hat{\mathbf{X}}$$




Widrow-Hoff (Hebbian) Learning:

$$\mathbf{W} \leftarrow \mathbf{W} + \eta\,(\mathbf{X} - \mathbf{W}\,\mathbf{X})\,\mathbf{X}^T$$

$$\mathbf{W} \leftarrow \mathbf{X}\mathbf{X}^T = \mathbf{\Sigma}$$

---

$$\mathbf{W} = \mathbf{U}\mathbf{L}\mathbf{U}^T \quad \text{with} \quad \mathbf{L} \rightarrow \mathbf{I}$$

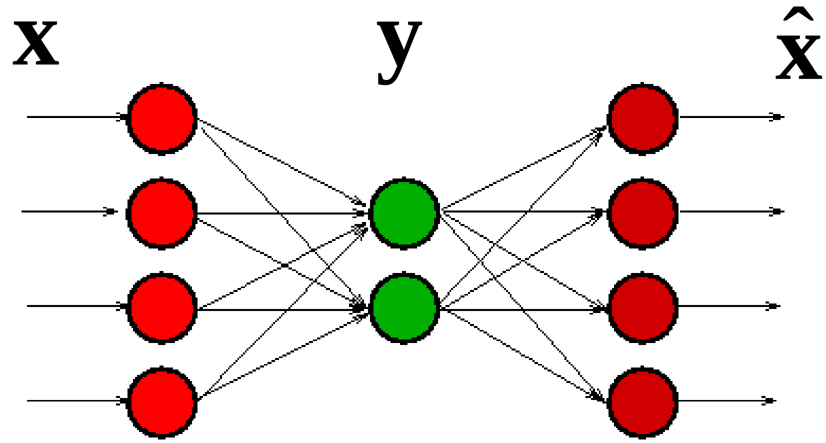
# Auto-encoder Network

Cottrell et al 1987   Cottrell & Fleming 1990   Golomb et al 1991

$$\mathbf{y} = f(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{x}} = f(\mathbf{W}_2 \mathbf{y})$$

$$f(x) = \frac{1}{1-e^{-\alpha x}}$$




MLP: "Backpropagation" Learning (Werbos 1974)

---

$\mathbf{W}_1$ "receptive field"

$\mathbf{W}_2$ "projective field"

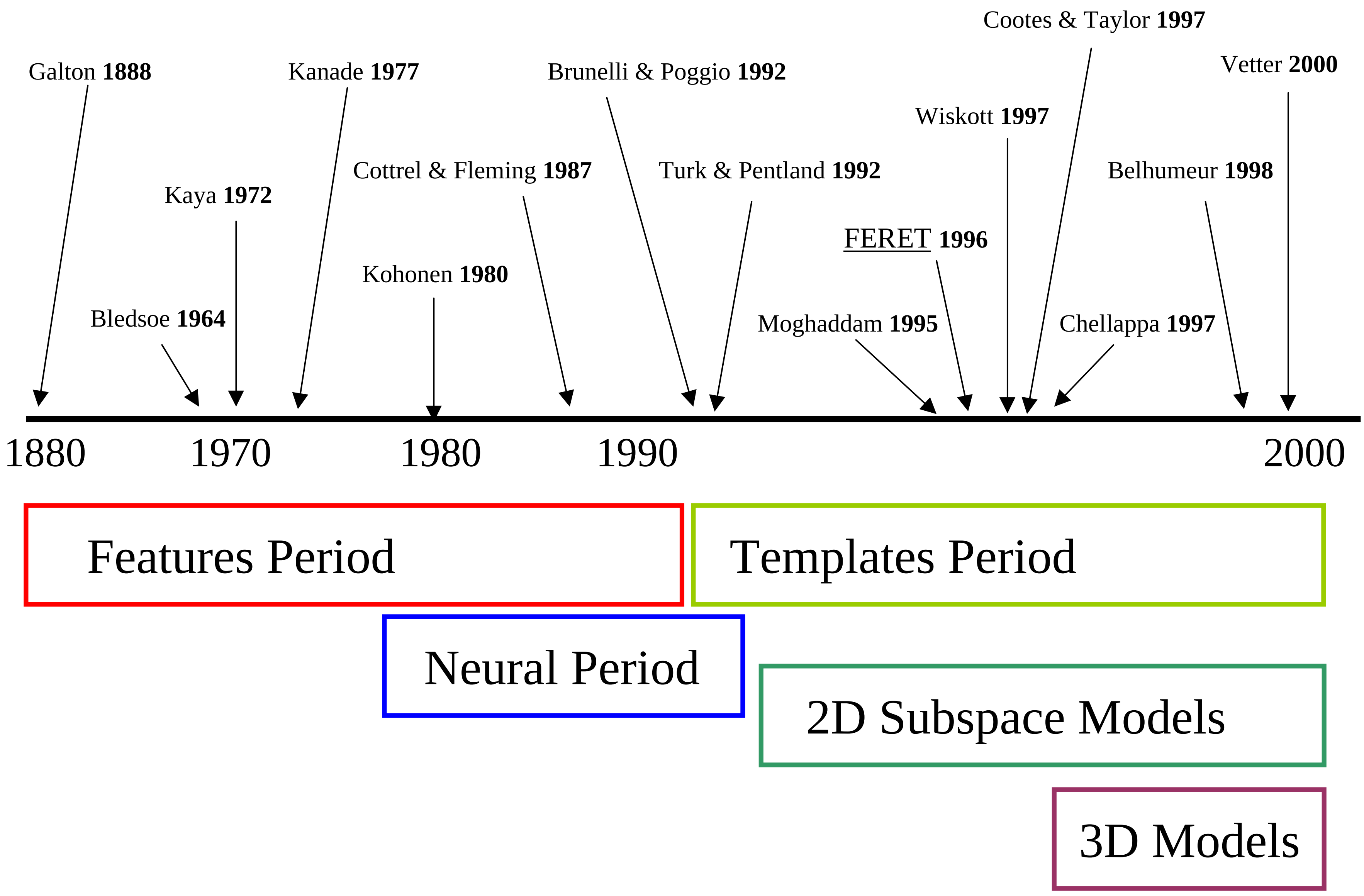
$\mathbf{y} \approx$ principal components

→ identification & classification

# Course Outline

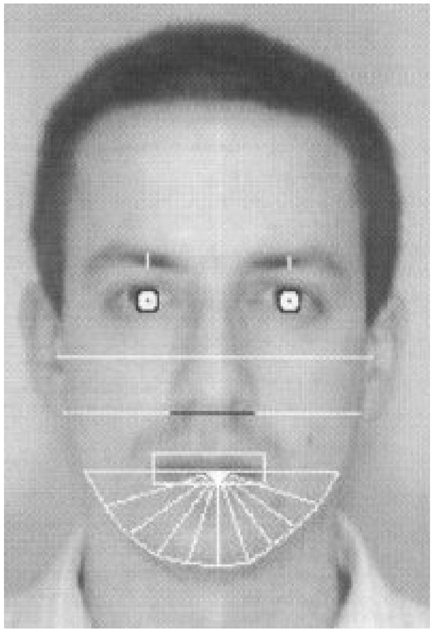
- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# A Brief History (1900-2000)



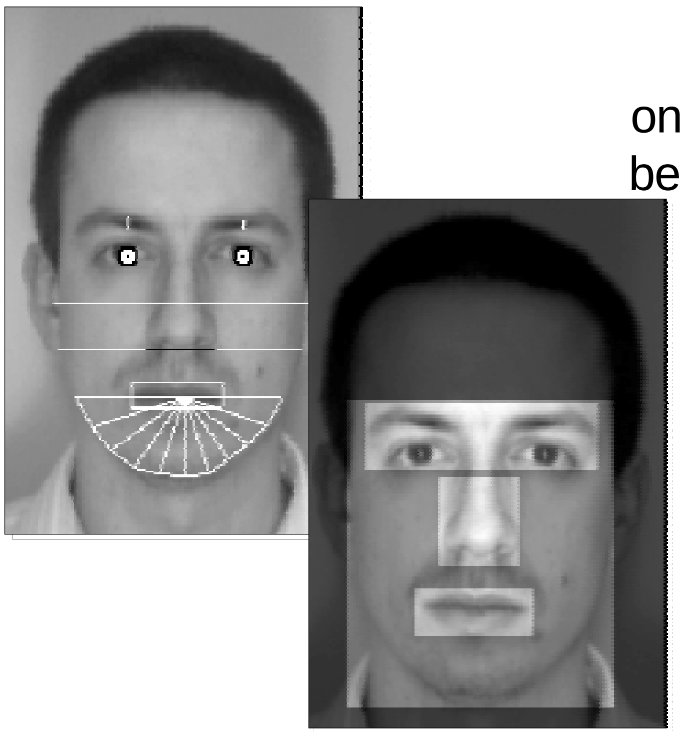
# Brunelli & Poggio 1993 PAMI-15(10)

**Features** *vs.* **Templates**



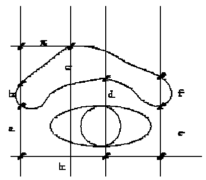
Psychological studies of human **face recognition** suggest that virtually every type of information is used.

Two ways can be distinguished to get a one-to-one correspondence between the stimulus (face to be recognized) and the stored representation (face in the database):

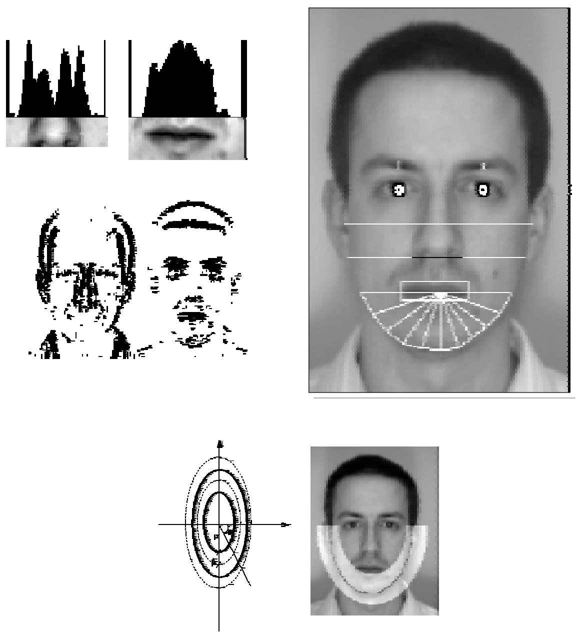


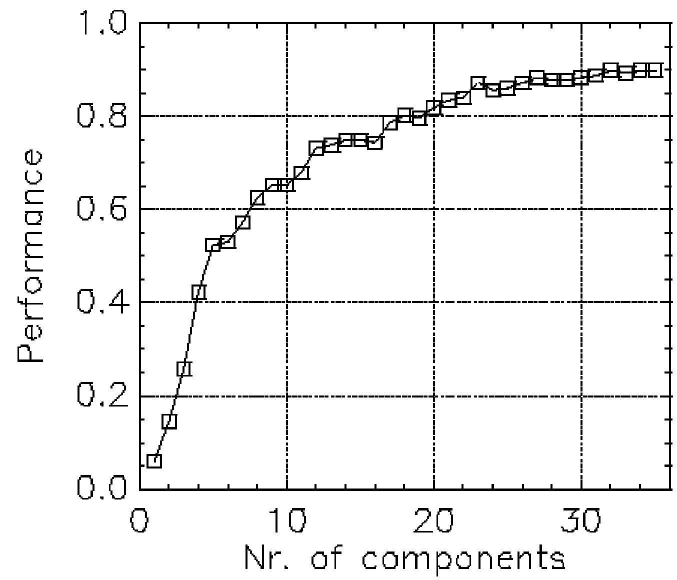
- **geometric**, feature-based matching relying on a restricted set of numbers describing position and shape of face features
- **iconic**, template-matching using bidimensional array of values providing a dense representation of faces

R. Brunelli & T.Poggio, *Face Recognition: Features vs. Templates*



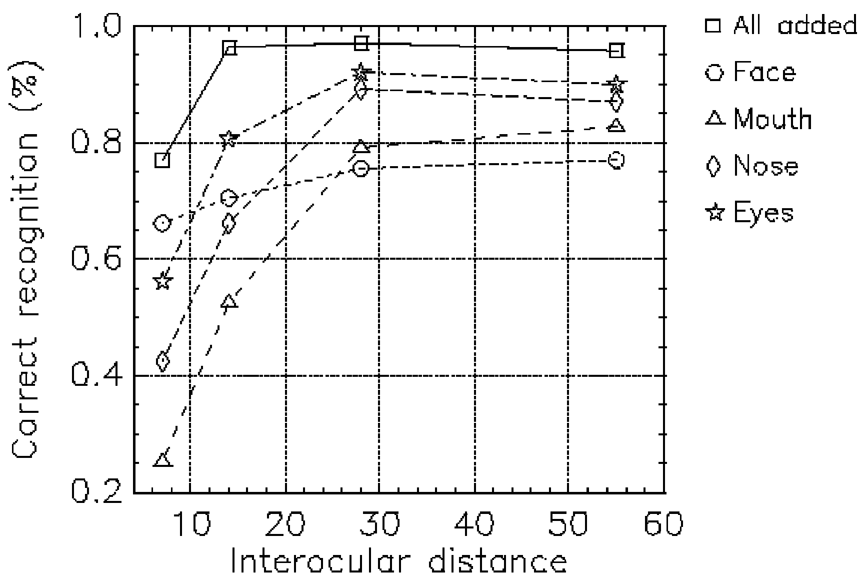
## ... geometric, feature-based matching






After scale and rotation **normalization** of the eye-to-eye segment **35 numbers** are computed through **edge projection** analysis: **principal component** analysis shows that all of them are necessary to get optimal performance with a Bayes classifier under the assumption of **Gaussian distribution** for the feature vectors of a single person.

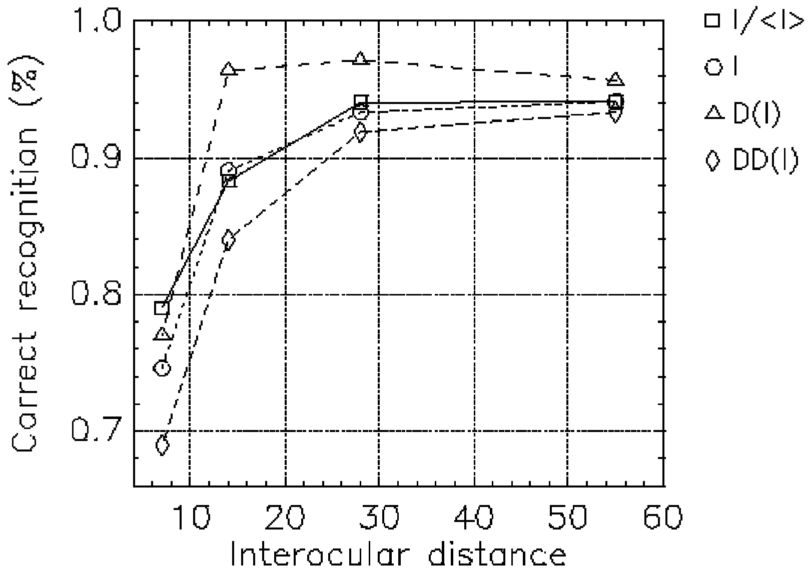
## … template matching




Template matching is based on the average normalized correlation of several face patches.

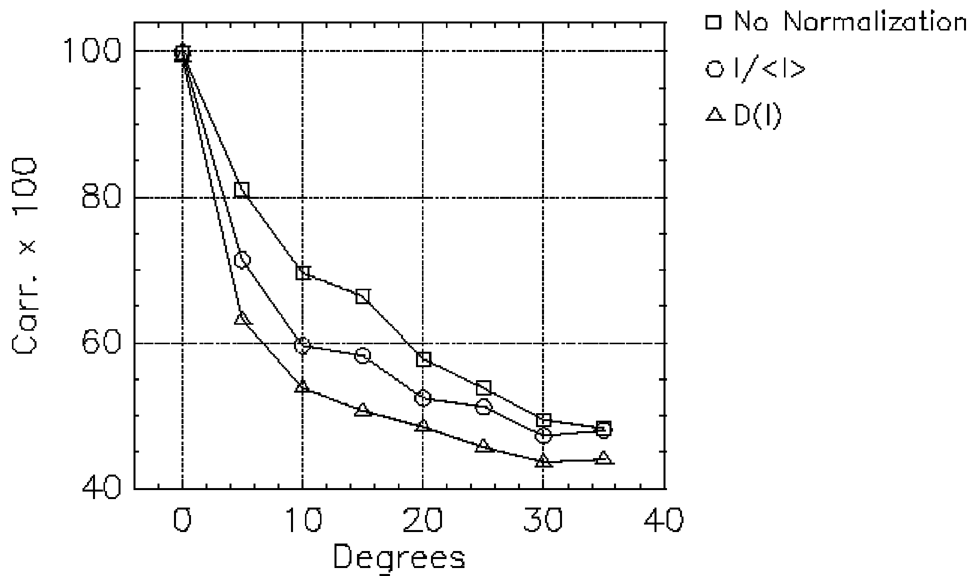
Using multiple features (and examples) provides increased robustness and performance even at very low resolution.

## … template matching




Several preprocessing techniques have been compared to optimize correlation performance against illumination variations and best results were obtained with gradient intensity

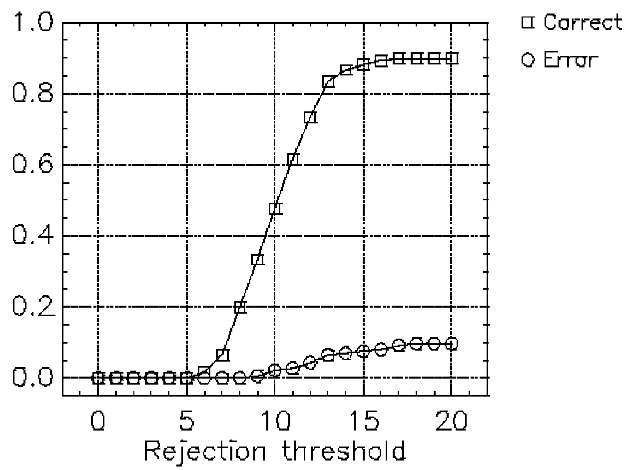
## … template matching



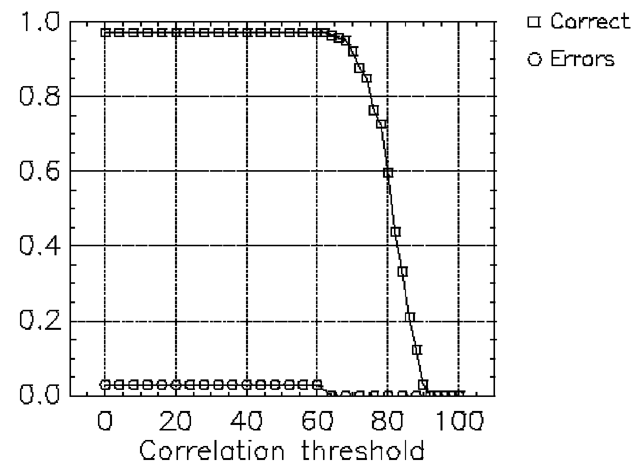

Correlation matching performance decreases markedly with rotation if only frontal templates are available.

Scale variations also adversely affect performance if not managed properly.

# … and the winner is: **TEMPLATE MATCHING**



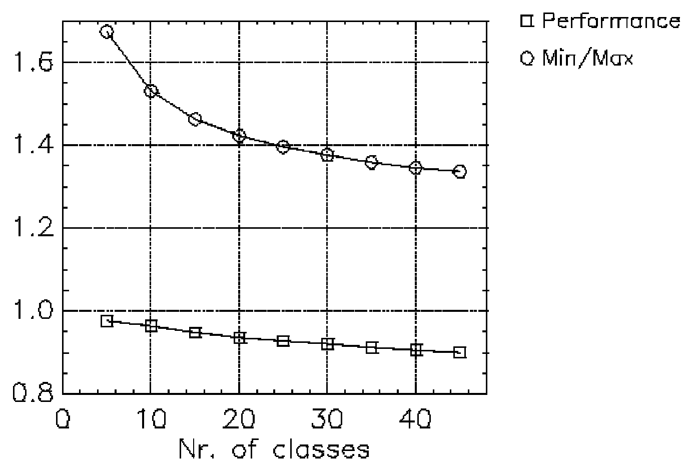

Geometrical features



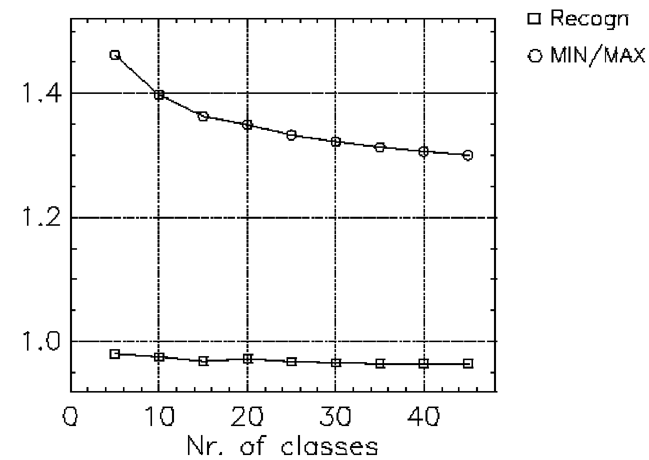

Template Matching

The approach based on **template matching** provides the best performance with or without rejection.

# … and the winner is: **TEMPLATE MATCHING**




Geometrical features

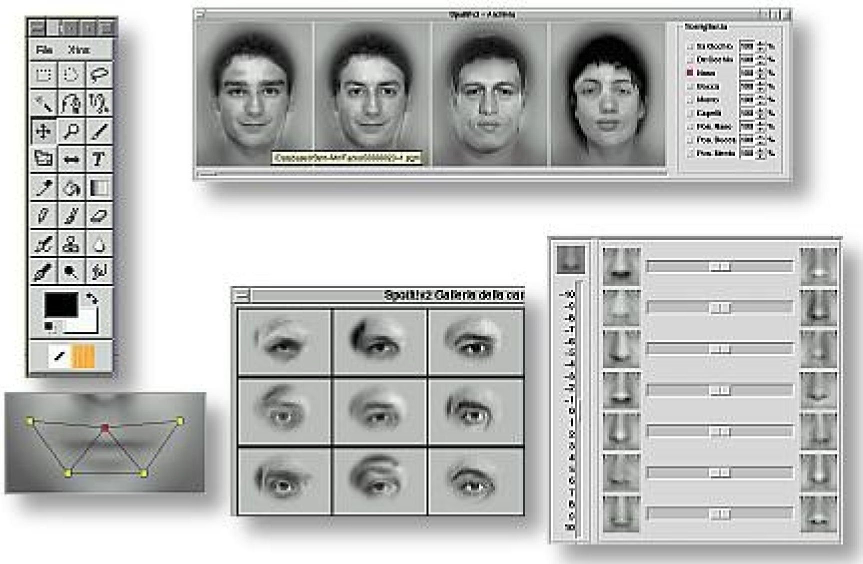



Template Matching

The **template matching** approach also provides **better scalability** as recognition experiments with subsets of different cardinality show (the average MIN/MAX ratio is a rough estimate of how easy it is discriminating different people).

# SpotIt!

http://spotit.itc.it



The feature based approach has been applied to the problem of joint **image synthesis and recognition** based on principal components:

the result is a novel face compositing system where the user can build images interactively moving sliders related to principal projections while getting immediate feedback from the system showing the most similar images in the database.

# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# Subspace Methods

- PCA (“Eigenfaces”)
- “Dual PCA”
  - Bayesian matching
- LDA/FLD (“Fisherfaces”)
- ICA
- LFA

# Principal Component Analysis

Joliffe (1986)

- data modeling & visualization tool
- discrete (partial) Karhunen-Loeve expansion
- dimensionality reduction tool $R^N \to R^M$
- makes <u>no</u> assumption about *p(x)*
- if *p(x)* is Gaussian, then $p(x) = \prod_i N(y_i; 0, \lambda_i)$

# Eigenfaces (PCA)

## Kirby & Sirovich (1990), Turk & Pentland (1991)

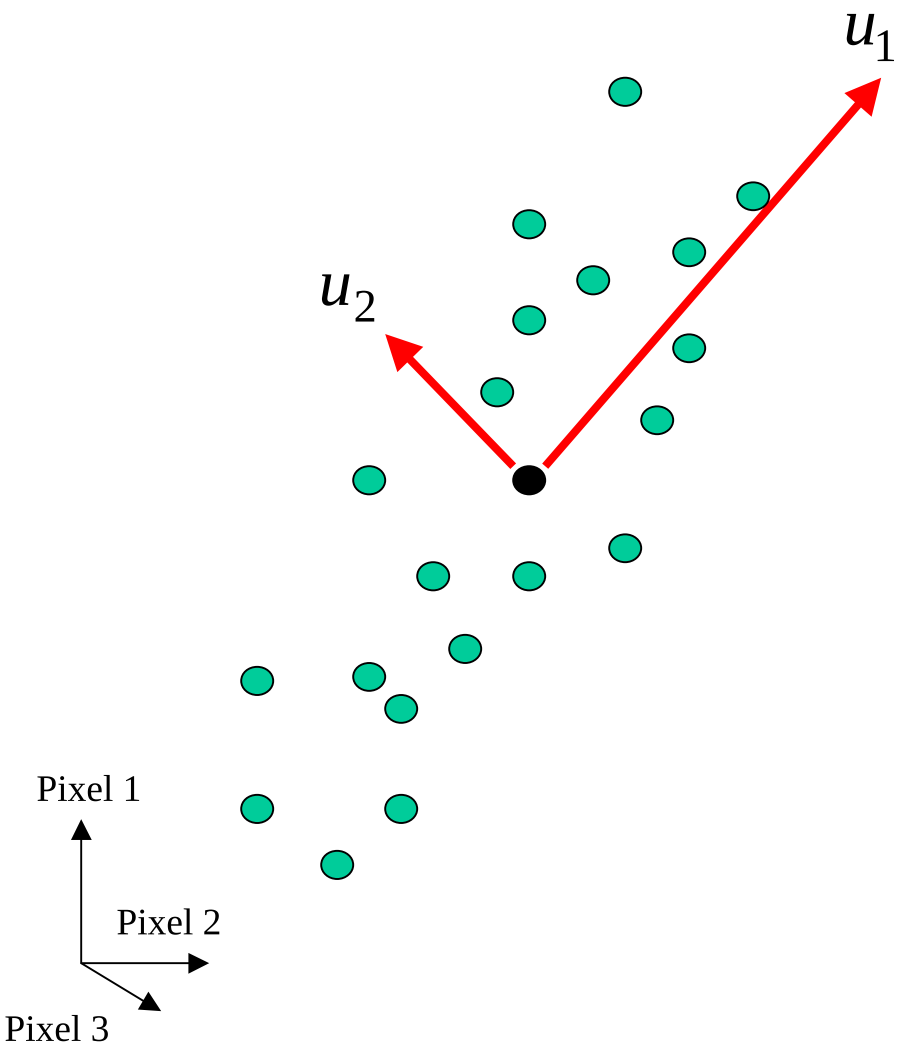
$$\{x_i\}_{i=1}^{M} \quad x \in R^N \qquad M < N$$

$$\mu = \frac{1}{M}\sum_{i=1}^{M} x_i$$

$$S = \sum_{i=1}^{M} (x_i - \mu)(x_i - \mu)^T$$

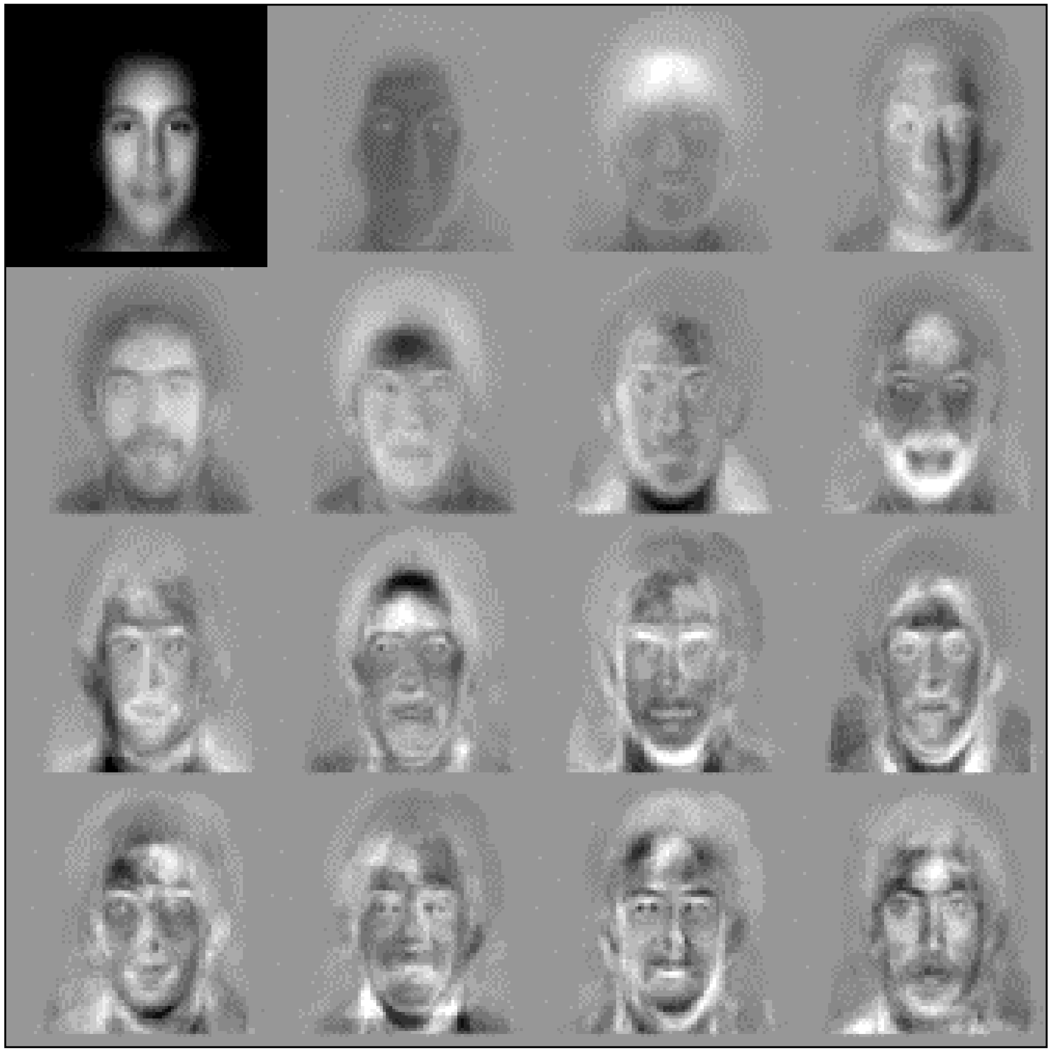
$$S = ULU^T$$

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



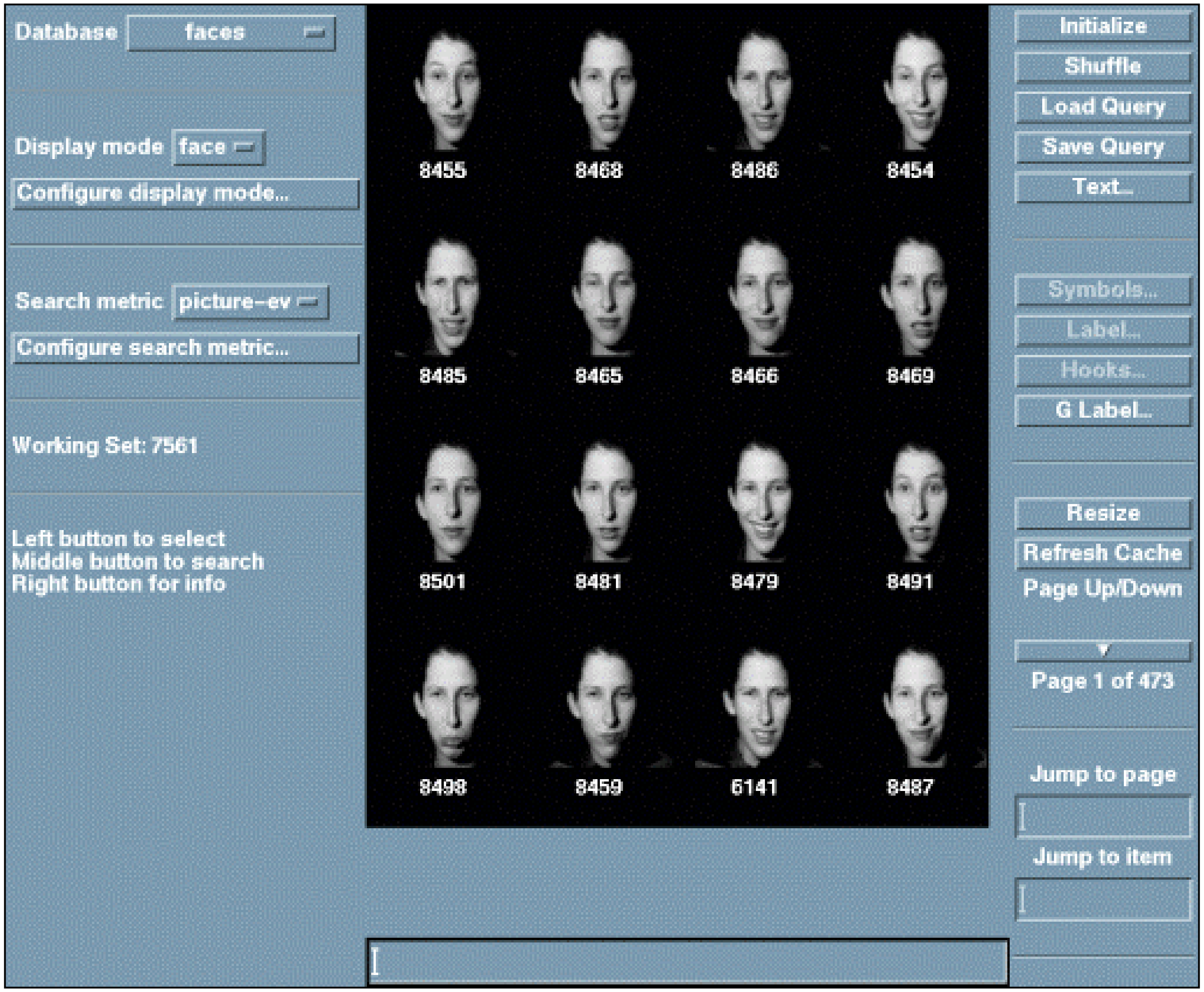
# Eigenfaces

## Turk & Pentland (1992)



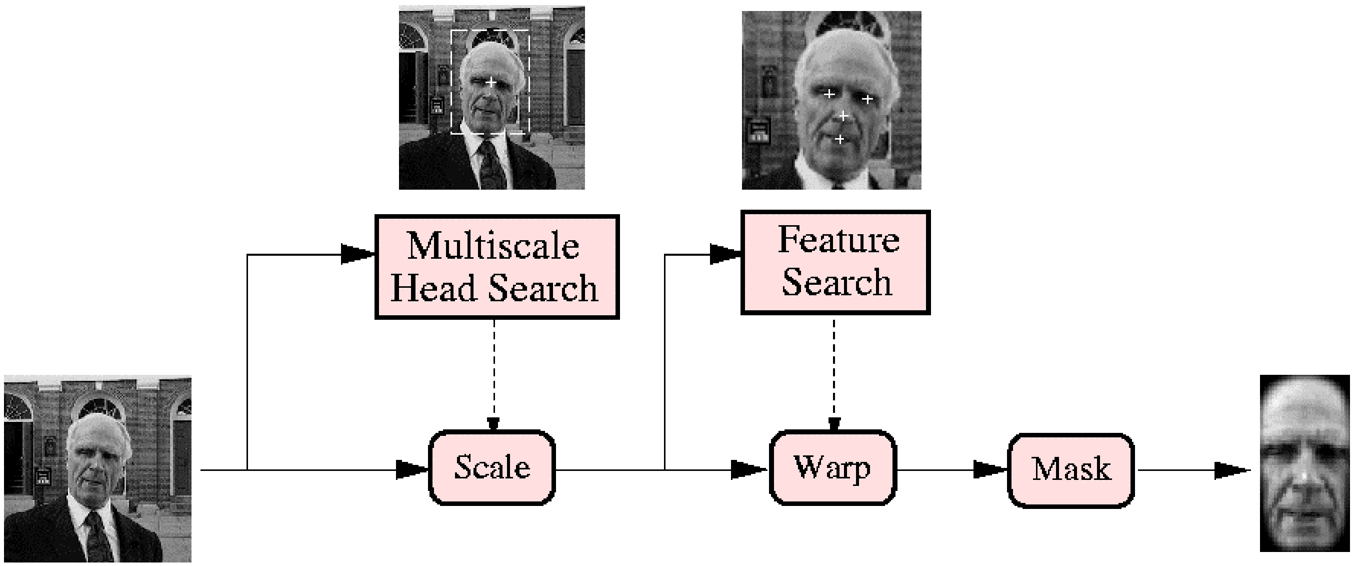
# Eigenfaces

## *Photobook* (MIT)

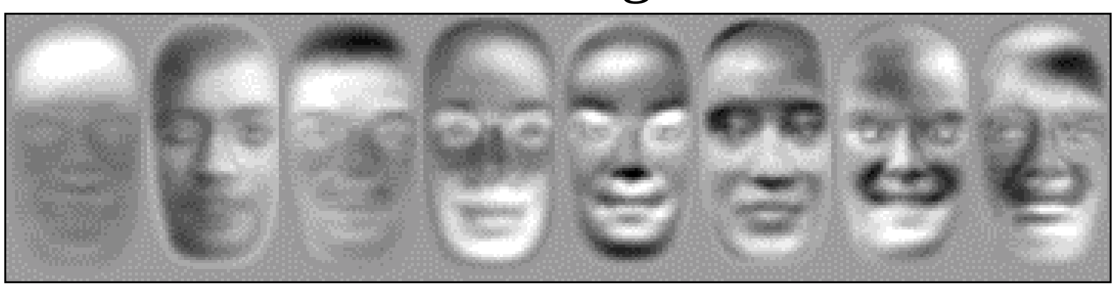


# Eigenfaces

Moghaddam & Pentland (1995)




Normalized Eigenfaces



# FERET Photobook

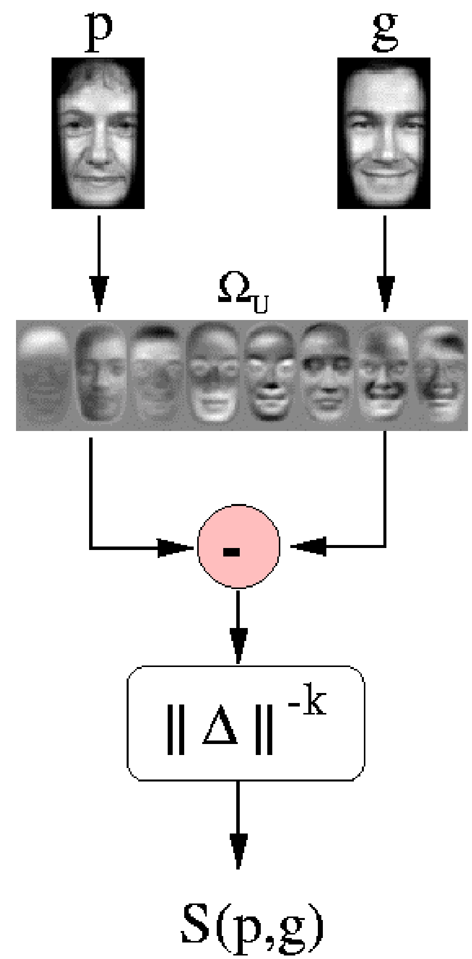
Moghaddam & Pentland (1995)

# FERET “Eigenfaces”

Moghaddam & Pentland (1995)

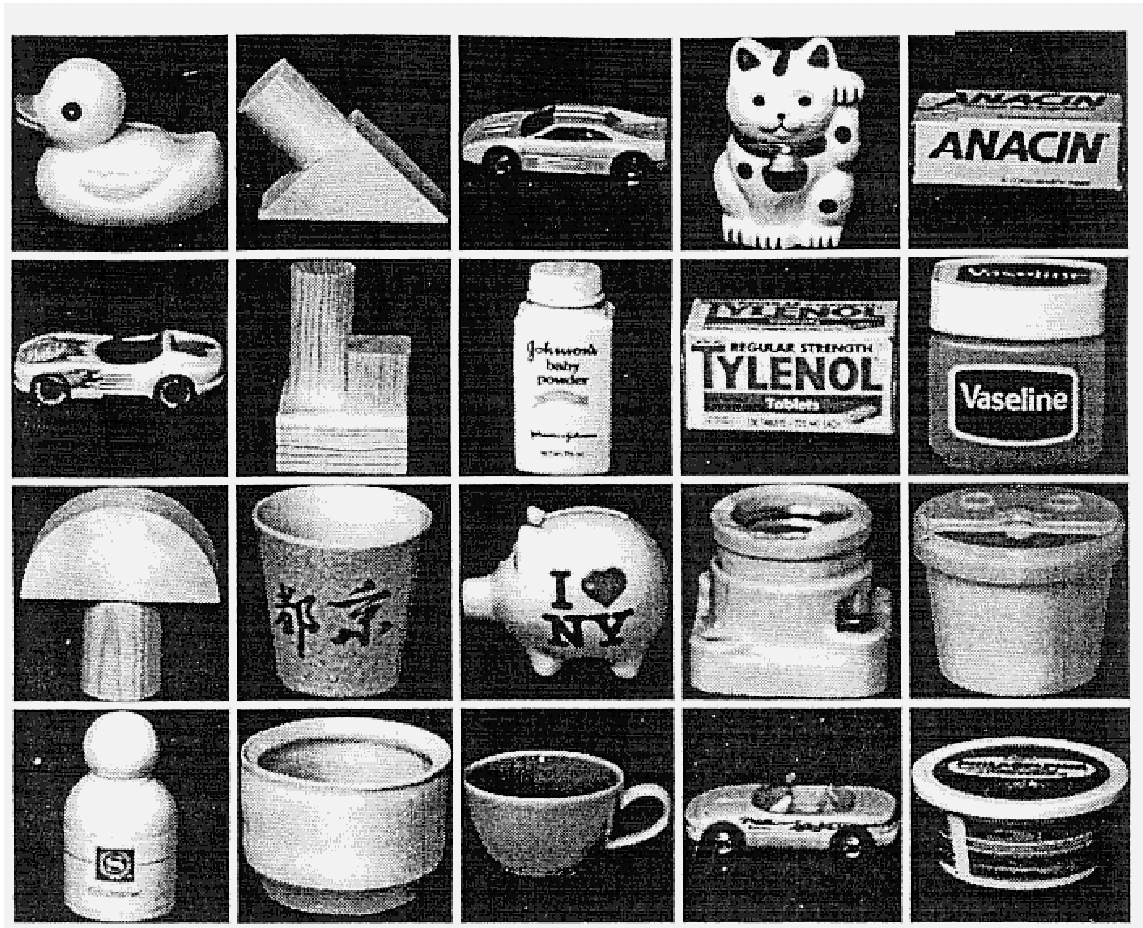
Projects all faces
onto a universal
eigenspace to “encode”
via principal components

Uses inverse-distance
as a similarity measure
for matching & recognition



# Appearance-Based Models: Parametric Eigenspaces

Murase & Nayar (1995)




Columbia “COIL” Object Database (Pose)

# Appearance-Based Models: Parametric Eigenspaces

Murase & Nayar (1995)


Columbia “COIL” Object Database (Pose)

# Bayesian Face Recognition

Moghaddam et al (1996)

*Intrapersonal* $\Omega_{\mathrm{I}}$

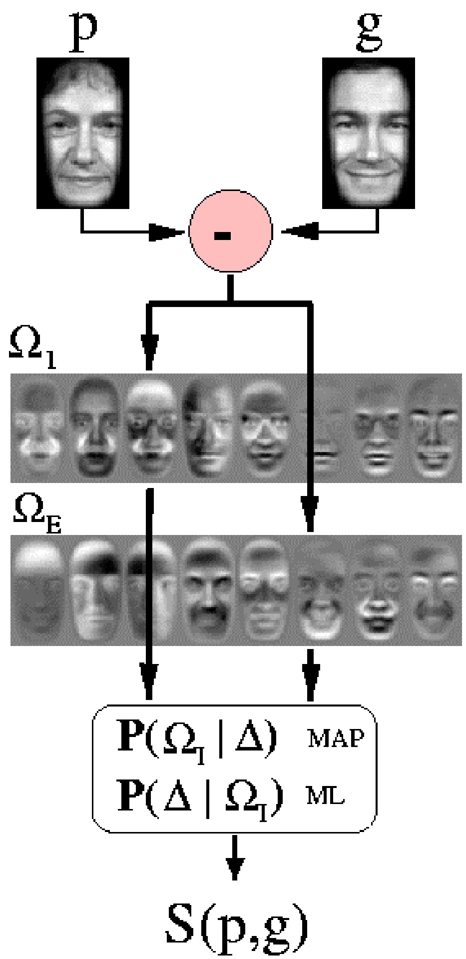
*Extrapersonal* $\Omega_{\mathrm{E}}$

$$\Omega_{\mathrm{I}} \equiv \{\Delta = x_i - x_j : L(x_i) = L(x_j)\}$$

$$\Omega_{\mathrm{E}} \equiv \{\Delta = x_i - x_j : L(x_i) \neq L(x_j)\}$$

$$\boxed{S = \frac{P(\Delta \mid \Omega_{\mathrm{I}})P(\Omega_{\mathrm{I}})}{P(\Delta \mid \Omega_{\mathrm{I}})P(\Omega_{\mathrm{I}}) + P(\Delta \mid \Omega_{\mathrm{E}})P(\Omega_{\mathrm{E}})}}$$

$P(\Delta \mid \Omega) \longrightarrow$ [Moghaddam ICCV'95]



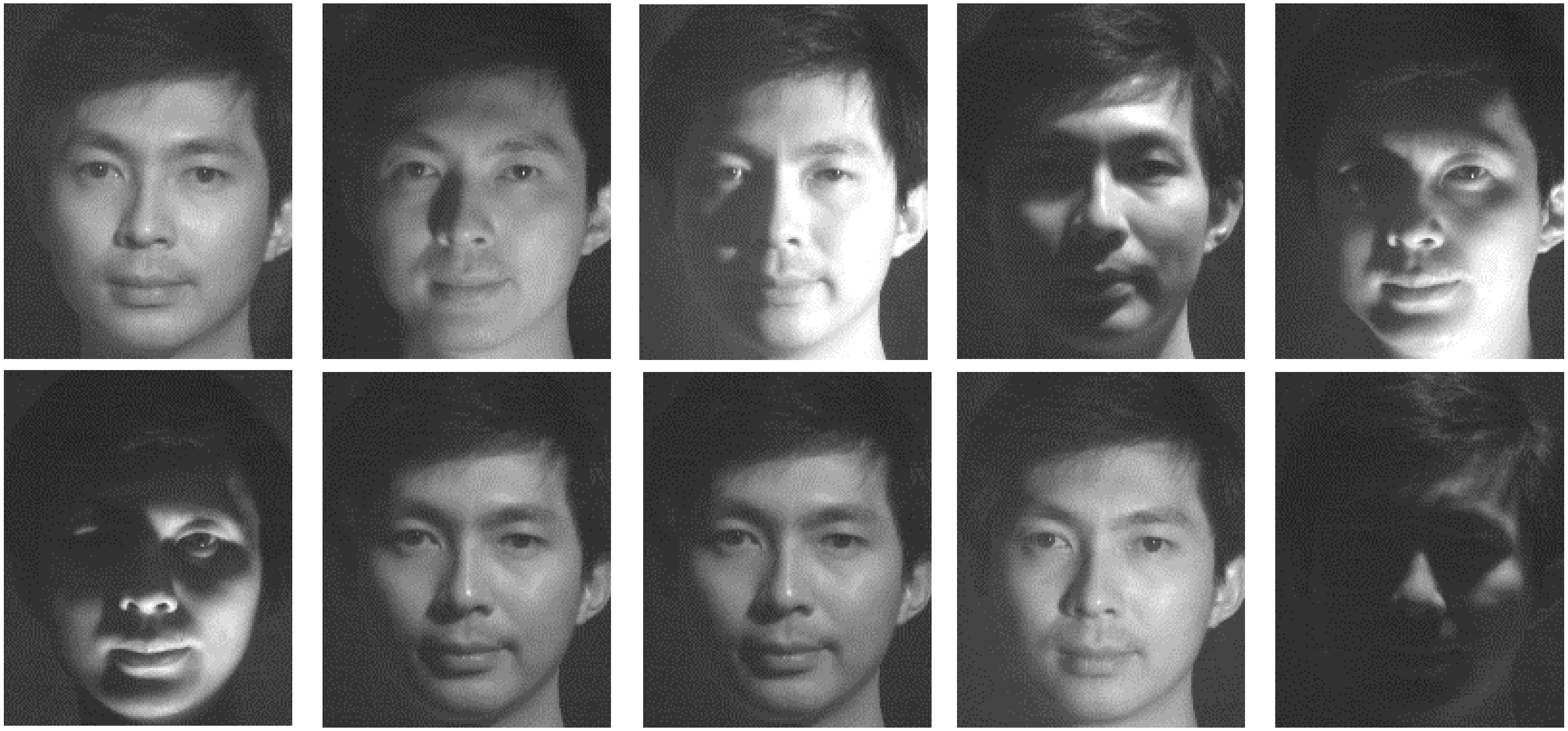
# Linear Discriminant Analysis
## LDA or “Fisherfaces”

Etemad & Chellappa (1994)

Swets & Yang (1996)

Belhumeur et al (1997)

# Illumination Variability



"The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity."

-- Moses, Adini, Ullman, ECCV '94

# Fisherfaces: Class specific linear projection

P. Belhumeur, J. Hespanha, D. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*, PAMI, July 1997, pp. 711--720.

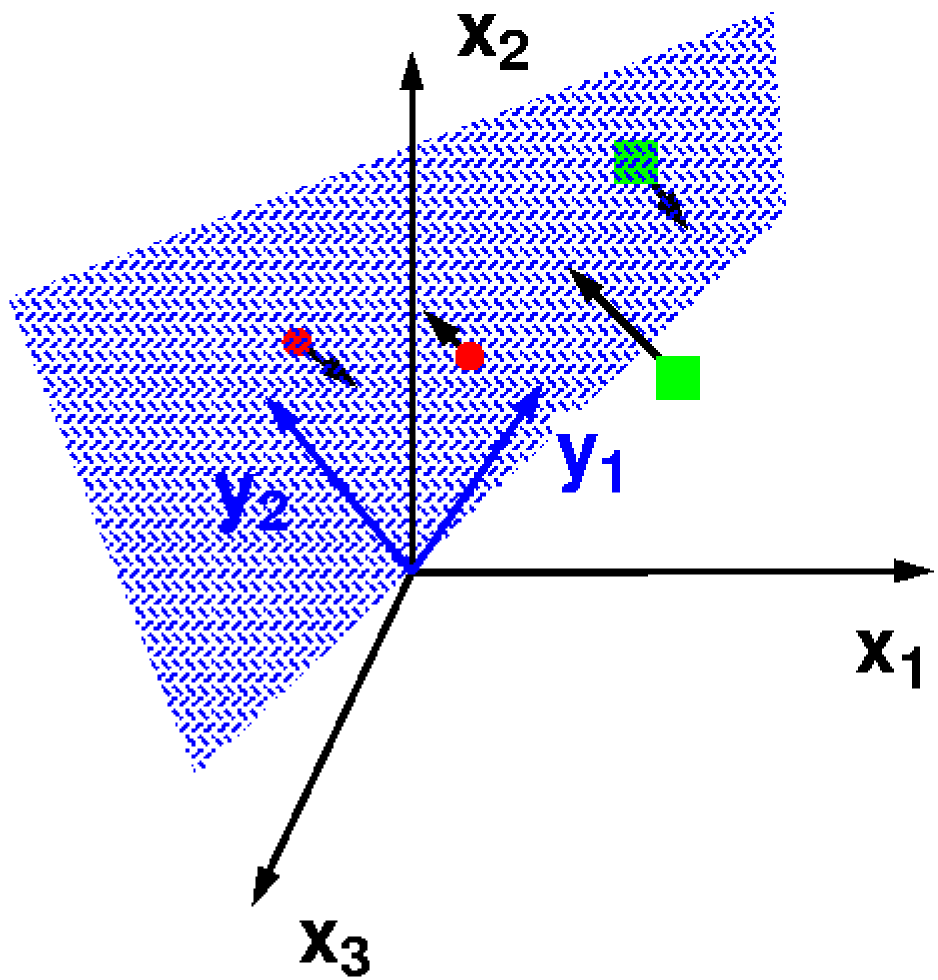
- An $n$-pixel image $\boldsymbol{x} \in \mathbf{R}^n$ can be projected to a low-dimensional feature space $\boldsymbol{y} \in \mathbf{R}^m$ by

$$\boldsymbol{y} = W\boldsymbol{x}$$

where $W$ is an $n$ by $m$ matrix.

- Recognition is performed using nearest neighbor in $\mathbf{R}^m$.

- How do we choose a good $W$?



# PCA & Fisher's Linear Discriminant

- Between-class scatter

$$S_B = \sum_{i=1}^{c} |\chi_i| (\mu_i - \mu)(\mu_i - \mu)^T$$

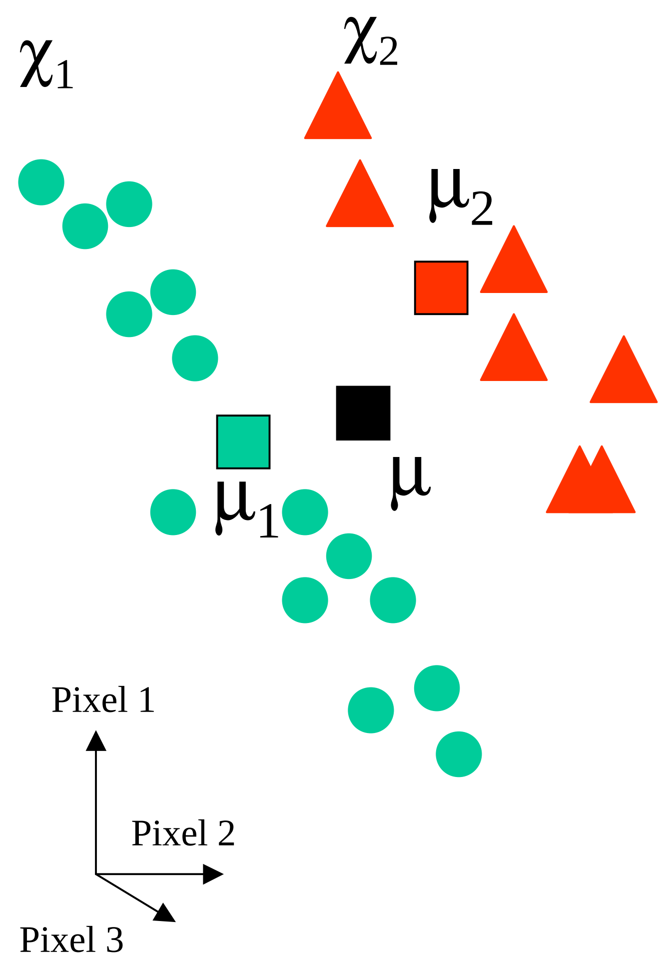
- Within-class scatter

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in \chi_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

- Total scatter

$$S_T = \sum_{i=1}^{c} \sum_{x_k \in \chi_i} (x_k - \mu)(x_k - \mu)^T = S_B + S_W$$

- Where
  - $c$ is the number of classes
  - $\mu_i$ is the mean of class $\chi_i$
  - $|\chi_i|$ is number of samples of $\chi_i$.



# PCA & Fisher's Linear Discriminant

- PCA (Eigenfaces)

Maximizes projected total scatter

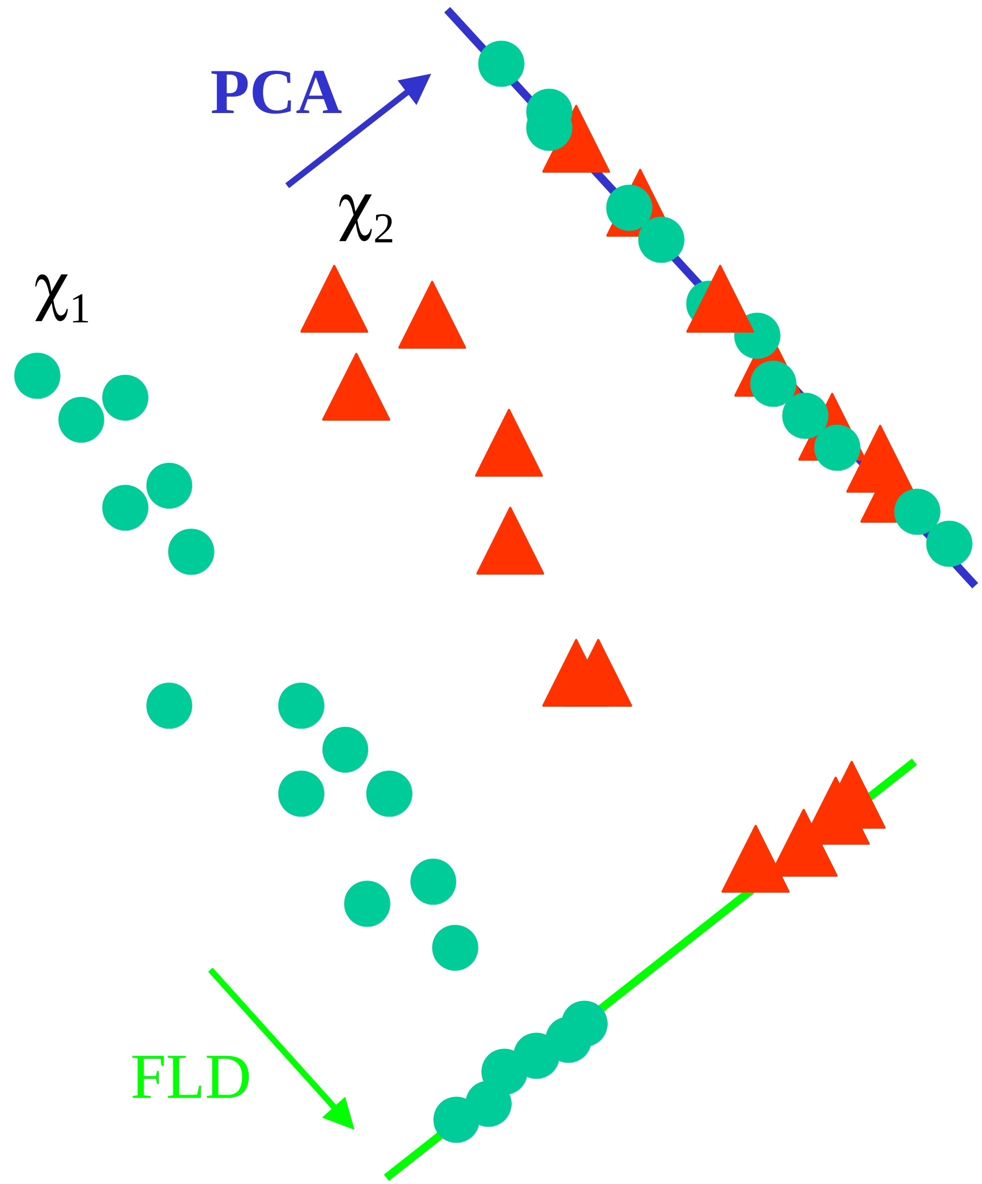
$$W_{PCA} = \arg\max_{W} \left| W^T S_T W \right|$$

- Fisher's Linear Discriminant

Maximizes ratio of projected between-class to projected within-class scatter

$$W_{fld} = \arg\max_{W} \frac{\left| W^T S_B W \right|}{\left| W^T S_W W \right|}$$

# PCA & Fisher's Linear Discriminant



# Fisherfaces

$$W = W_{fld} W_{PCA}$$

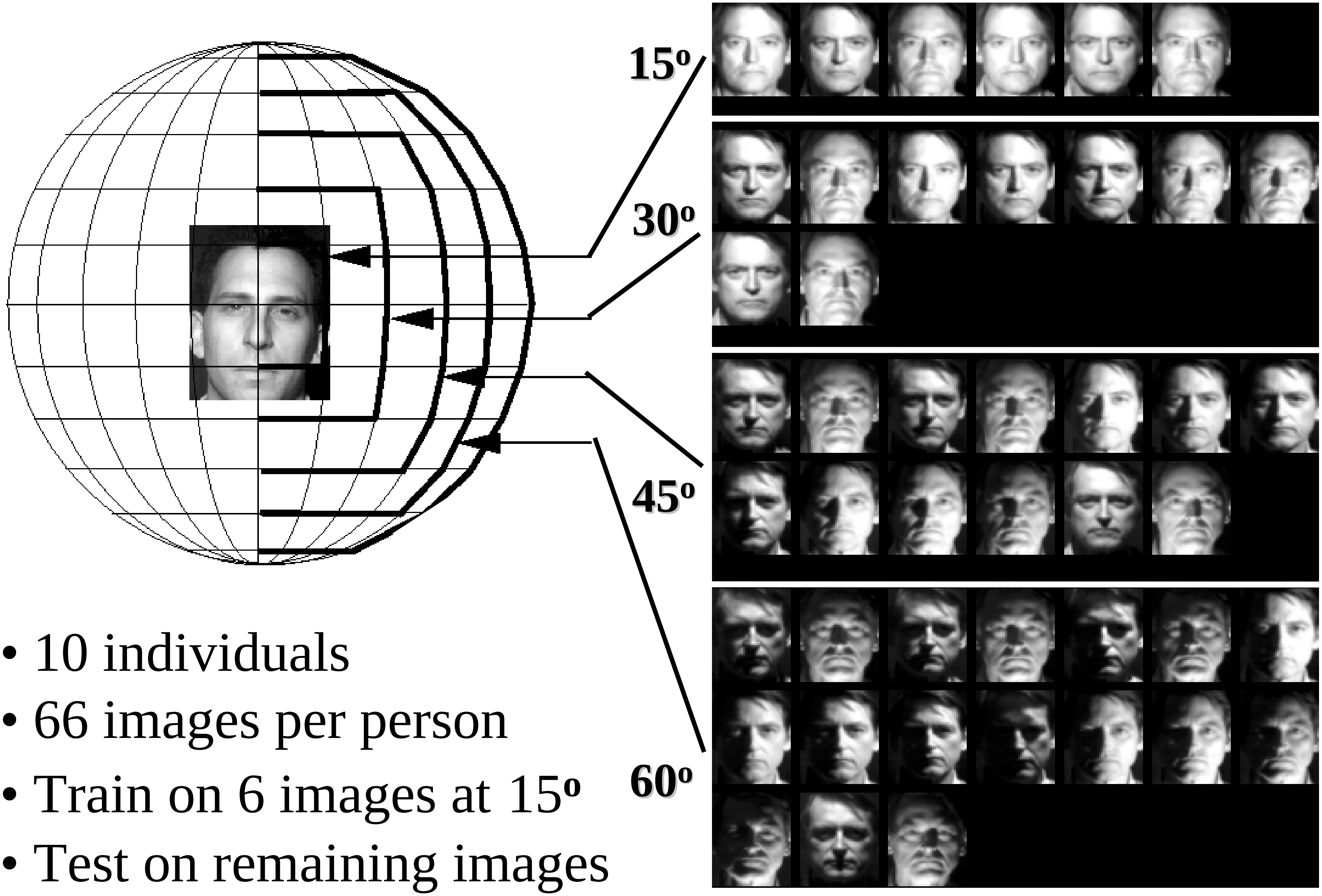
$$W_{PCA} = \arg\max_{W} \left| W^{T} S_{T} W \right|$$

$$W_{fld} = \arg\max_{W} \frac{\left| W^{T} W_{PCA}^{T} S_{B} W_{PCA} W \right|}{\left| W^{T} W_{PCA}^{T} S_{W} W_{PCA} W \right|}$$

- Since $S_W$ is rank N-c, project training set to subspace spanned by first *N-c* principal components of the training set.
- Apply FLD to *N-c* dimensional subspace yielding *c-1* dimensional feature space.

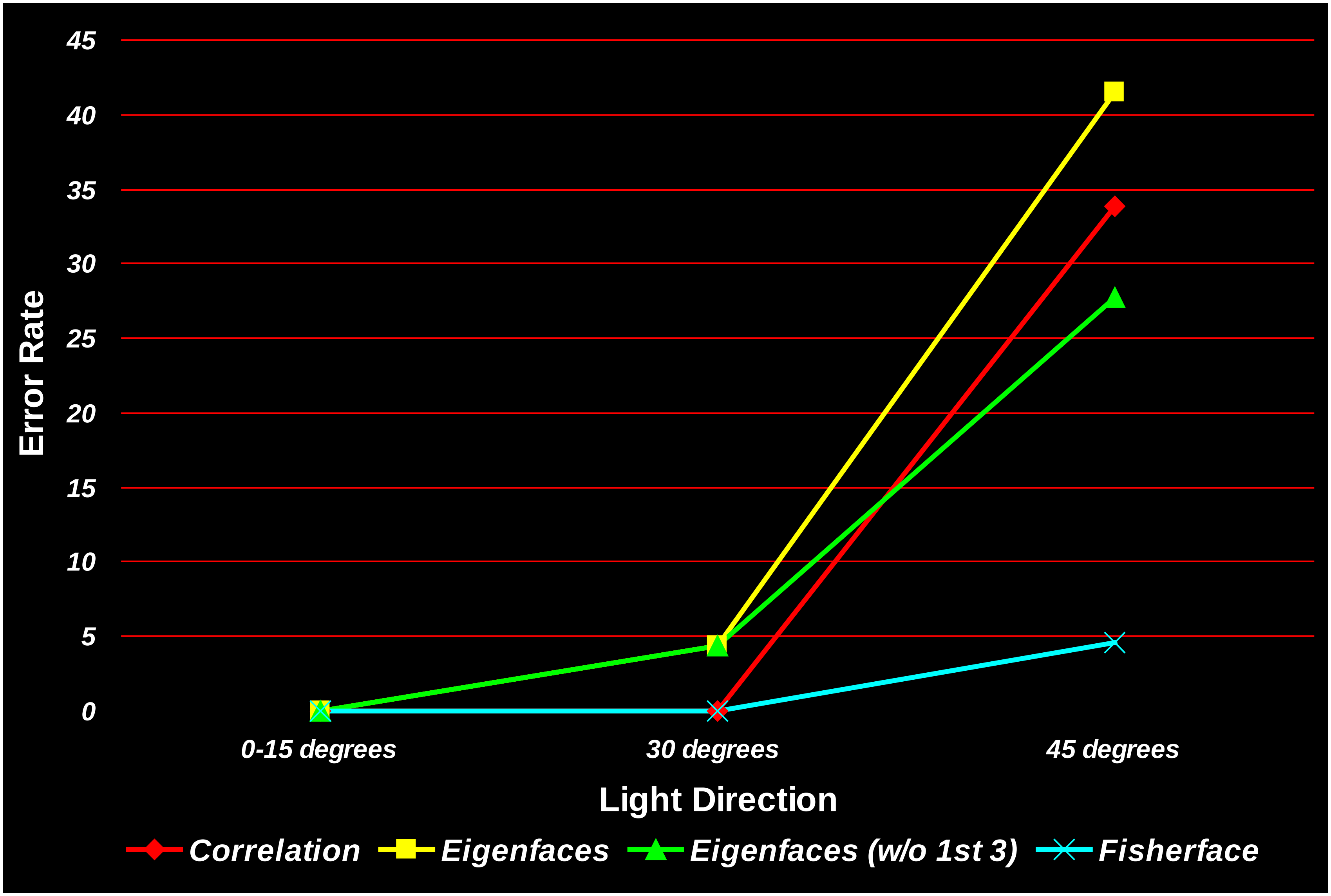
- Fisher's Linear Discriminant projects away the within-class variation (lighting, expressions) found in training set.
- Fisher's Linear Discriminant preserves the separability of the classes.

# Harvard Face Database




- 10 individuals
- 66 images per person
- Train on 6 images at 15°
- Test on remaining images

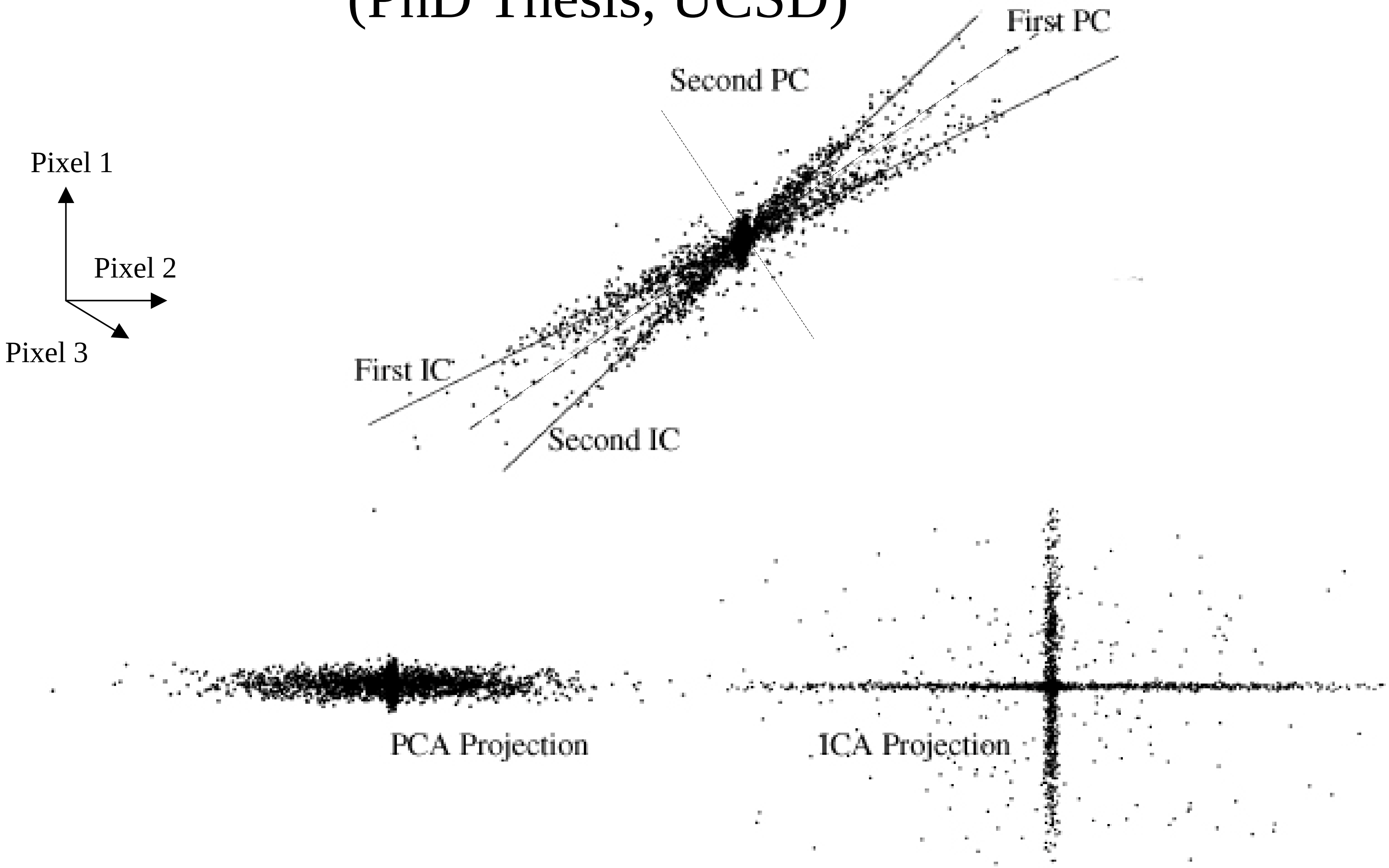
# **Recognition Results:** Lighting Extrapolation



# Independent Component Analysis

# ICA: Bartlett (1998)

## (PhD Thesis, UCSD)



# ICA: Bartlett (1998)
## Motivation

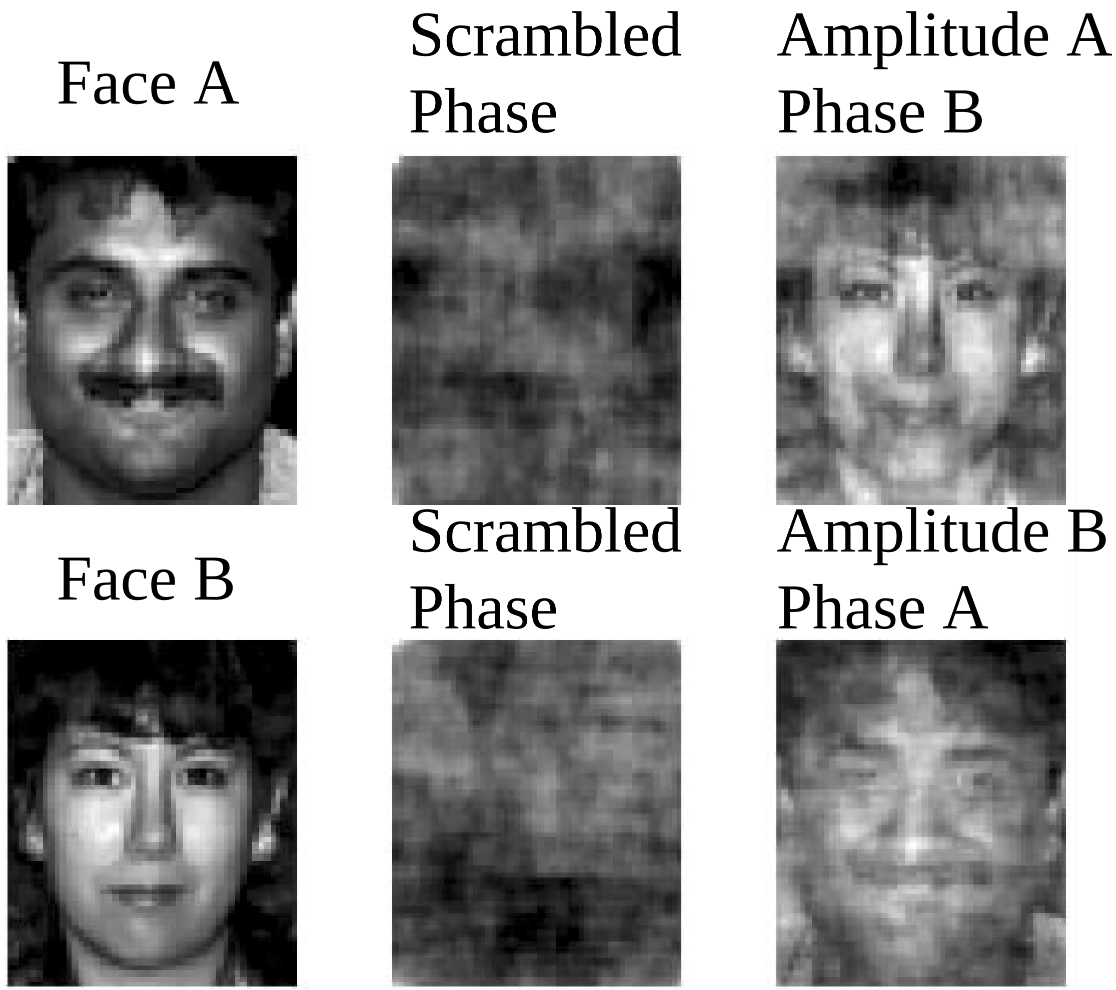
- Successful face recognition based on learning structure from the dependencies among pixels
  - Eigenfaces (Turk & Pentland, 1991)

- The more dependencies that are learned, the more structure that is learned

- ICA can be more robust to noise than PCA

- Phase is contained in the high order statistics

# ICA: Bartlett (1998)




The phase spectrum, not the power spectrum, contains the structural information that drives drives human perception. Phase is contained in the high order statistics.

# ICA: Bartlett (1998)

## *Infomax* Method

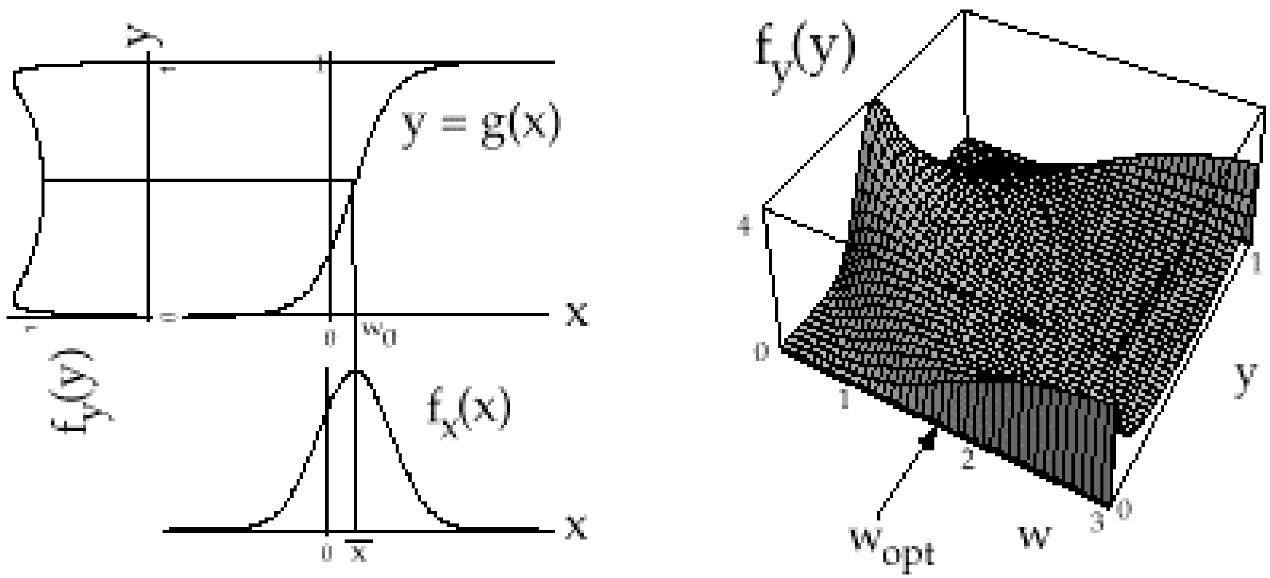



Figure 2.1: Optimal information flow in sigmoidal neurons. The input $x$ is passed through a non-linear function, $g(x)$. The information in the output density $f_y(y)$ depends on matching the mean and variance of $f_x(x)$ to the slope and threshold of $g(x)$. Right: $f_y(y)$ is plotted for different values of the weight, $w$. The optimal weight, $w_{opt}$ transmits the most information. Figure from Bell & Sejnowski (1995), reprinted with permission from *Neural Computation*, copyright 1995, MIT Press.

# ICA: Bartlett (1998)

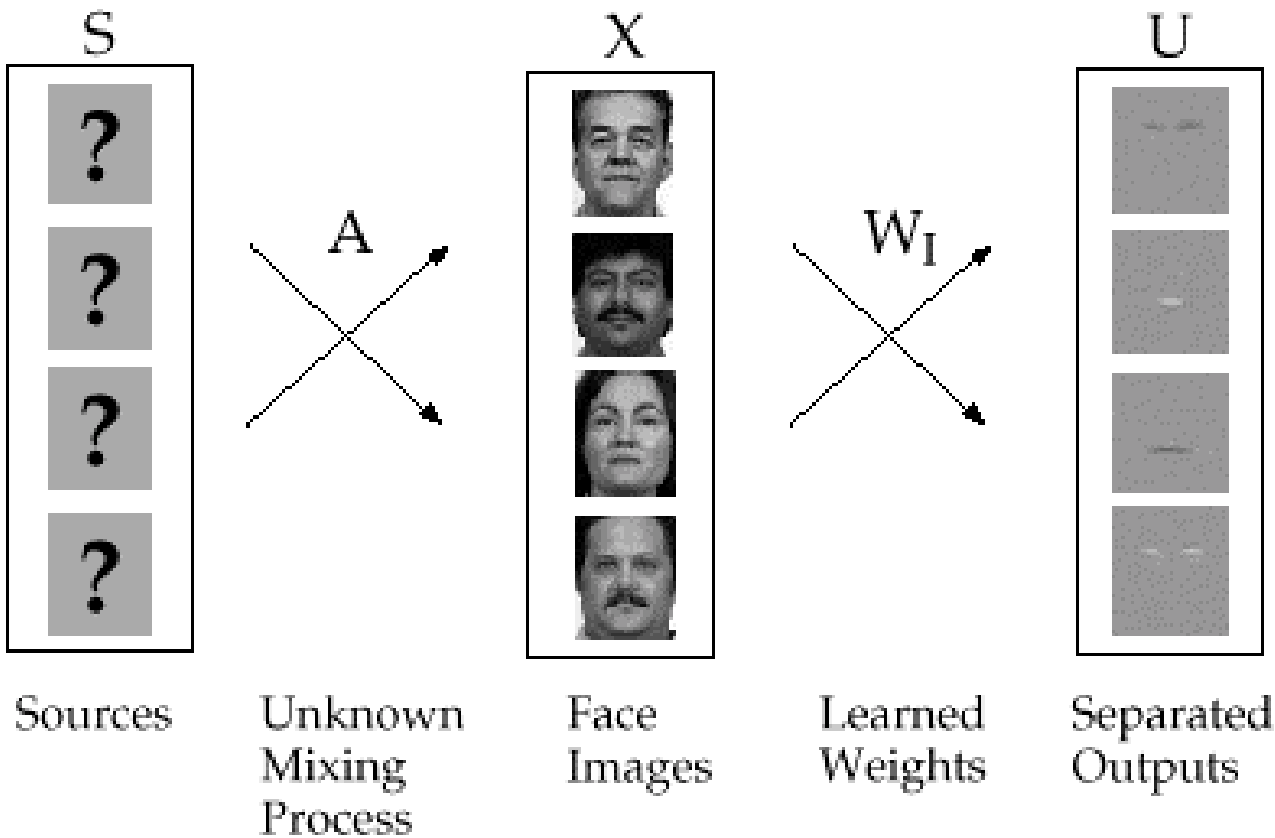



Figure 2.3: Image synthesis model. For finding a set of independent component images, the images in $X$ are considered to be a linear combination of statistically independent basis images, $S$, where $A$ is an unknown mixing matrix. The basis images were recovered by a matrix of learned filters, $W_I$, that produced statistically independent outputs, $U$.

# ICA: Bartlett (1998)

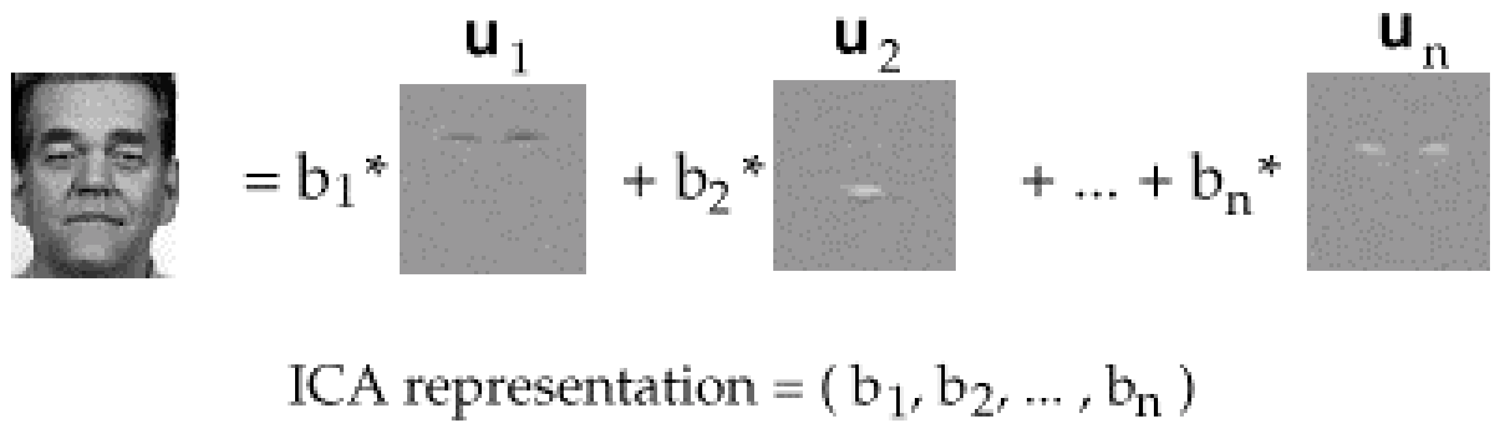



Figure 2.4: The independent basis image representation consisted of the coefficients, **b**, for the linear combination of independent basis images, **u**, that comprised each face image **x**.

# ICA:  Bartlett (1998)

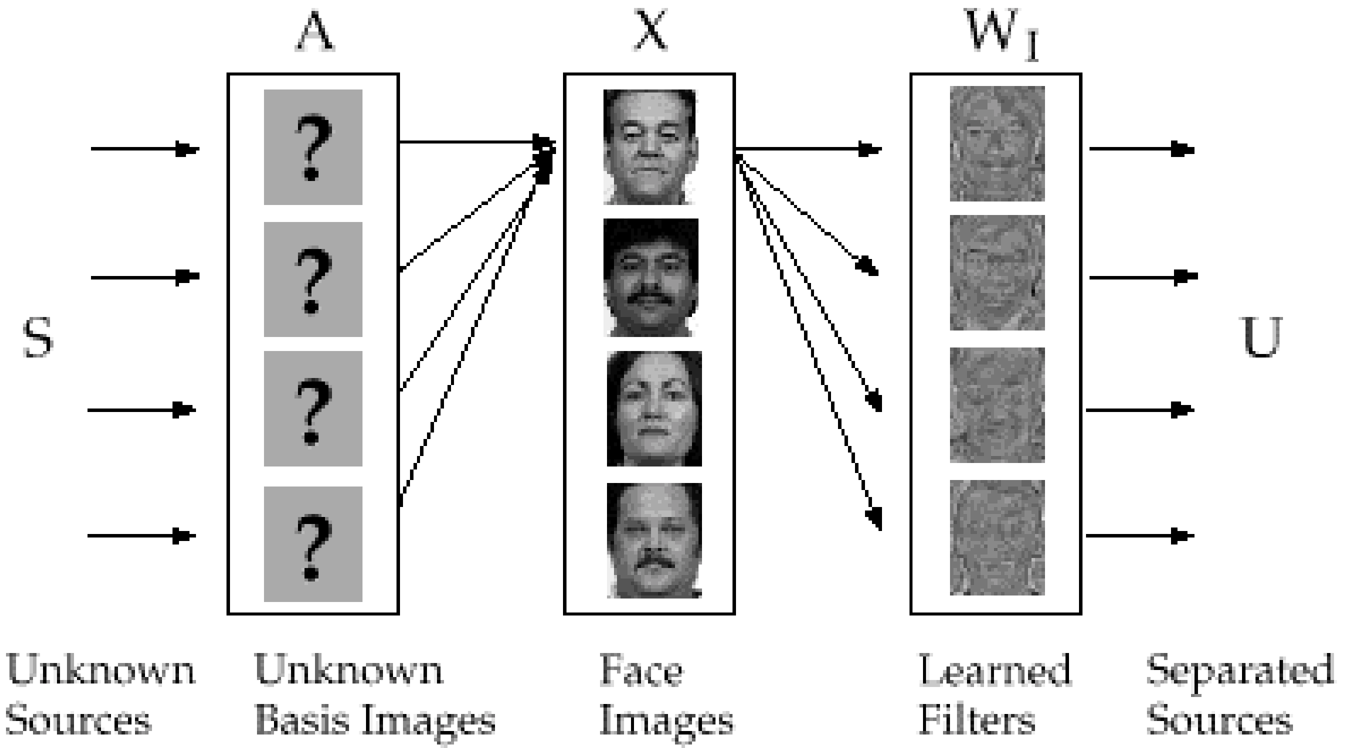
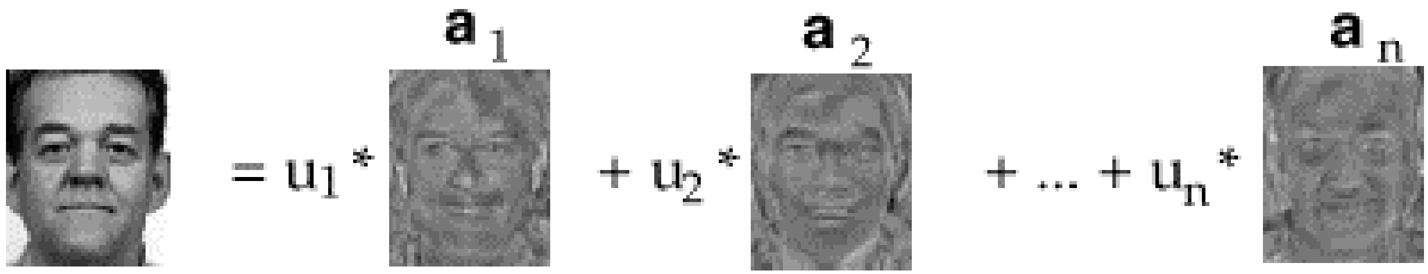



Figure 2.6: Image synthesis model for Architecture 2, based on Olshausen & Field (1996) and Bell & Sejnowski (1997). Each image in the dataset was considered to be a linear combination of underlying basis images in the matrix $A$. The basis images were each associated with a set of independent "causes", given by a vector of coefficients in $S$. The causes were recovered by a matrix of learned filters, $W_I$, which attempts to invert the unknown basis functions to produce statistically independent outputs, $U$.

# ICA: Bartlett (1998)




ICA factorial representation = $( u_1, u_2, \ldots, u_n )$

Figure 2.7: The factorial code representation consisted of the independent coefficients, $\mathbf{u}$, for the linear combination of basis images in $A$ that comprised each face image $\mathbf{x}$.

# ICA:  Bartlett (1998)

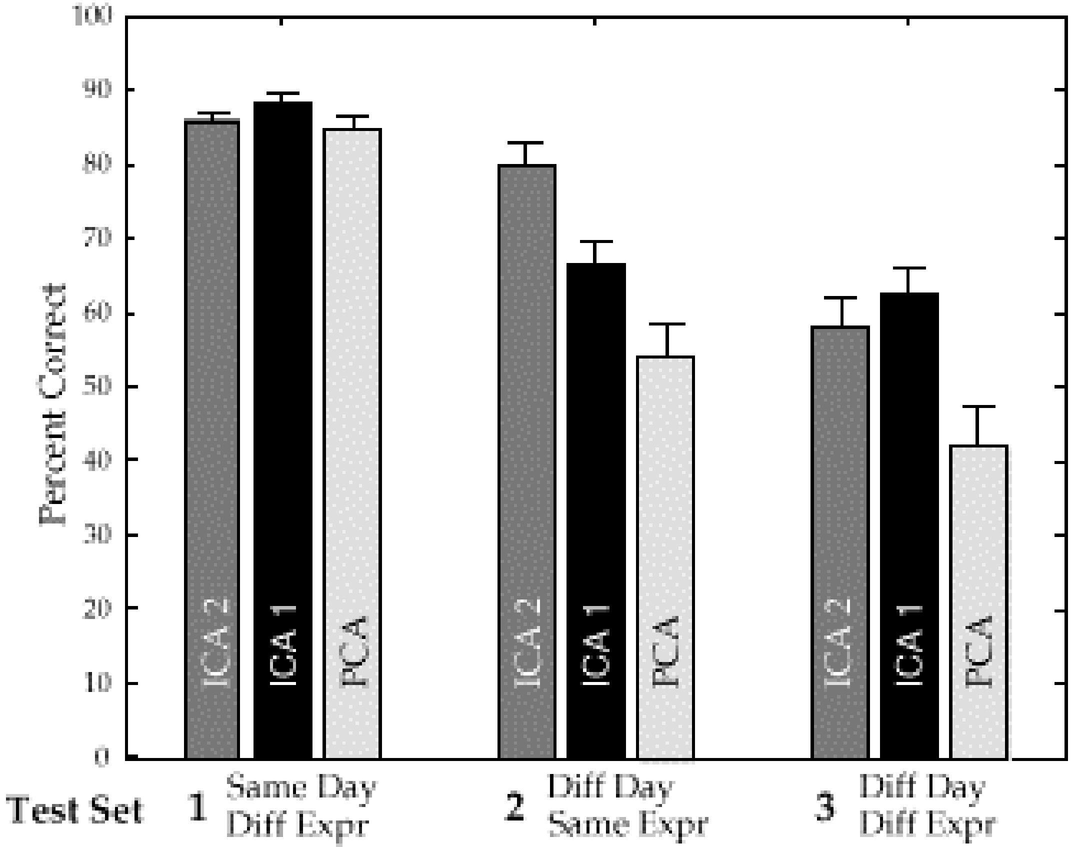
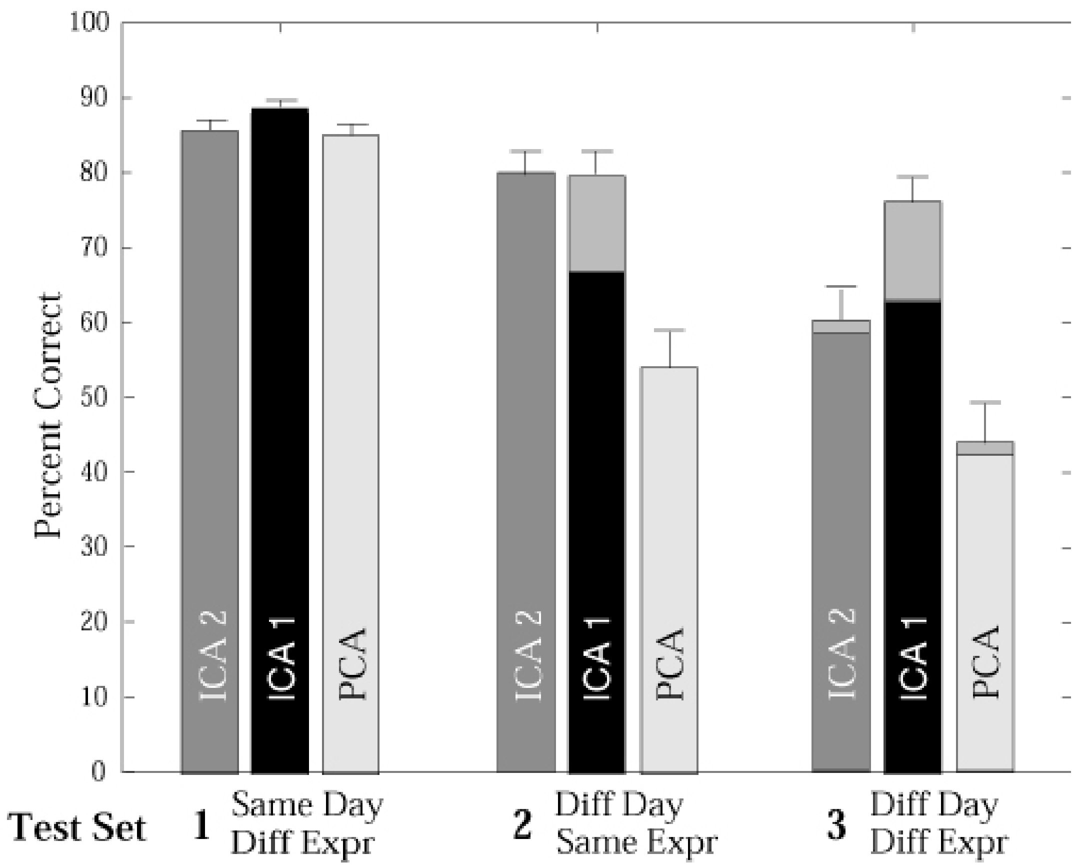



Figure 2.14: Recognition performance of the factorial code ICA representation (ICA2) using all 200 coefficients, compared to the ICA independent basis representation (ICA1), and the PCA representation, also with 200 coefficients.

# ICA: Bartlett (1998)




Improvement in recognition performance by selecting subsets of components by class discriminability. Gray extensions show improvement.

# ICA: Bartlett (1998)

## See also:


Bartlett, M.S. Face Image Analysis by Unsupervised Learning. Kluwer (2001).

- Foreword by Terrence J. Sejnowski

- For more information, see
  http://inc.ucsd.edu/~marni

# Local Feature Analysis (LFA)

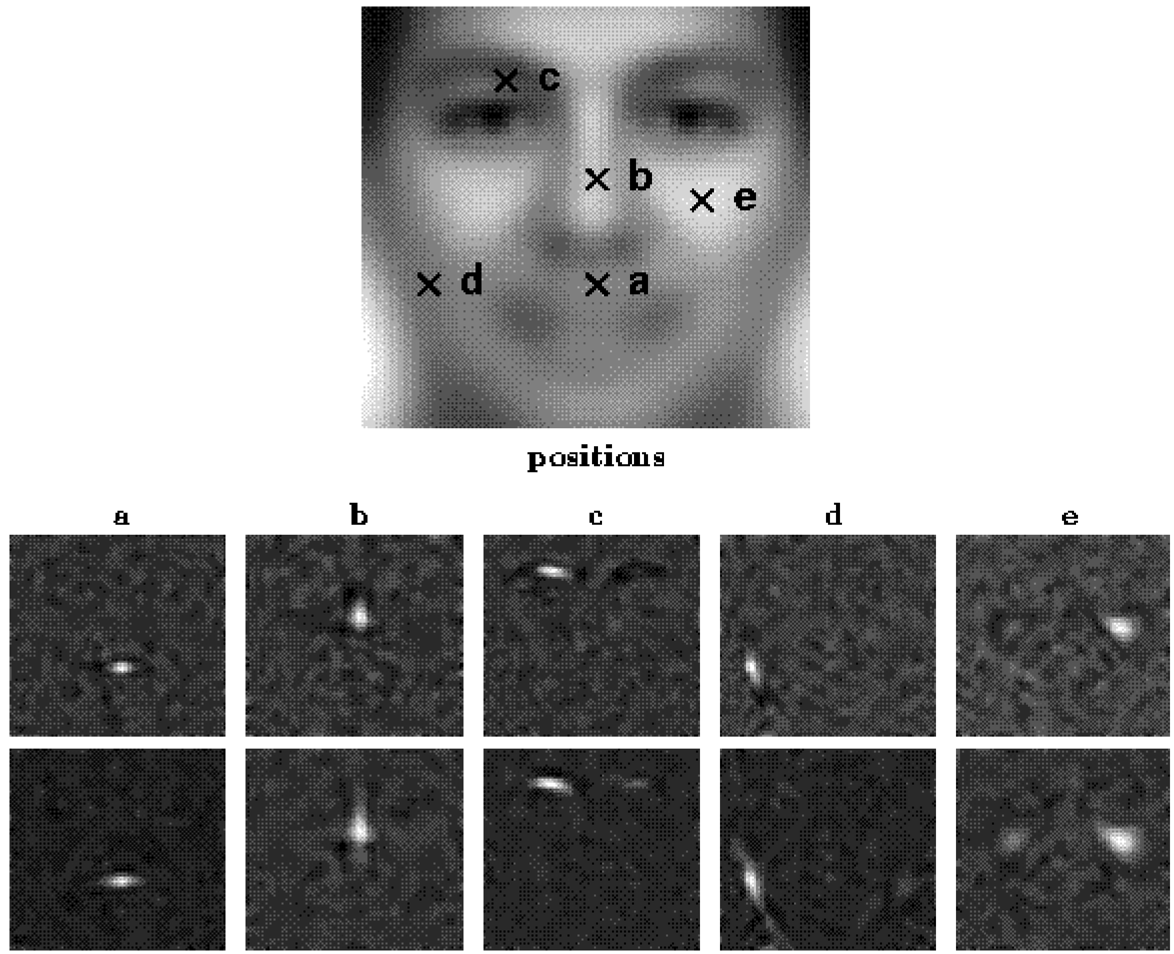
Atick & Penev (1996)

## PCA

* Global
* No Topology
* Dense

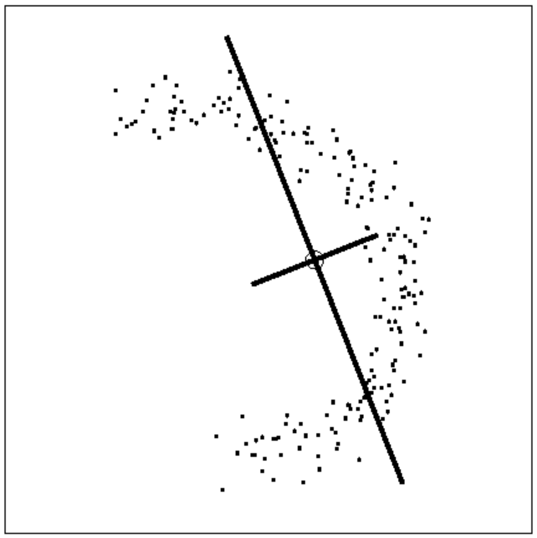
## LFA

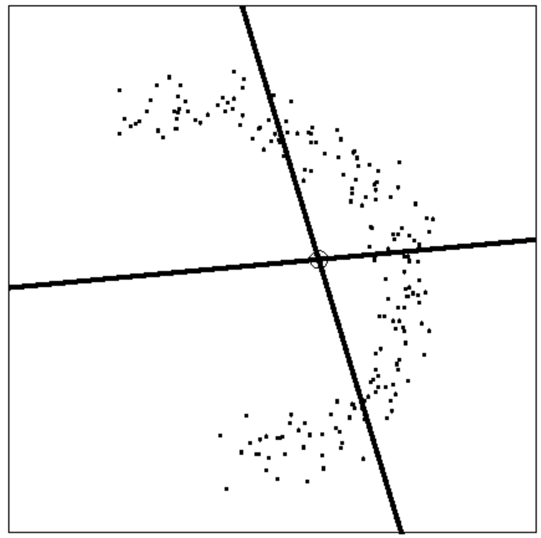
* Local
* Topographic
* Sparse

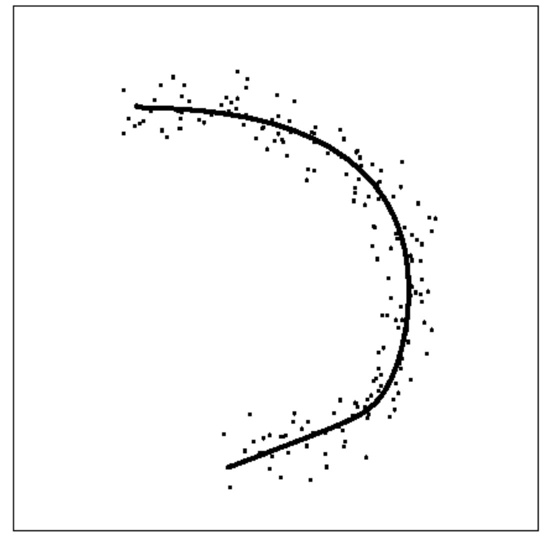



www.visionics.com

# Subspace Modeling

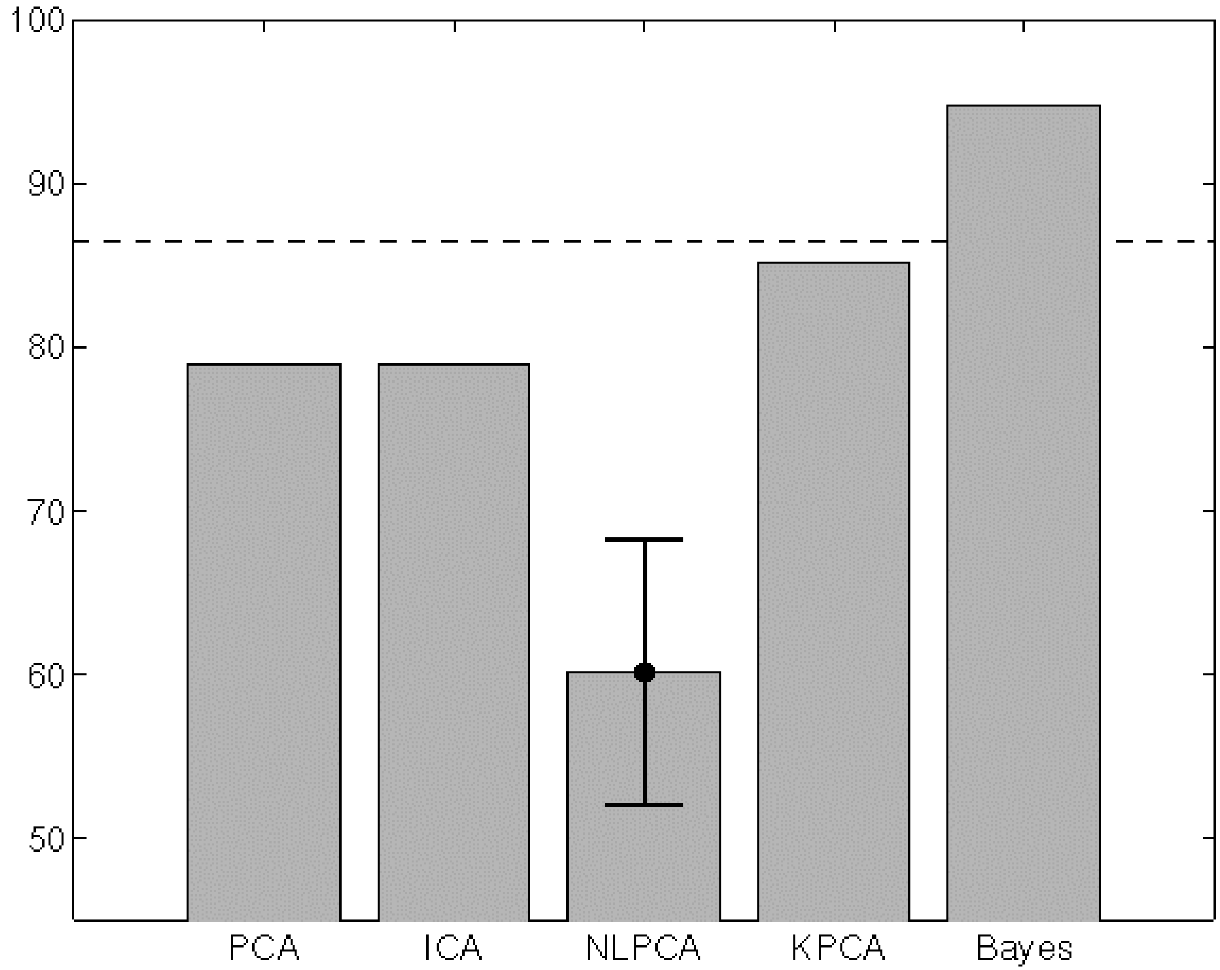



| PCA | ICA | NLPCA |
|---|---|---|
| $x \approx Uy$ | $x \approx Ay$ | $y = f(x)$ |
| $U^T U = I$ | $A^T A \neq I$ | $x \approx g(y)$ |
| $E\{y_i y_j\}_{i \neq j} = 0$ | $P(y) \approx \prod p(y_i)$ | $P(y) = ?$ |

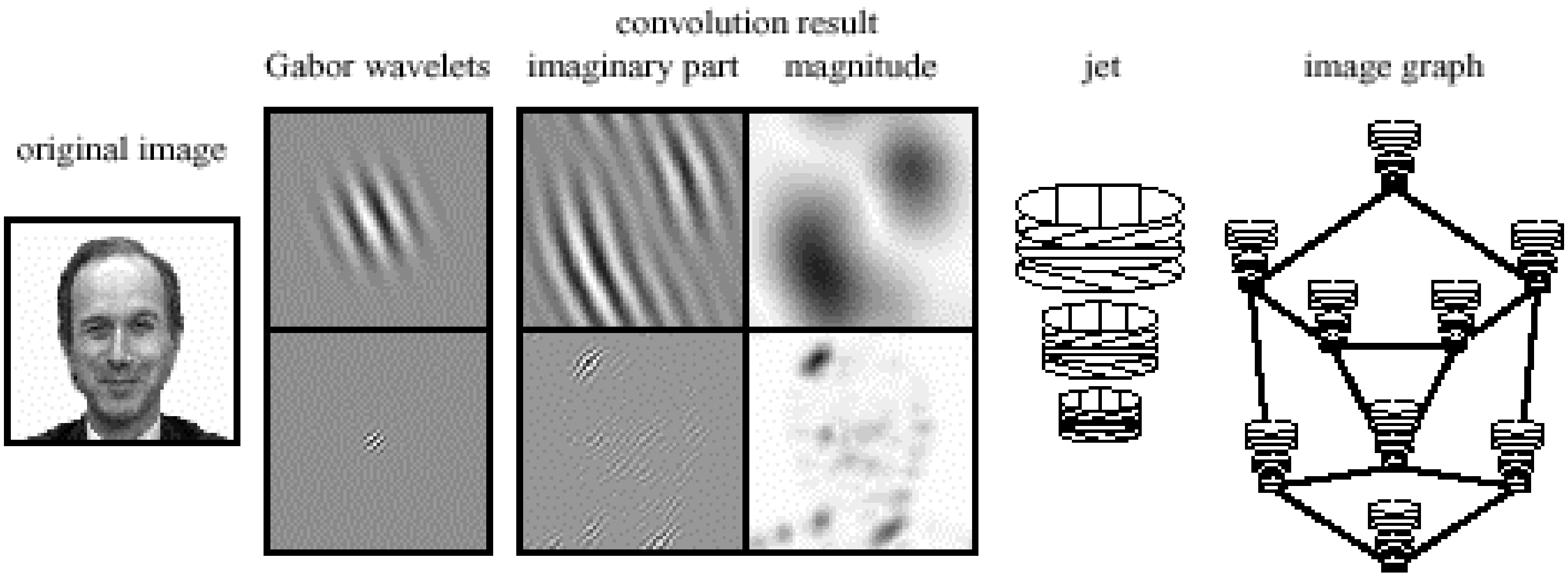
# Subspace Modeling



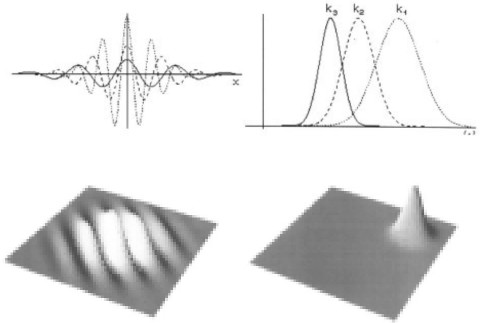

**Moghaddam ICCV'99 + PAMI (to appear)**

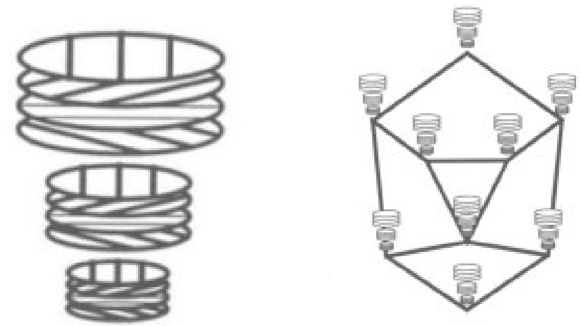
# Elastic Bunch Graphs (EBG)

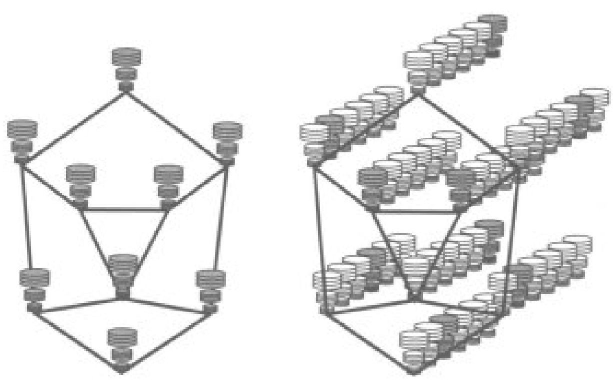
Wiskott et al (1997)

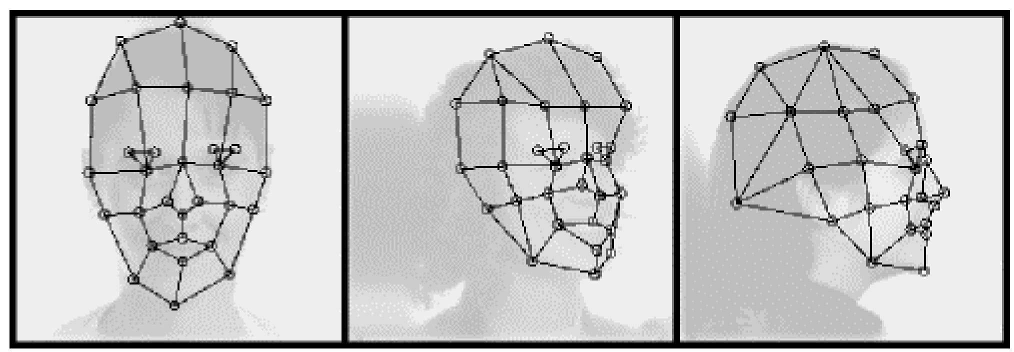


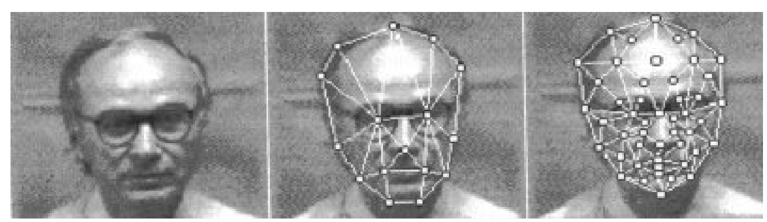
# Face Modeling with EBGs











# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# The FERET Evaluations

*courtesy of*

P. Jonathon Phillips
National Institute of Standards and Technology

***jonathon@nist.gov***

# FERET Publications

**August 1994 and March 1995 Tests:**

FERET (face recognition technology) recognition algorithm development and test report, Phillips, Rauss, and Der, Tech. Report ARL-995, 1996

The FERET database and evaluation procedure for face recognition algorithms, Phillips, Wechsler, Huang, and Rauss, Image and Vision Computing J. 16(5):295-306, 1998.

**September 1996 Test:**

*The FERET evaluation methodology for face recognition algorithms, Phillips, Moon, Rauss, and Rizvi*, IEEE trans. PAMI 22(10):1090-1104 October 2000

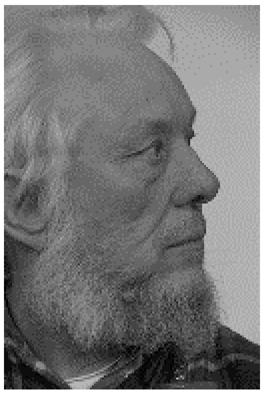
*A verification protocol and statistical performance analysis for face recognition algorithms, Rizvi, Phillips, and Moon, CVPR'98; Invited paper, Special Issue of Image and Vision Computing J. on face and gesture recognition.*

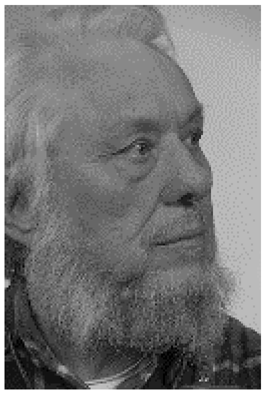
# FERET Database

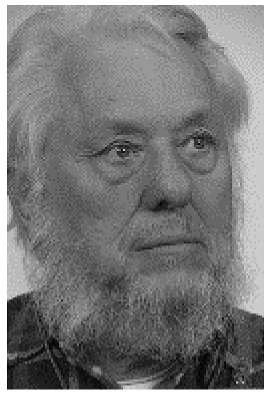
- George Mason U. / ARL
- September 1993 - August 1996
- Standard database for developing and testing
- Development portion
  - Given to researchers
- Sequestered portion
  - Used for testing

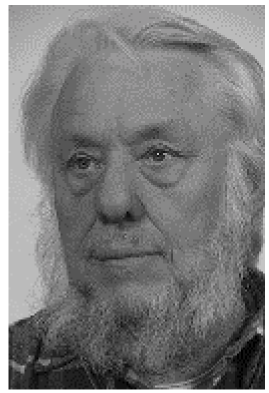
# FERET Database

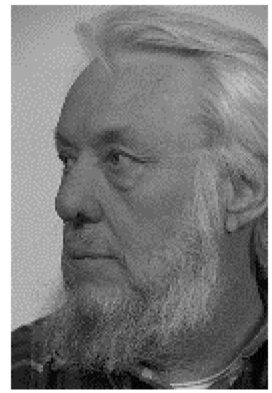
## Sample image set

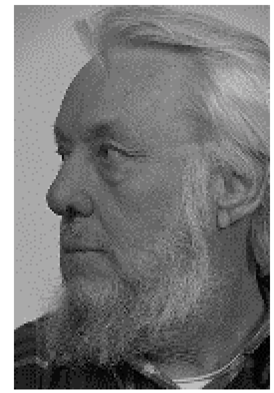

pr


hr

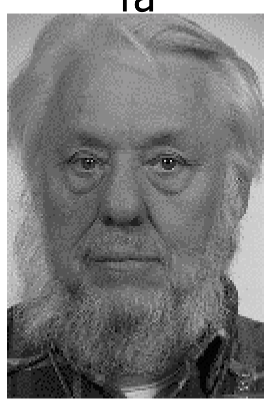

qr


ql

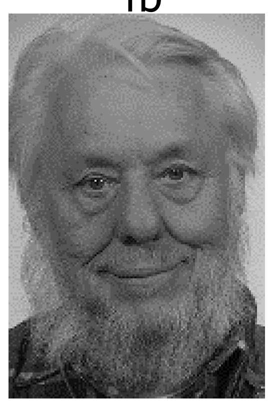

hl


pl

fa



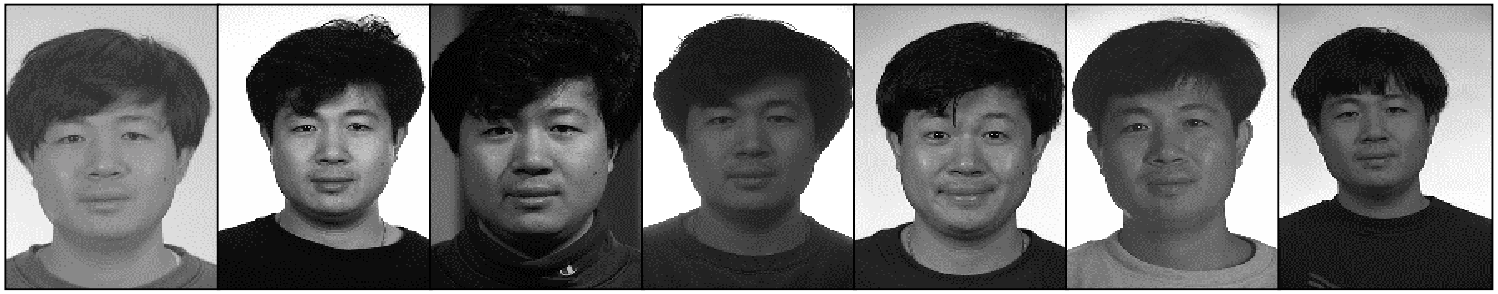
fb



# Size of FERET Database

Images in the FERET database as of September 1996

| | |
|---|---|
| 14,126 | Images |
| 1,564 | Sets of images |
| 1,199 | Individuals |
| 365 | Duplicate sets |
| 503 | Sets of images in the development portion |



Variations among duplicate images

# FERET Evaluations

- Independent evaluation of face recognition algorithms
- Test allows
  - Assessment of state-of-the-art
  - Identification of future research directions
- Series to advance face recognition
  - Each test more advanced
  - Supported by data collections
  - Based on previous tests

# Three Evaluations

- Aug 94 test
  - August 1994
- Mar 95 test
  - March 1995
  - November 1995
  - August 1996
- Sep 96
  - September 1996
  - March 1997

# Face Recognition
## State-of-the-art: September 93

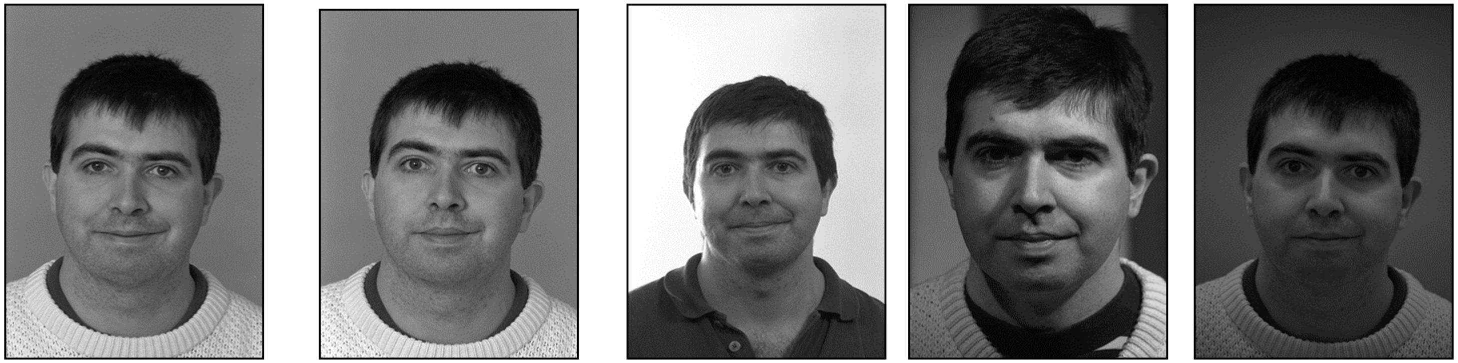
- No method to assess state-of-the-art
  - Algorithm results reported on small (<50) internal databases
- Algorithms NOT fully automatic
- Leading researchers reported on following databases
  - Pentland
    - Database of 7,500
    - Collected in a booth, eyes registered
  - Wilder
    - Database of 250 images
    - Very controlled conditions, chin registered
  - von der Malsburg
    - Database of 100 images
    - Controlled size, some variation in rotation and pose

# Groups Tested

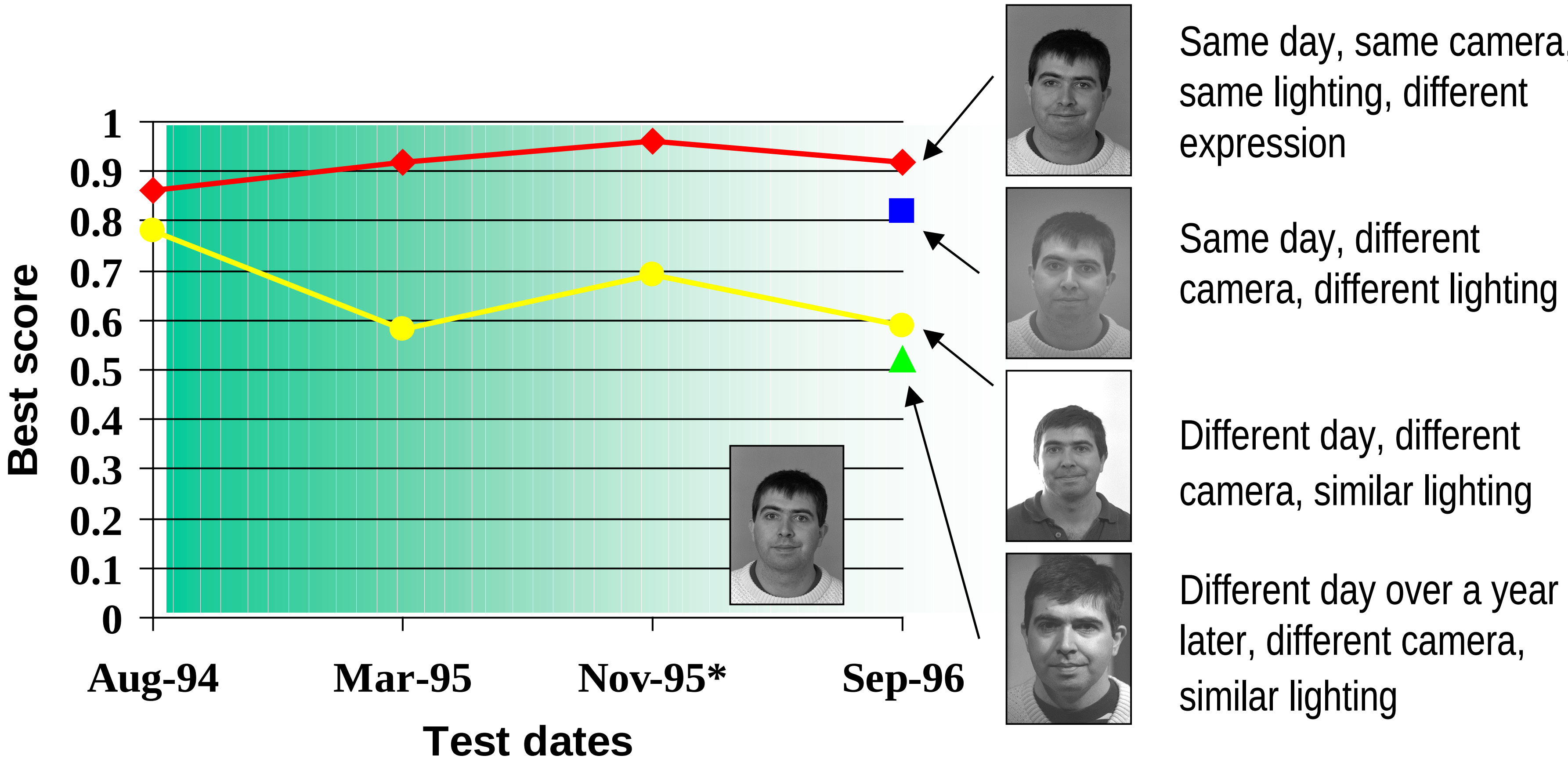
| | | Test date | | | | | |
|---|---|---|---|---|---|---|---|
| Version of test | Group | Aug 94 | Mar 95 | Nov 95 | Aug 96 | Sep 96 | Mar 97 |
| Fully automatic | MIT | * | * | | * | * | |
| | Rockefeller U. | | | * | | | |
| | Rutgers | * | | | | | |
| | TASC | * | | | | | |
| | USC | * | * | | | | * |
| Eye coordinates given | Baseline PCA | * | | | | | |
| | Excalibur | | | | | * | |
| | MSU | | | | | * | |
| | Rutgers | | | | | * | |
| | UMD | | | | | * | * |

# Test Size

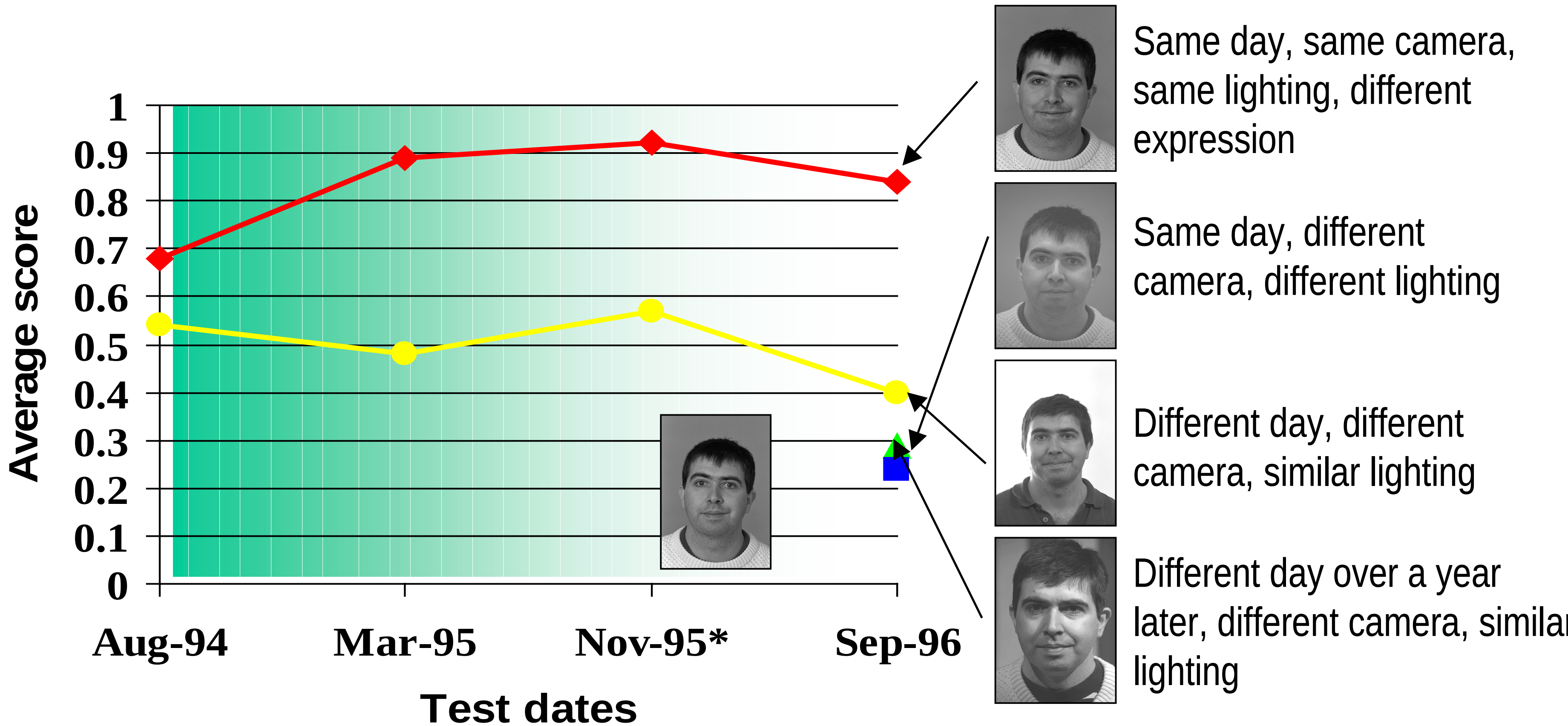
| | Gallery images | FB probes | Duplicate I probes | Duplicate II probes | fc probes |
| --- | --- | --- | --- | --- | --- |
| Aug 94 test | 317 | 316 | 50 | 0 | 0 |
| Mar 95 test | 831 | 780 | 463 | 0 | 0 |
| Sep 96 test | 1196 | 1195 | 722 | 234 | 194 |



# Best Scores by Image Type

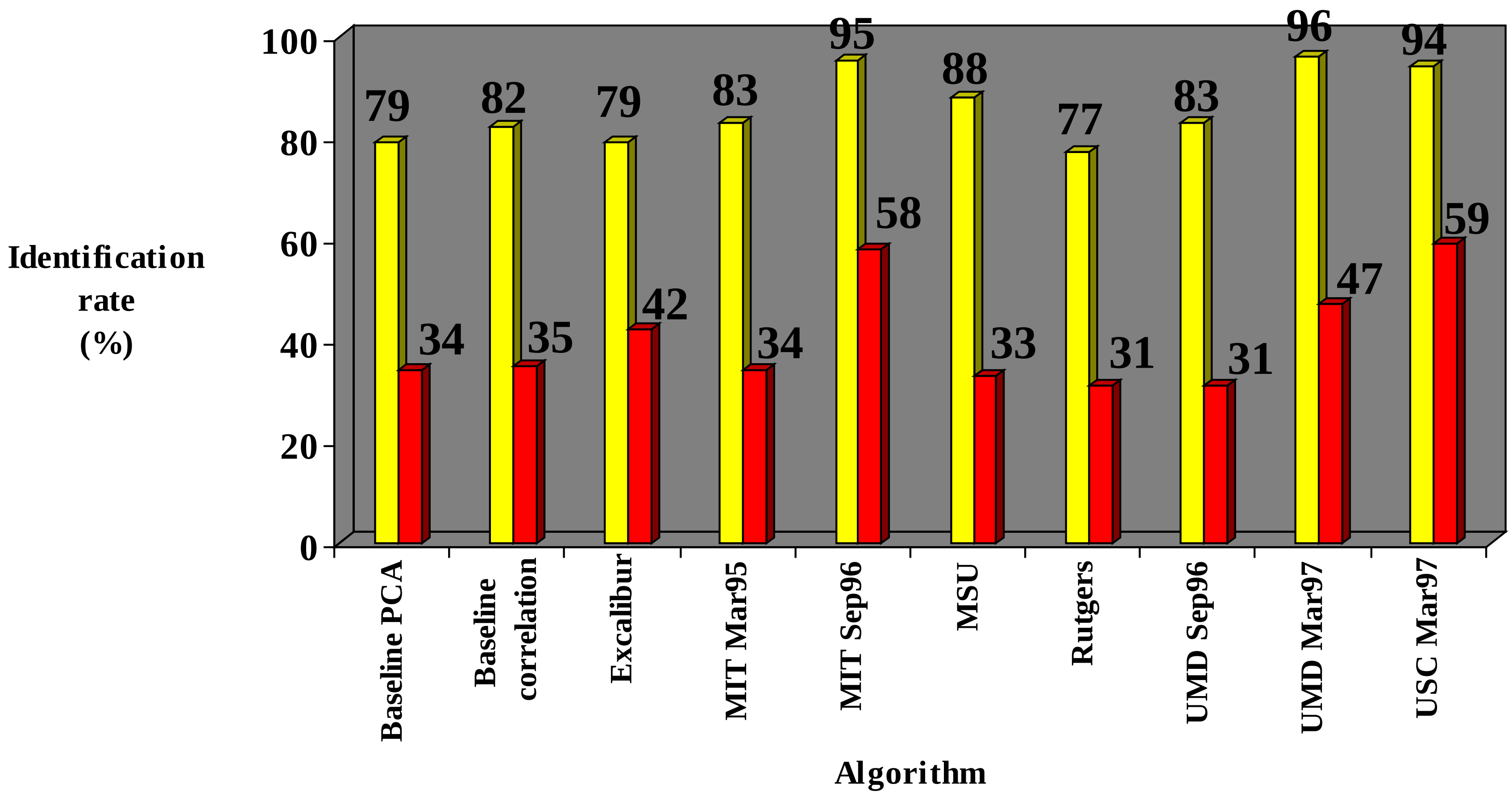


# Average Scores by Image Type



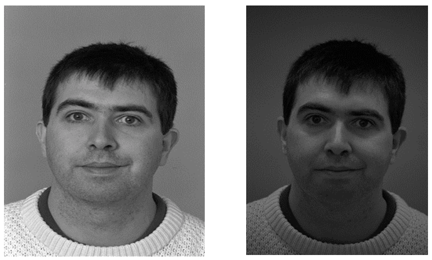
# Identification Performance

Identification rate for FB and duplicate probes
(gallery: 1196, FB probes: 1195, Dup I probes: 722)



# Conclusions from FERET

- Further research directions:



Lighting changes

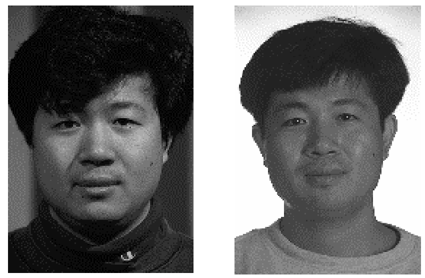
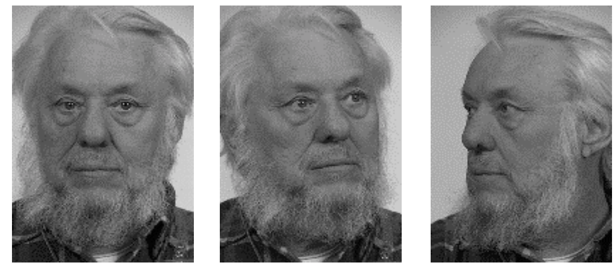


Image taken 1+ years apart



Pose changes

- What are the critical factors that influence performance?
  - Gallery and probe sets
  - Algorithm design
  - Algorithm implementation

# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# From Few To Many: Illumination Cones for Recognition

Image variability confounds recognition:
Images of a single person under variable lighting and viewpoint

➡ ***Model*** image variability due to lighting and viewpoint using a ***small*** number of captured images.

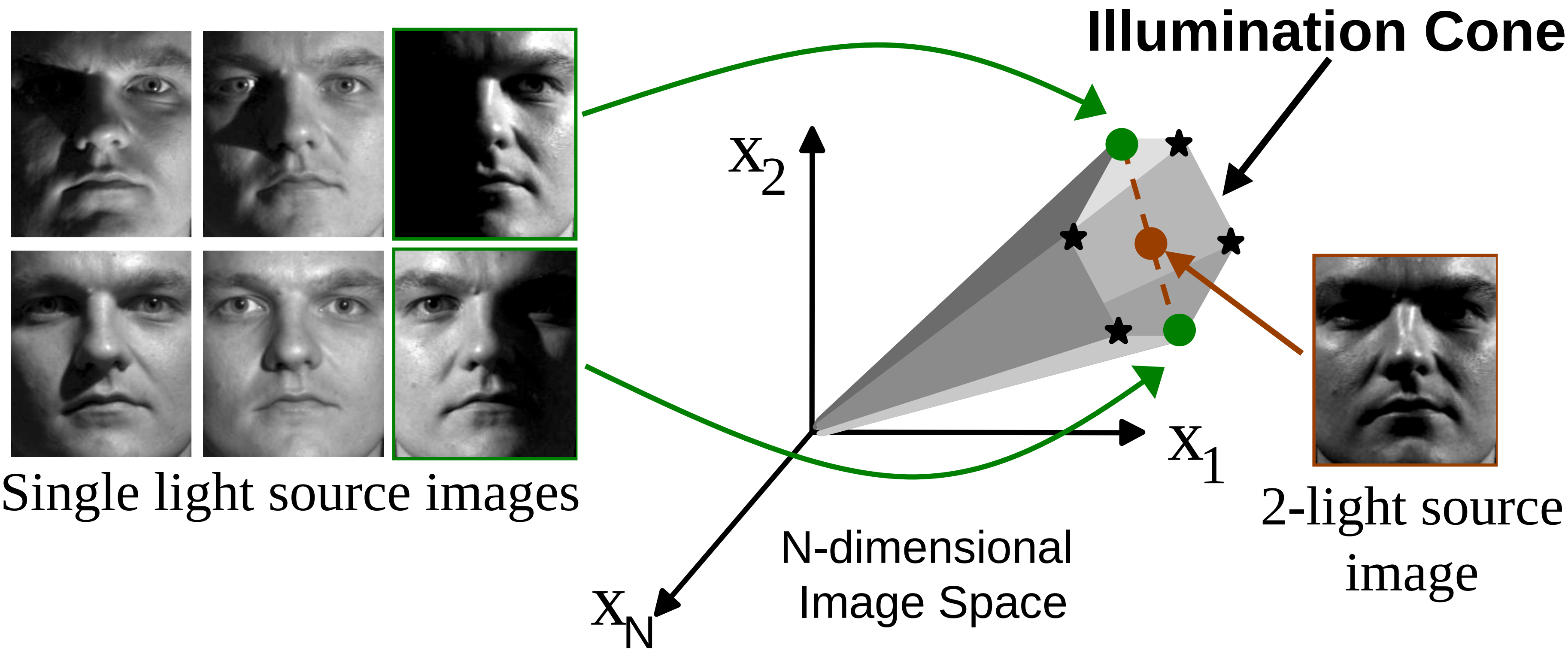
(Georghiades, Belhumeur, Kriegman, PAMI, June 2001, pp. 643--660)

# The Illumination Cone

What is the set of *n*-pixel images of an object under all possible lighting conditions (but fixed pose)?

**Proposition:** *Due to the superposition of images, the set of images is a* **convex cone** *in the image space.*

(Belhumeur and Kriegman, IJCV, July ‘98)




Single light source images

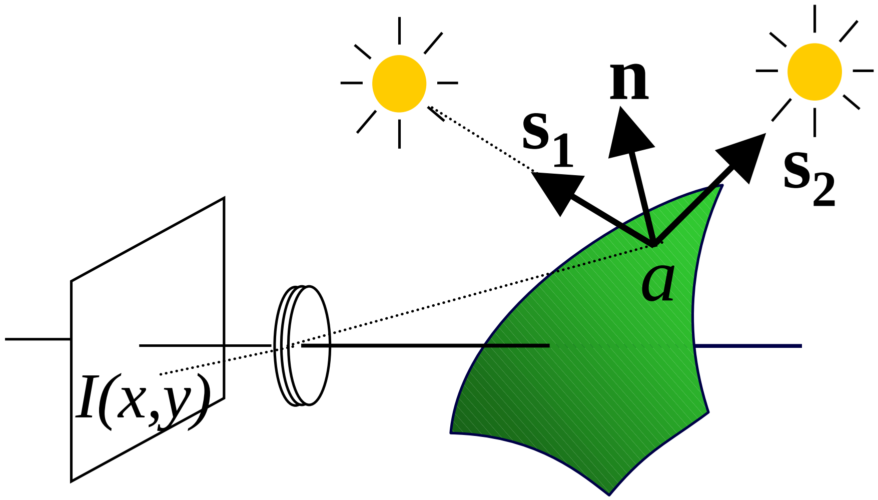
2-light source image

[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

# Generating the Illumination Cone

Assume:

- Objects have Lambertian (matte) reflectance functions.
- Objects have convex shape.
- Point light sources at infinity.
- Orthographic projection.




At image location *(x,y)* the intensity of a pixel *I(x,y)* is

$$I(x,y) = \sum_i \max\left( a(x,y)\, \mathbf{n}(x,y) \bullet \mathbf{s}_i, 0\right)$$

where

- *a(x,y)* is the albedo of the surface *f (x,y)* projecting to *(x,y)*.
- $\mathbf{n}(x,y) = [f_x(x,y), f_y(x,y), -1] \Big/ \sqrt{f_x^2(x,y) + f_y^2(x,y) + 1}$ is the unit surface normal.
- $\mathbf{s}_i$ are the directions and strengths of the light sources.

[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

# Generating the Illumination Cone (II)

- For Lambertian surfaces, the illumination cone is determined by the 3D linear subspace (of the function space) **B***(x,y)*, where

$$I(x,y) = \sum_i \max\left( \overbrace{\boxed{a(x,y)\,\mathbf{n}(x,y)}}^{\mathbf{B}(x,\,y)} \cdot \mathbf{s}_i, 0 \right)$$

- When no shadows, then $I(x,y) = \sum_i \mathrm{B}(x,y) \bullet \mathbf{s}_i$

- With a small number of captured (single light source) images with no shadows, use least-squares to find best 3D linear subspace **B***(x,y)*, subject to the constraint $f_{xy}=f_{yx}$. (The constraint forces the estimated **B***(x,y)* to correspond to a surface.)

Original (Training) Images

Least-squares

3D linear subspace, **B***(x,y)*

Integrate

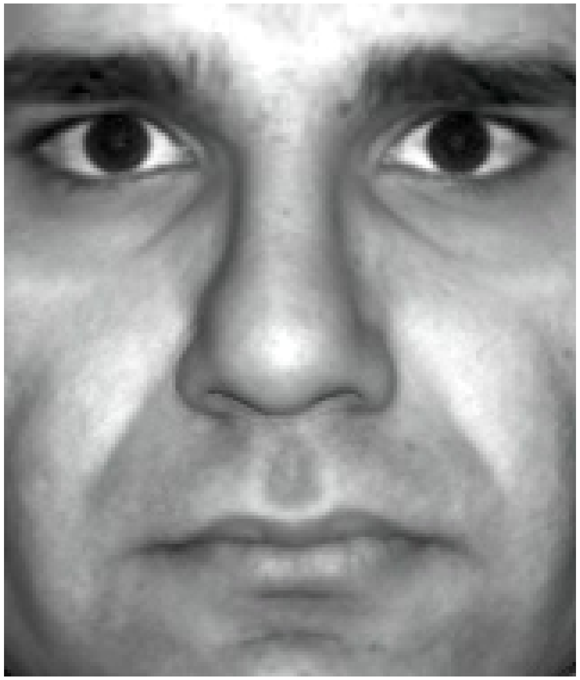
Surface, $f(x,y)$

[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

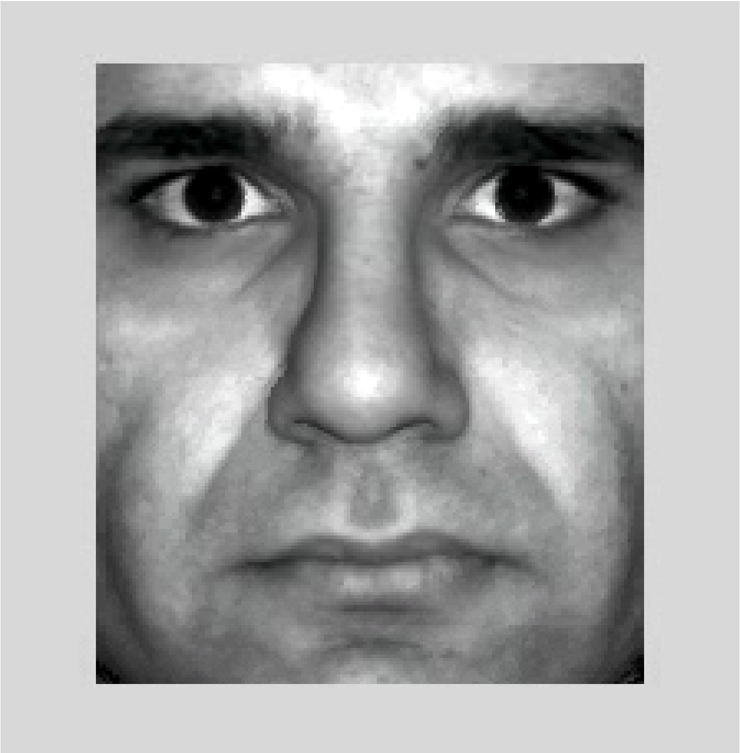
# Synthesizing Illumination Cone Images

Use estimated 3D linear subspace $\mathbf{B}(x,y)$ and surface $f(x,y)$ to generate synthetic images under variable lighting and viewpoint.

**Use synthetic images for recognition.**



Variable Lighting Movie



Variable Viewpoint Movie

[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

# Face Recognition Algorithm

## For fixed viewpoint:

- Use reconstructed surface $f(x,y)$ and linear subspace $\mathbf{B}(x,y)$ to synthesize extreme rays (images) of the illumination cones.
- Since a cone lies near a low-dimensional linear subspace, then approximate cone by a subspace.
- Classification is performed by computing distance to cone or distance to linear subspace.

## For variable viewpoint:

- Systematically sample the viewpoint space generating a linear subspace (cone approximation) per sample viewpoint.
- The union of the linear subspaces forms the face representation.
- Classification is performed by computing distance to the union of subspaces. This distance is equal to distance to nearest subspace in the representation.

[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

# The Yale Face Database B

The Geodesic Lighting Rig with 64 computer controlled strobes.

- 5,760 (single light source) images: 10 subjects, 576 images/subject
- 64 lighting directions (frontal to 90° off-center)
- 9 poses (frontal to 24° off-center)

http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html

[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

# Face Recognition:
## Experimental Protocol

Subset 1: 0-12°

Subset 2: 12-25°

Subset 3: 25-50°

Subset 4: 50-77°

Test images divided into 4 subsets with increasing extremity in illumination.

**Training**: Train on 70 images of Subset 1 (7/person) where lighting is within 12° of the camera's optical axis and viewpoint is frontal.

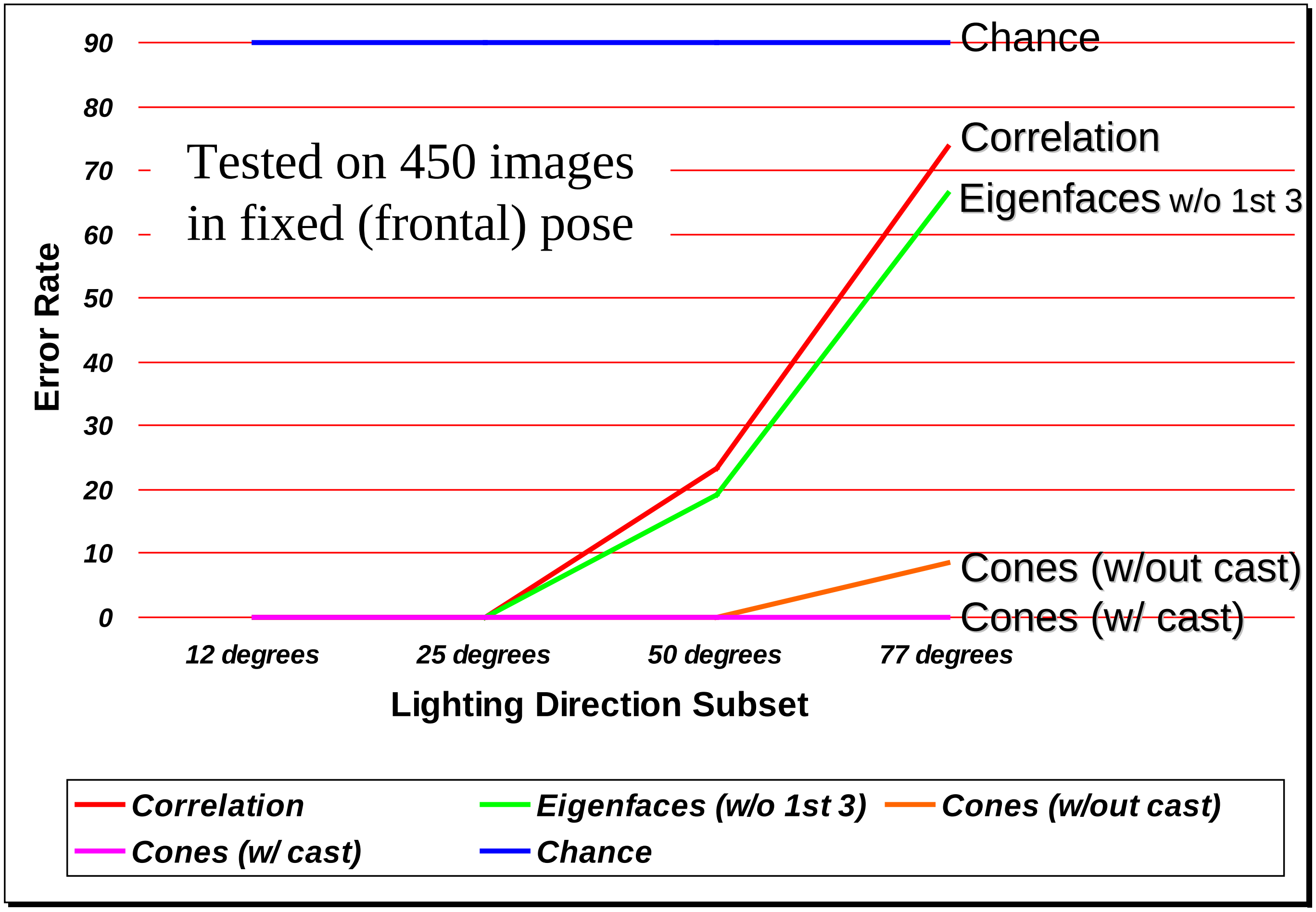
**Testing**: Test on 4050 images (405/person): 10 persons X 45 lighting directions ranging from frontal to 77° X 9 viewpoints ranging from frontal to 24°.

**Classification Method**: Nearest Neighbor (i.e., assign to test image the identity of the nearest face representation.)

[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

# Face Recognition Results:
## Variable Lighting

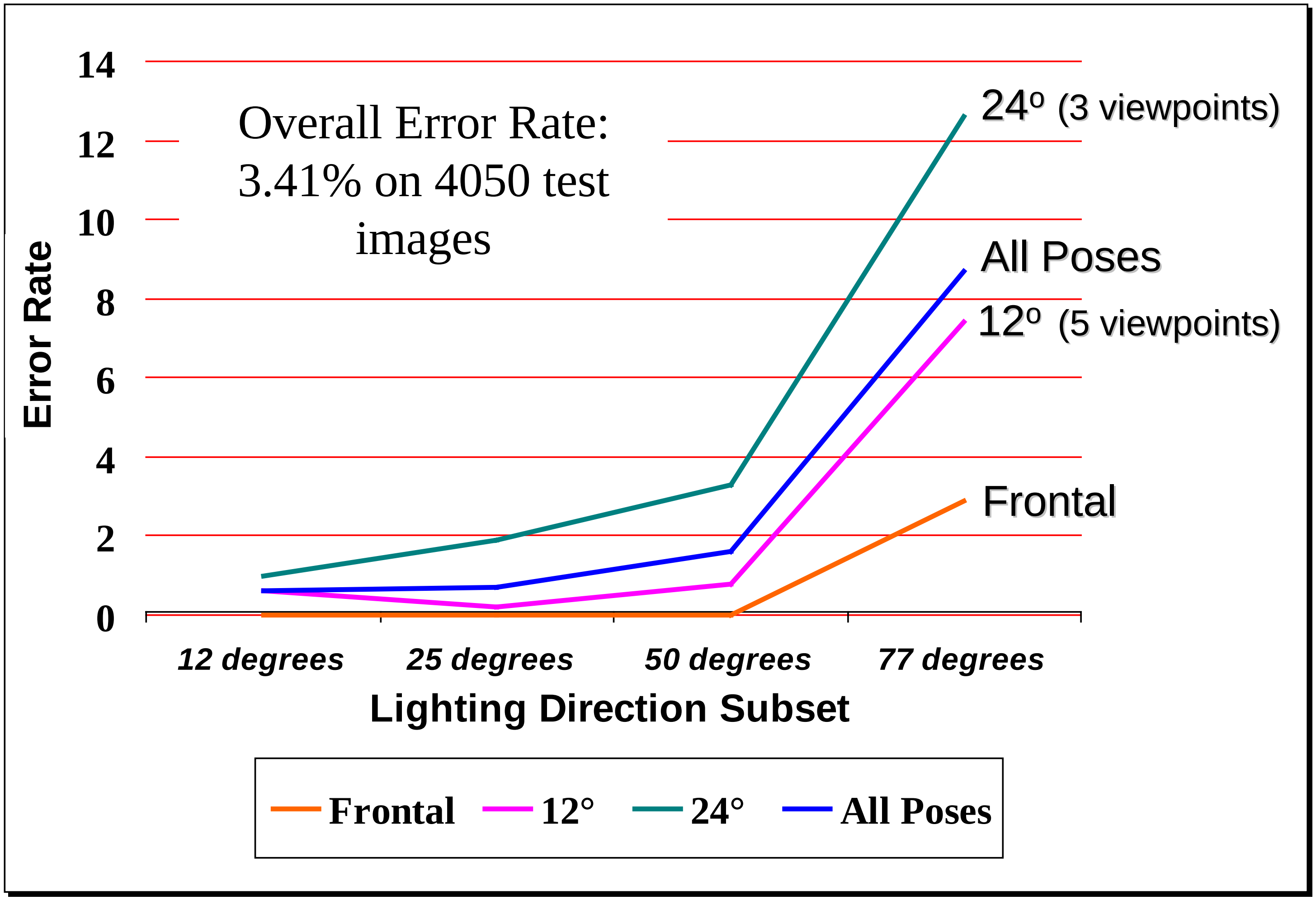
Comparison of Illumination Cones (with and without cast shadows) against Normalized Cross-Correlation and Eigenfaces (w/out the 1st 3 principal components).




[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

# Face Recognition Results:
## Variable Lighting *and* Viewpoint

Test of the union of linear subspaces with all 9 viewpoints




[Georghiades, Belhumeur, Kriegman, PAMI, June 2001]

# Dealing with Pose

- "View-Based"
  - multiple view-tuned models
- Use components
  - less sensitive
- Use 3D models

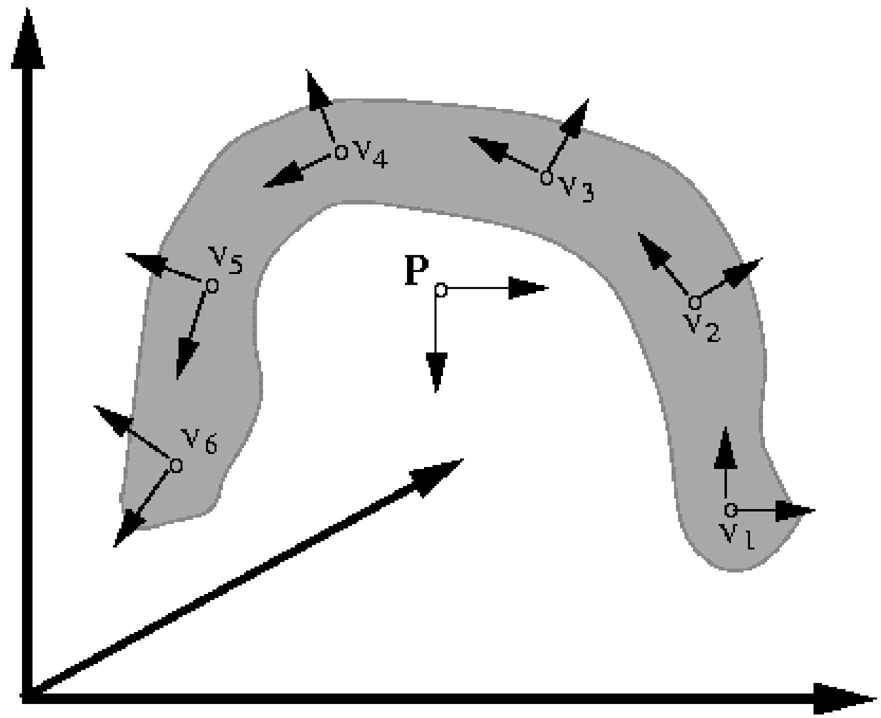
# 2D View-Based Models

Form specific ("view-tuned") models for each view/pose (orientation)

Note: memory-based technique!

# 2D View-Based Models

Compute an Eigenspace
for each view (column)




Pentland, Moghaddam, Starner, "View-Based and Modular Eigenspaces for Face Recognition"
IEEE Conf. on Computer Vision & Pattern Recognition, **CVPR'94**. Seattle, WA, July 1994.

# Virtual Views via Morphing

Beymer & Poggio (1995)

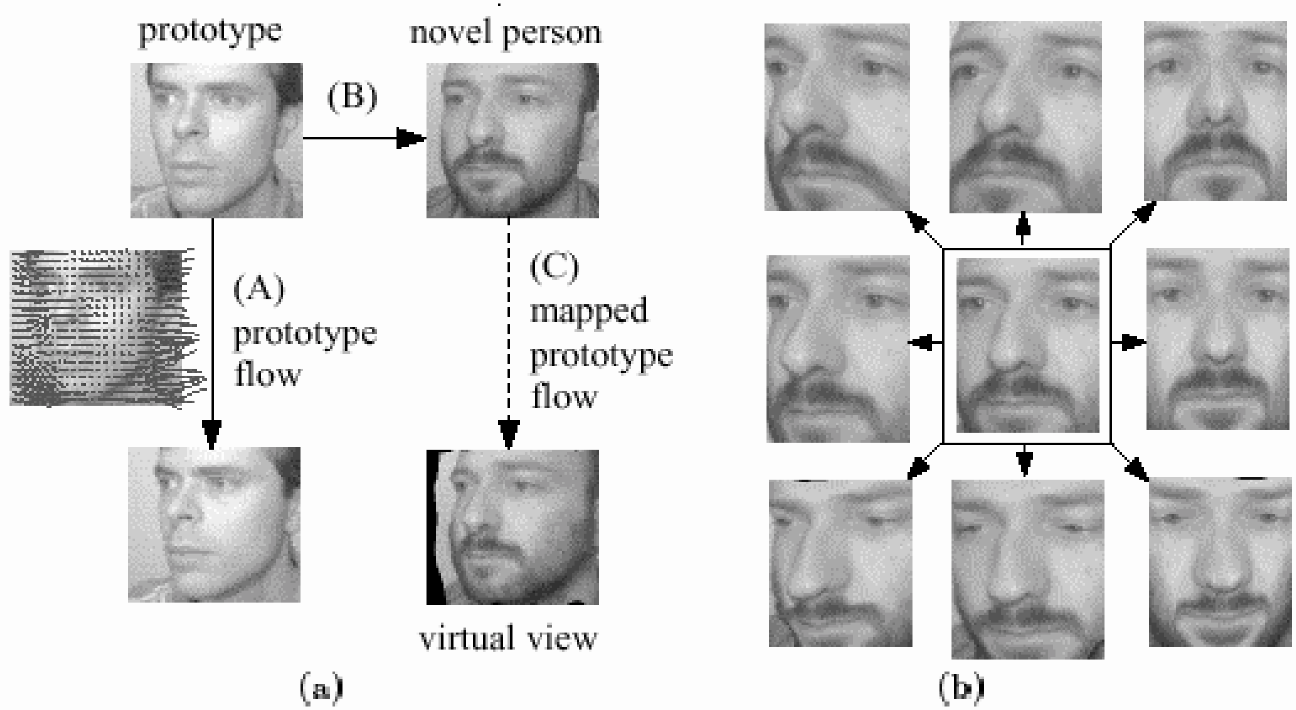
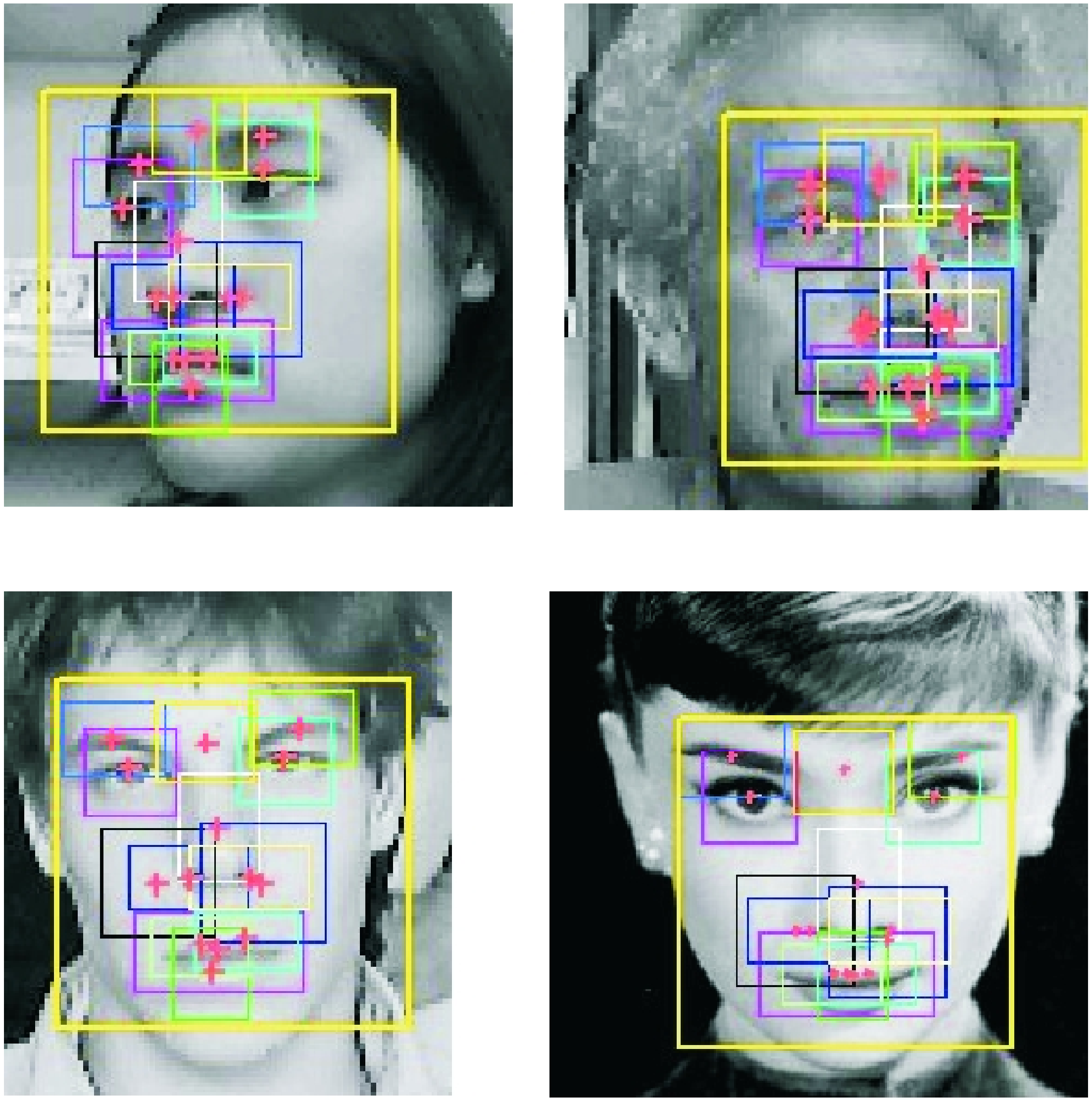



**Figure 7.** **(a):** In parallel deformation, (A) a 2D deformation representing a transformation is measured by finding correspondence among prototype images. In this example, the transformation is rotation and optical flow was used to find a dense set of correspondences. Next, in (B), the flow is mapped onto the novel face, and (C) the novel face is 2D warped to a "virtual" view. From Beymer and Poggio (1995b) . **(b):** A real view (center) surrounded by virtual views derived from it using parallel deformation. From Beymer and Poggio (1995b) .

# Component-Based SVM Face Recognition

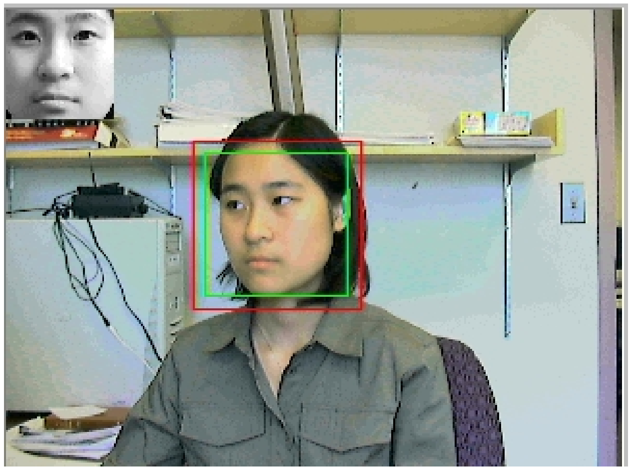


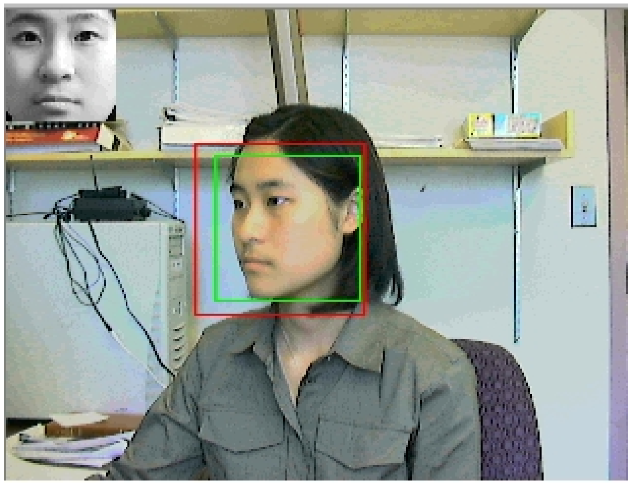
Hierarchical SVMs based on automatically learned components (14 here) yield a robust trainable object detection system tolerant to significant rotations with a ROC performance better than any of the existing systems (on the test data we used)

Heisele, Poggio, 2000

# Component-Based SVM Face Recognition

- Recognition of rotated faces up to about 45º
- Robust against changes in illumination and background
- Frame rate of 15 Hz





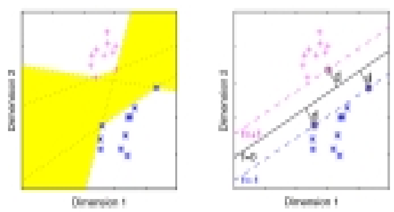
Ho, Heisele, Poggio et al., 2000

# Face Recognition with Support Vector Machines: Global vs. Component-based Approach

**Bernd Heisele†, Purdy P. Ho‡, Tomaso Poggio**
Massachusetts Institute of Technology
Center for Biological and Computational Learning
† Honda R&D Americas Inc., ‡ Hewlett-Packard

## 1 Support Vector Machine

### 1.1 Principle



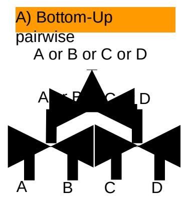

The SVM is a **maximum margin** classifier. It performs pattern recognition between two classes by finding a decision surface that has maximum distance to the closest points in the training set which are termed support vectors.

Decision function:

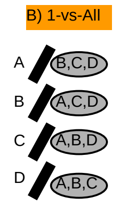
$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{N} y_i\, \alpha_i\, \mathbf{x}_i \cdot \mathbf{x} - b\right)$$

$y_i$: Label $\{-1,1\}$, $\alpha_i$: Lagrange multiplier
$\mathbf{x}_i$: Support vector, $\mathbf{x}$: Feature vector
$N$: Nb. of support vectors, $b$: Constant

### 1.2 Multi-class Classification with SVMs




Training: $L$ ($L$-1) / 2
Run-time: $L$-1




Training: $L$
Run-time: $L$

A) **Pairwise approach**:
Each SVM separates a pair of classes. The pairwise classifiers are arranged in trees.

B) **1-vs-all approach**:
Each of the SVMs separates a single class from all remaining classes. The decision is based on the maximum distance to the hyperplane.

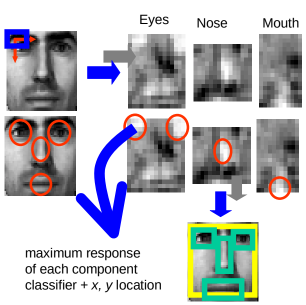
We opted for the 1-vs-all strategy where the number of SVMs is linear with the number of classes $L$.

## 2 Component-based Face Detection

### 2.1 System Overview

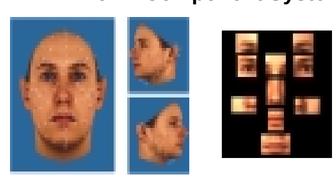
**1st Level: Component classifiers**

**2nd Level: Geometrical classifiers**

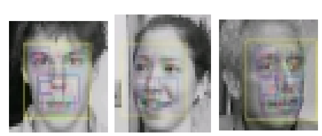



On the first level, windows of the size of the components (solid lined boxes) are shifted over the face image and classified by the component classifiers. On the second level, the maximum outputs of the component classifiers within predefined search regions (dotted lined boxes) and the positions of the detected components are fed into the geometrical configuration classifier.

### 2.2 The 14 Component System



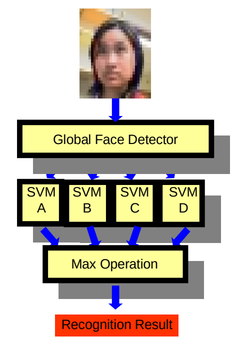
The shapes and positions of the components have been automatically learned from synthetic face images in order to provide maximum discrimination between face and non-face images. Overall 14 components have been learned.



Examples of component-based detection applied to real face images.

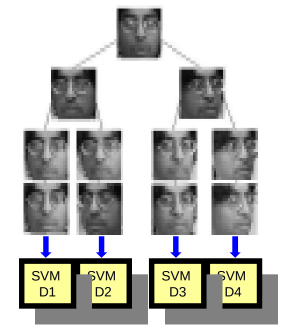
## 3 Face Identification

### 3.1 Global Approach

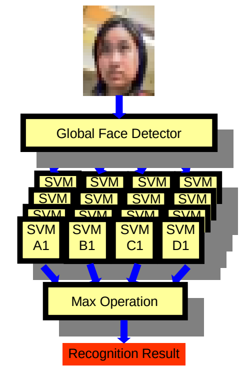



A global face detector localizes and extracts the face. The pixel values of the face pattern are combined into a feature vector which is then fed into $L$ linear SVMs. The classification is based on the maximum distance to the hyperplane.

### 3.2 Global Approach with Clustering

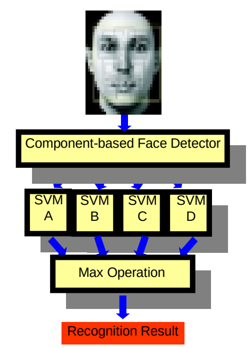



The training data is split into $N$ view-dependent clusters by a divisive clustering algorithm. A linear SVM is trained on each cluster.



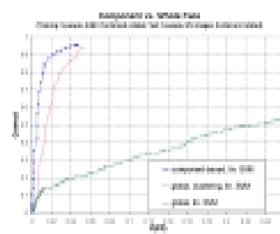

A global face detector extracts the face. The pixel values of the face pattern are combined into a feature vector which is then fed into $N \bullet L$ linear SVMs. The classification is based on the maximum distance to the hyperplane.

### 3.2 Component-based Approach

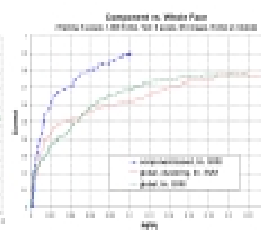



A component-based face detector localizes the face and extracts the components. The components are normalized in size and their pixel values are into a feature vector which is then fed into $L$ linear SVMs. The classification is based on the maximum distance to the hyperplane.

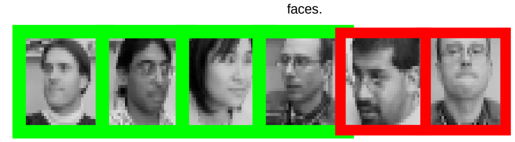
### 3.3 Results



ROC curves when trained and tested on frontal and rotated faces.



ROC curves when trained on frontal faces and tested on frontal and rotated faces.



Examples of component-based face recognition. The people in the first four images (green boxes) could be identified by the system. The last two images (red box) show misclassifications due to strong rotation and facial expression.

# Course Outline

- Brief History
- Introduction to Key Problems
- Face Perception in Humans
- Automatic Face Recognition
  - face detection
  - neural network methods
  - features vs. templates
  - subspace methods
  - FERET test protocol
  - lighting/pose techniques
  - 2D/3D models
- Future Directions

# 2D Shape + Texture

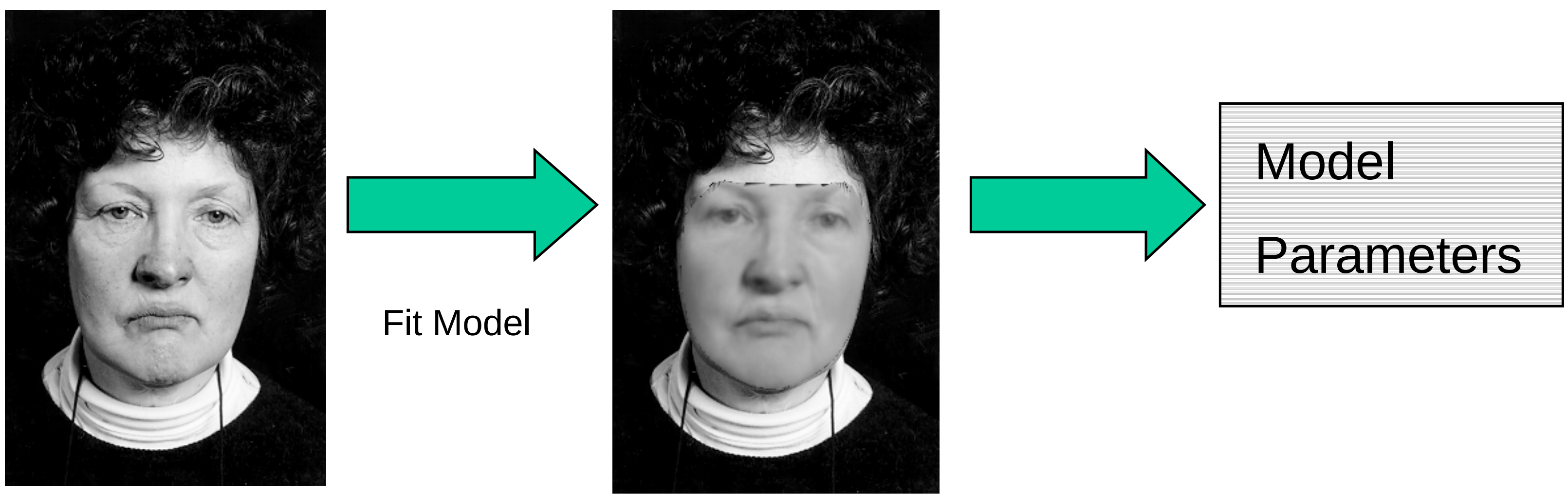
- Beier & Neely (SIGGRAPH 92)
- Craw et al (F&G 1995)
- Bichsel (F&G 1995-6)
  - 2.5D models & "morphing"
- Lanitis, Cootes & Taylor (1997)
  - "Active Appearance Models"


Thomas Vetter

# Active Appearance Models

Cootes *et.al.* University of Manchester, UK

*Interpret images using generative models of appearance – ‘explain’ the image*



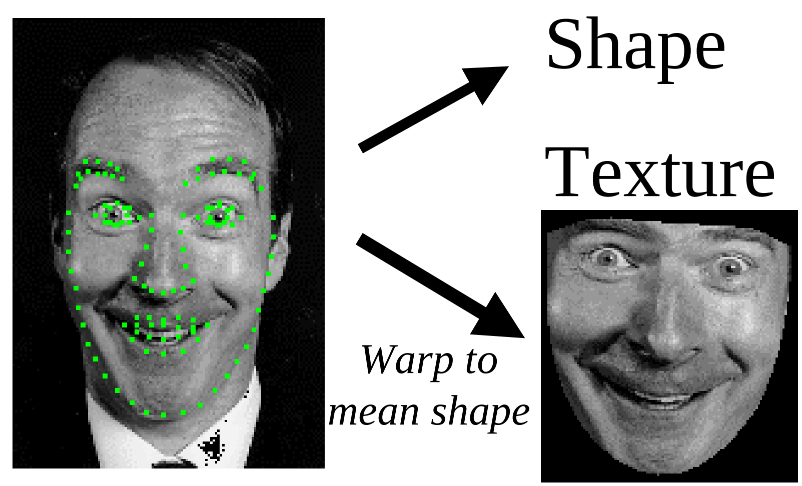
# Appearance Models

Cootes *et.al.* University of Manchester, UK

From a training set learn model of shape and texture variation

**Shape:** $\mathbf{x} = \mathbf{x}_{mean} + \mathbf{Q}_s\mathbf{c}$

**Texture:** $\mathbf{g} = \mathbf{g}_{mean} + \mathbf{Q}_g\mathbf{c}$




Varying one parameter of model:

# Face Interpretation with AAMs

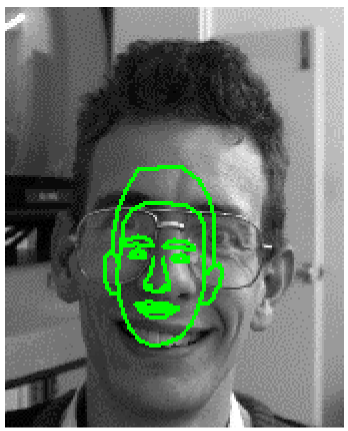
Cootes *et.al*. University of Manchester, UK

- Appearance models encode face in a small number of parameters
- Match to new image using Active Appearance Model (fast iterative algorithm)
- Model parameters can then estimate
  - Identity
  - Head pose
  - Expression etc
- Can also use model to manipulate face images (change expression, lighting etc)

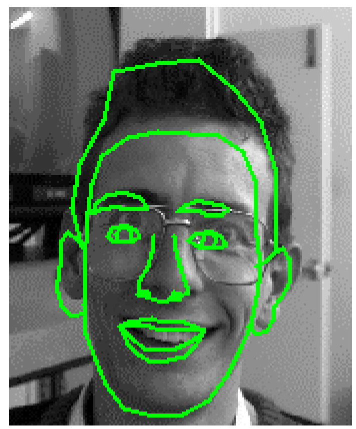
# Shape Matching with AAMs

Cootes *et.al.* University of Manchester, UK

initialized



5 iterations

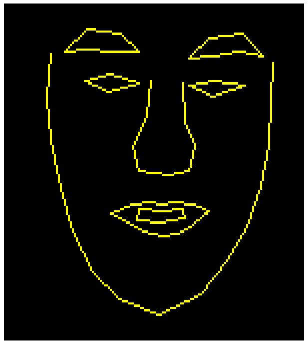


converged

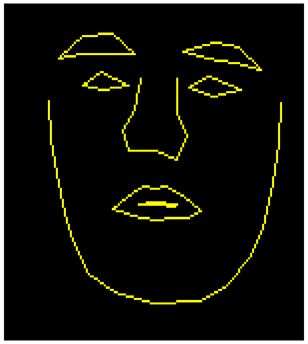
# Shape Deformations with AAMs

Cootes *et.al.* University of Manchester, UK

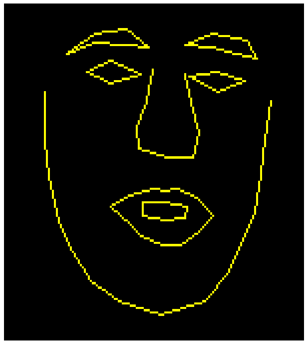
Mode 1



Mode 2

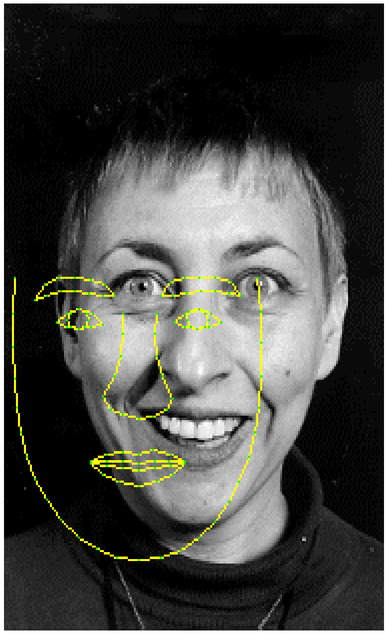


Mode 3



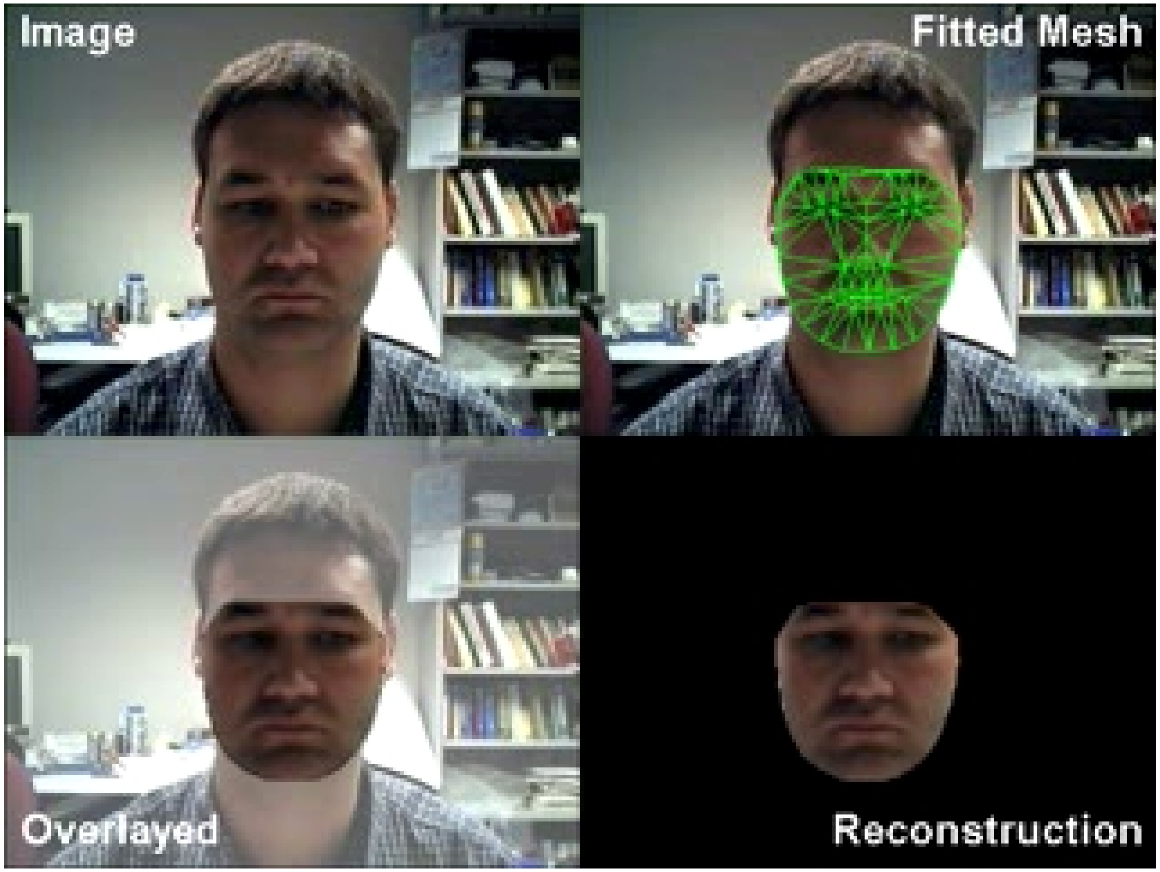
# Face Interpretation with AAMs

Cootes *et.al.* University of Manchester, UK



# Flexible Appearance Models

Baker & Mathews (CVPR'01)

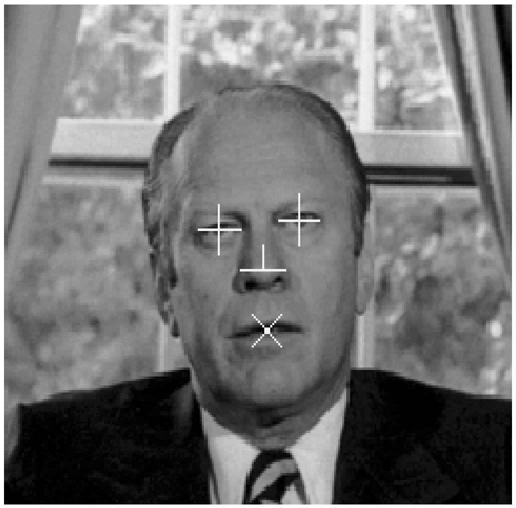


# Building 3D Models

# Jebara (1995)

"3D Pose Estimation and Normalization for Face Recognition"

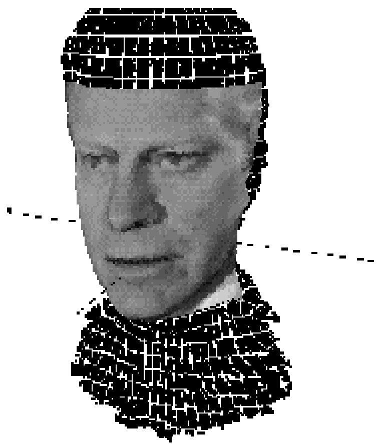
MS Thesis, McGill University, 1995

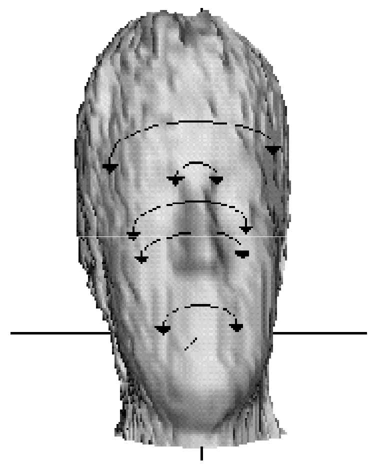


(a) (b) (c)

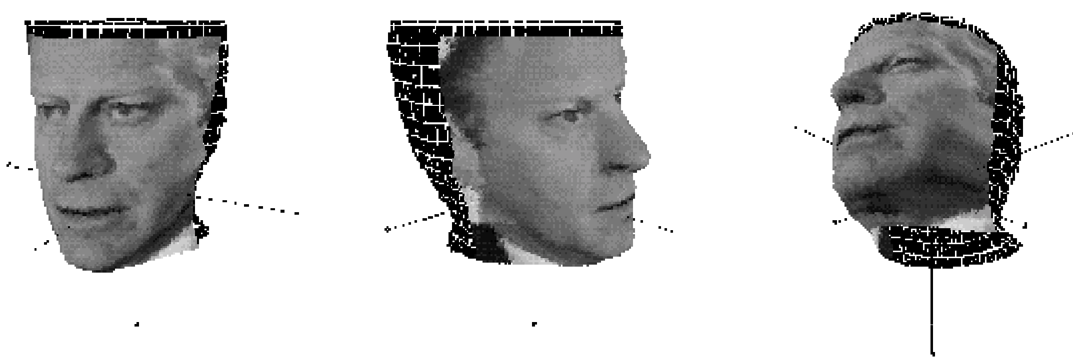
# Jebara (1995)

“3D Pose Estimation and Normalization for Face Recognition”





virtual “mugshot”



(a) (b) (c)

# Jebara (1995)

"3D Pose Estimation and Normalization for Face Recognition"

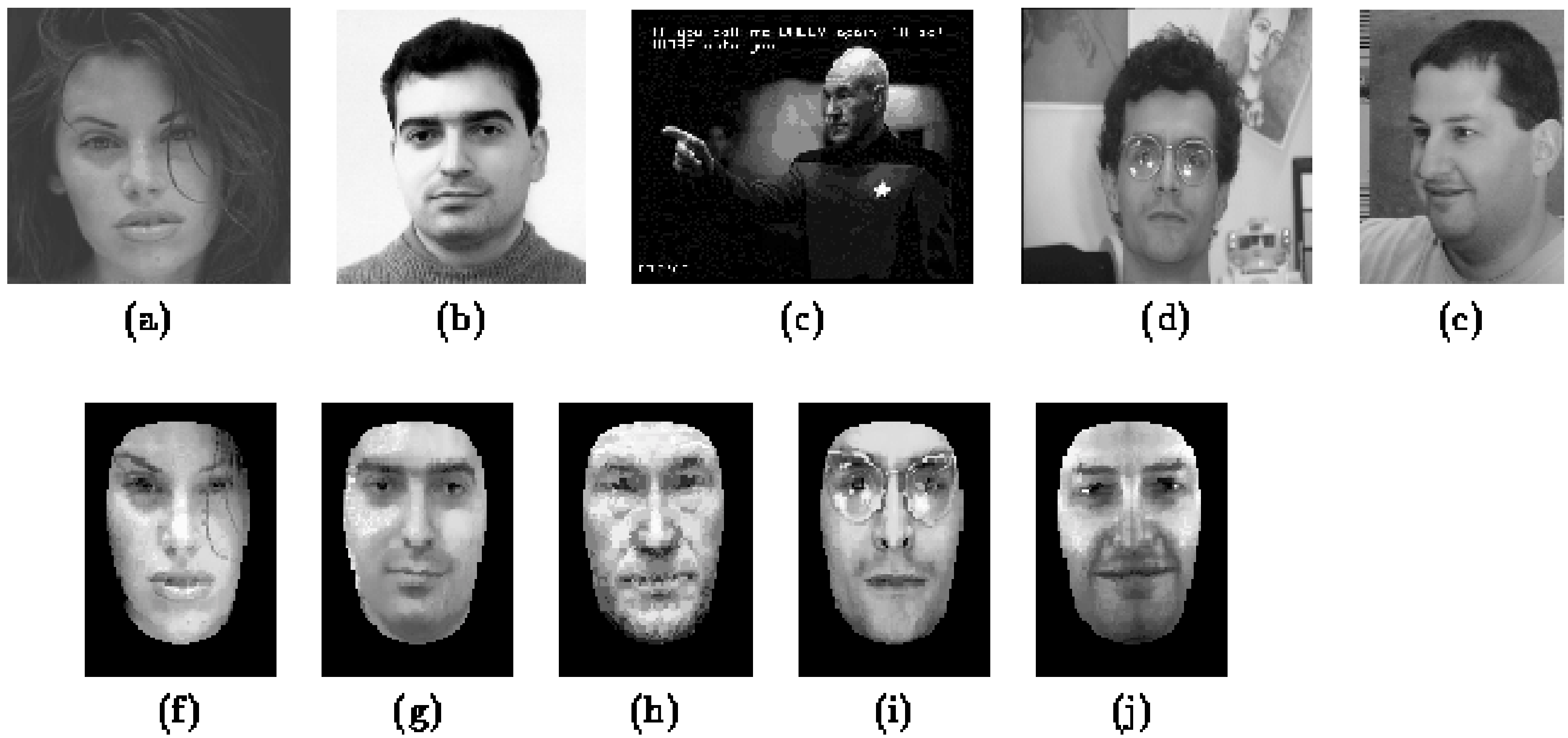


Figure 4.5: A gallery of face normalization results. (a) (b) (c) (d) (e) The original faces. (f) (g) (h) (i) (j) The corresponding synthesized mug-shots.

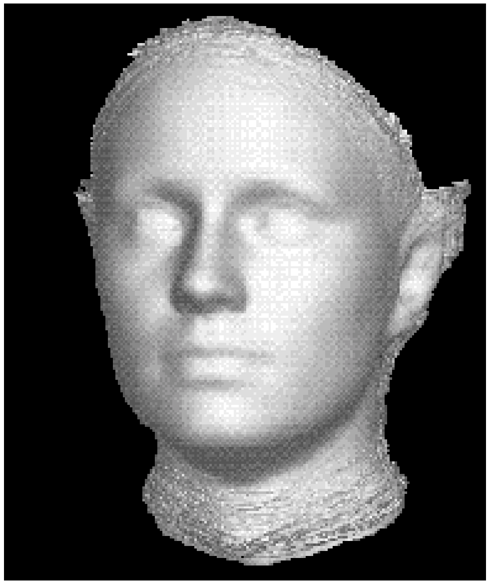
# “Eigenheads”

Atick, Griffin, Redlich (1996)

Model prior on 3D head shape and use it to for *shape-from-shading*

$$r = f(\theta, z)$$

$$r = f_0(\theta, z) + \sum_i \alpha_i \psi_i(\theta, z)$$



$f_0(\theta, z)$

Assume face is a Lambertian surface, with constant albedo and no self-shadowing:

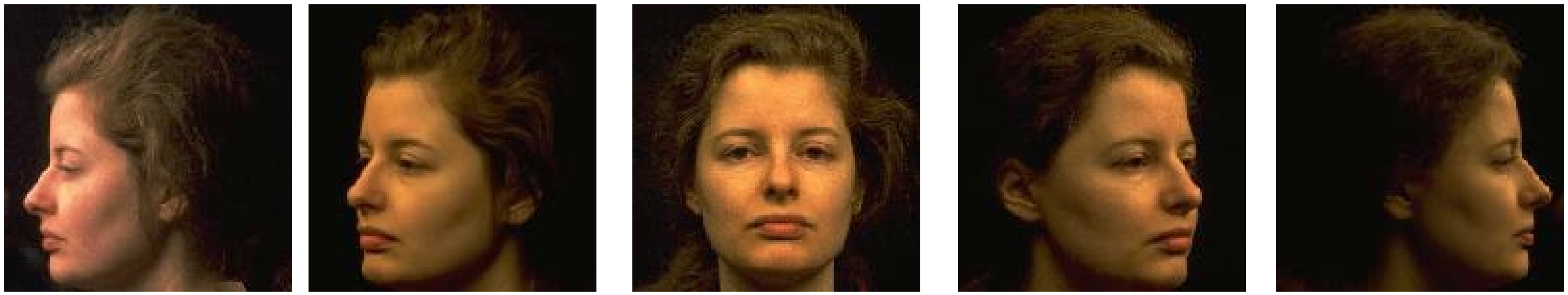
$$I(r\sin(\theta), z) \approx \vec{L} \cdot \vec{n}(\theta, z)$$

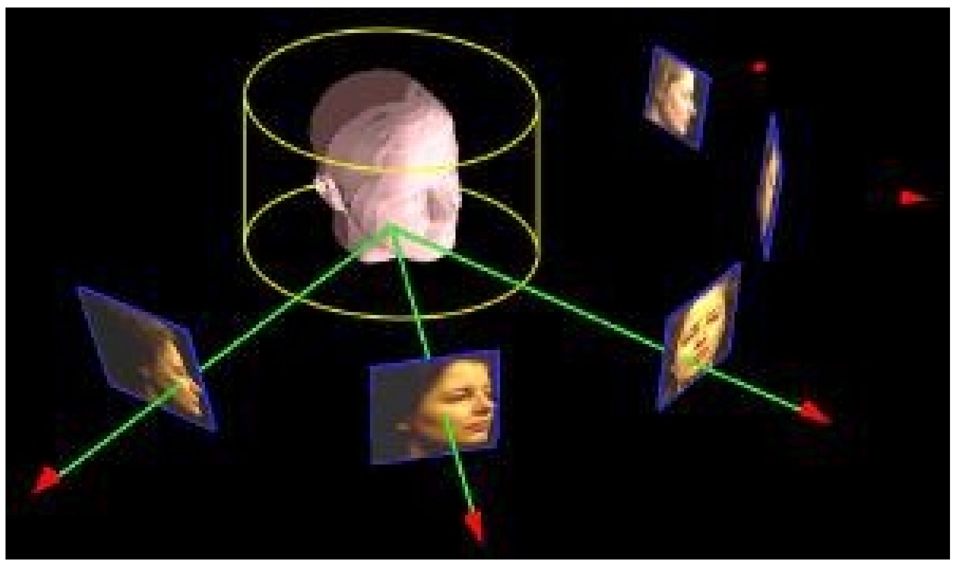
Estimate 3D shape (eigenhead expansion coefficients) and the light source from a single 2D image using prior shape construct (eigenhead surface normals), by minimizing:

$$J(\alpha, \vec{L}) = \iint (I - \vec{L} \cdot \vec{n}_\alpha)^2 \, d\theta dz$$

# 3D Models from Multiple Views

## Pighin et al (1998)



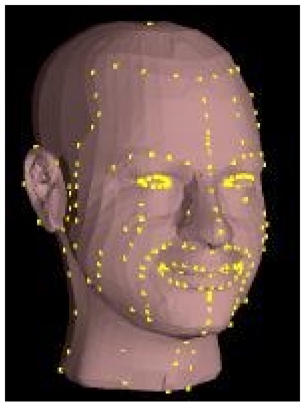


Frederic Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, and David Salesin. Synthesizing Realistic Facial Expressions from Photographs. Proceedings of **SIGGRAPH 98**, in Computer Graphics Proceedings, Annual Conference Series, 1998.

# 3D Models from Multiple Views

## Pighin et al (1998)

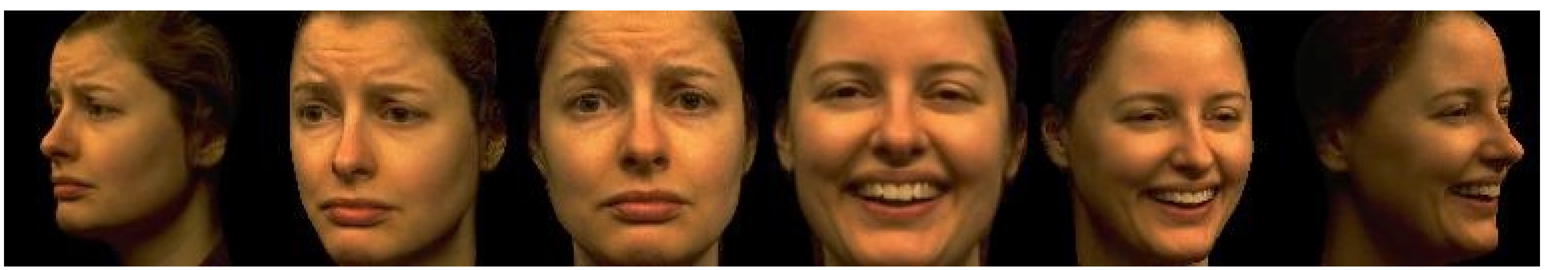
3D shape



+

2D texture

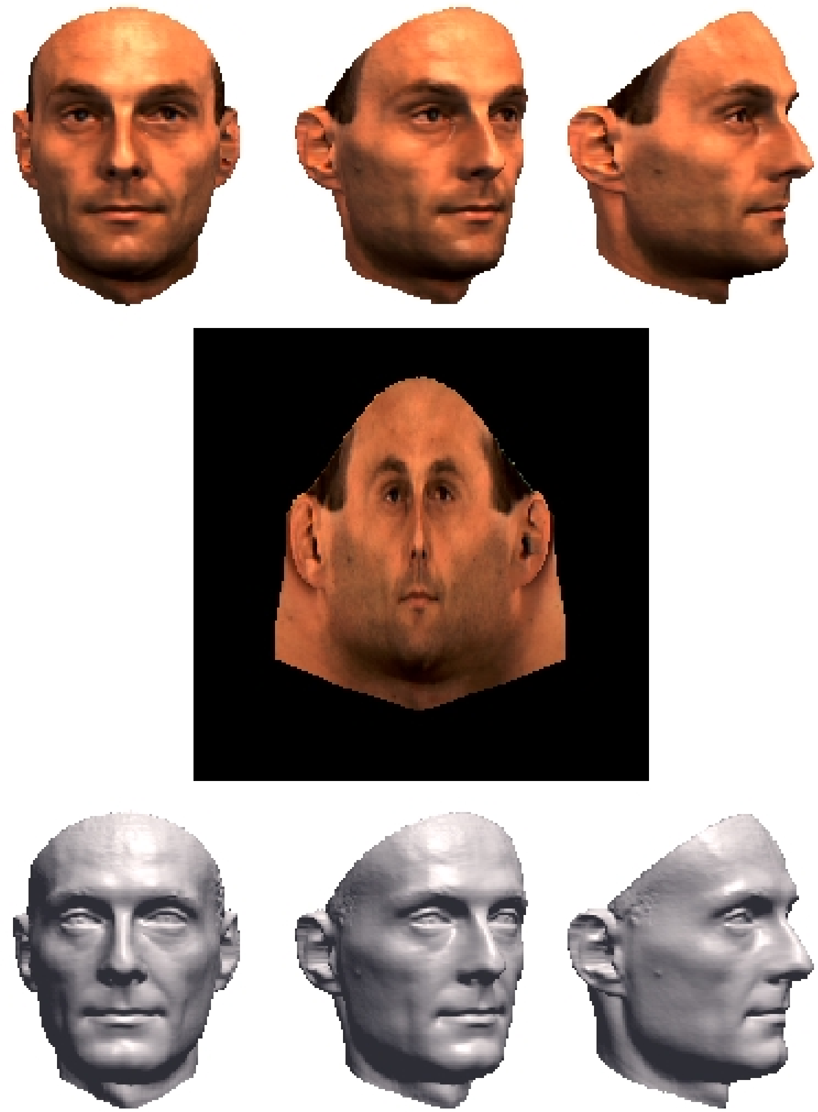
synthetic novel views



Frederic Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, and David Salesin. Synthesizing Realistic Facial Expressions from Photographs. Proceedings of **SIGGRAPH 98**, in Computer Graphics Proceedings, Annual Conference Series, 1998.

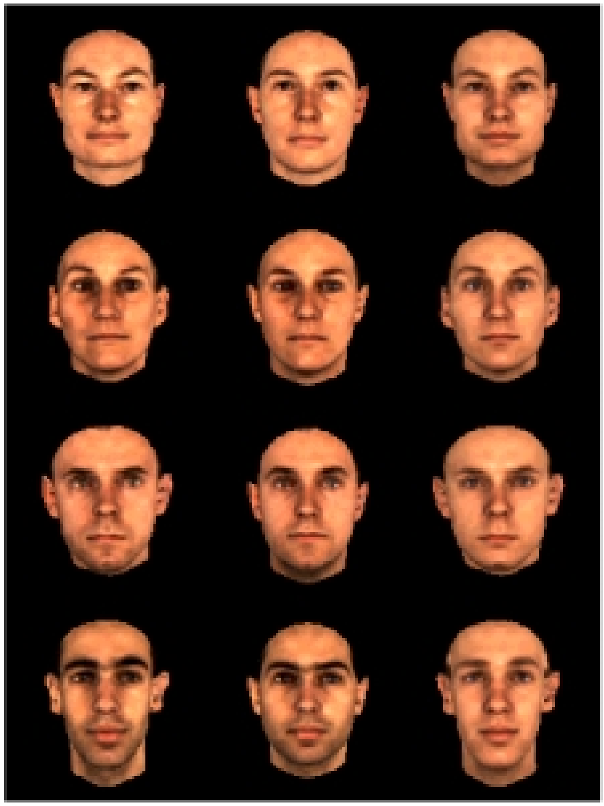
# 3D Laser Scans

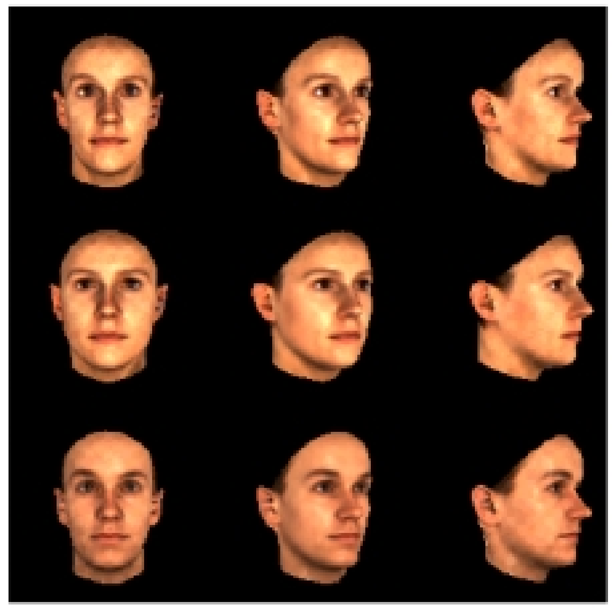
Combines 3D geometry
with skin texture map.
Use compute graphics
to render views under
arbitrary lighting and pose



www.cyberware.com

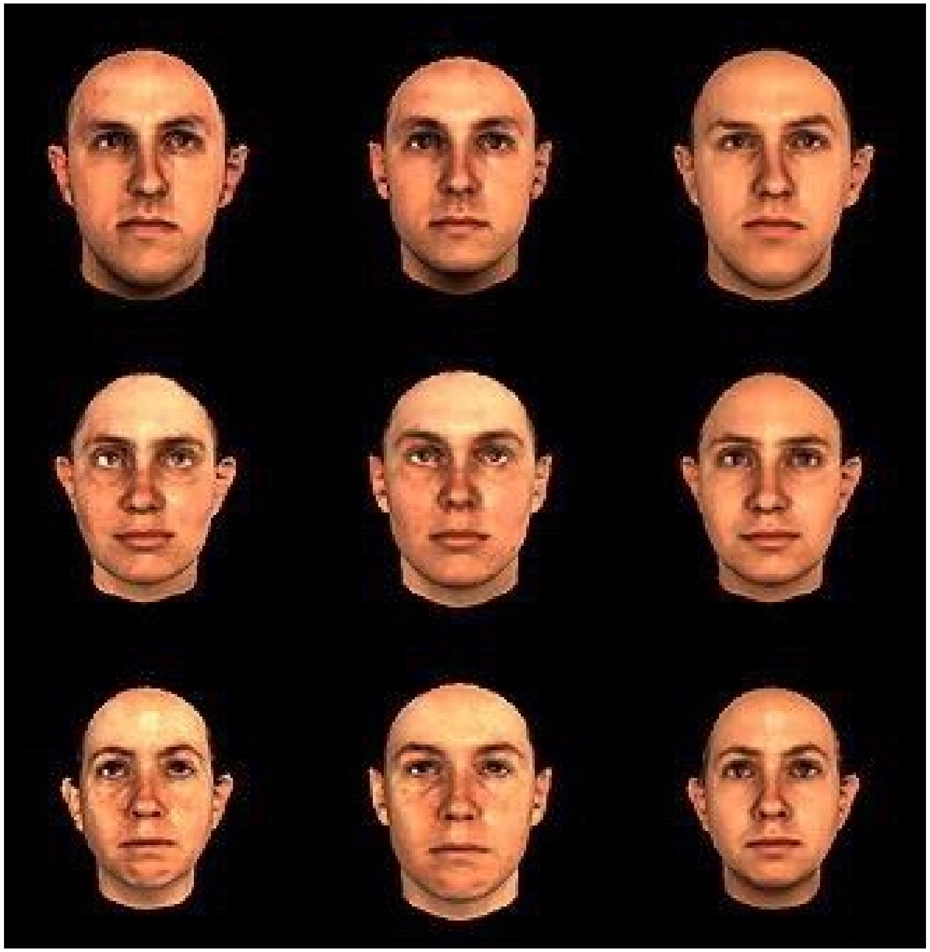
# 3D Shape + Texture





# Shape & Texture Norms
## O’Toole et al (2001)

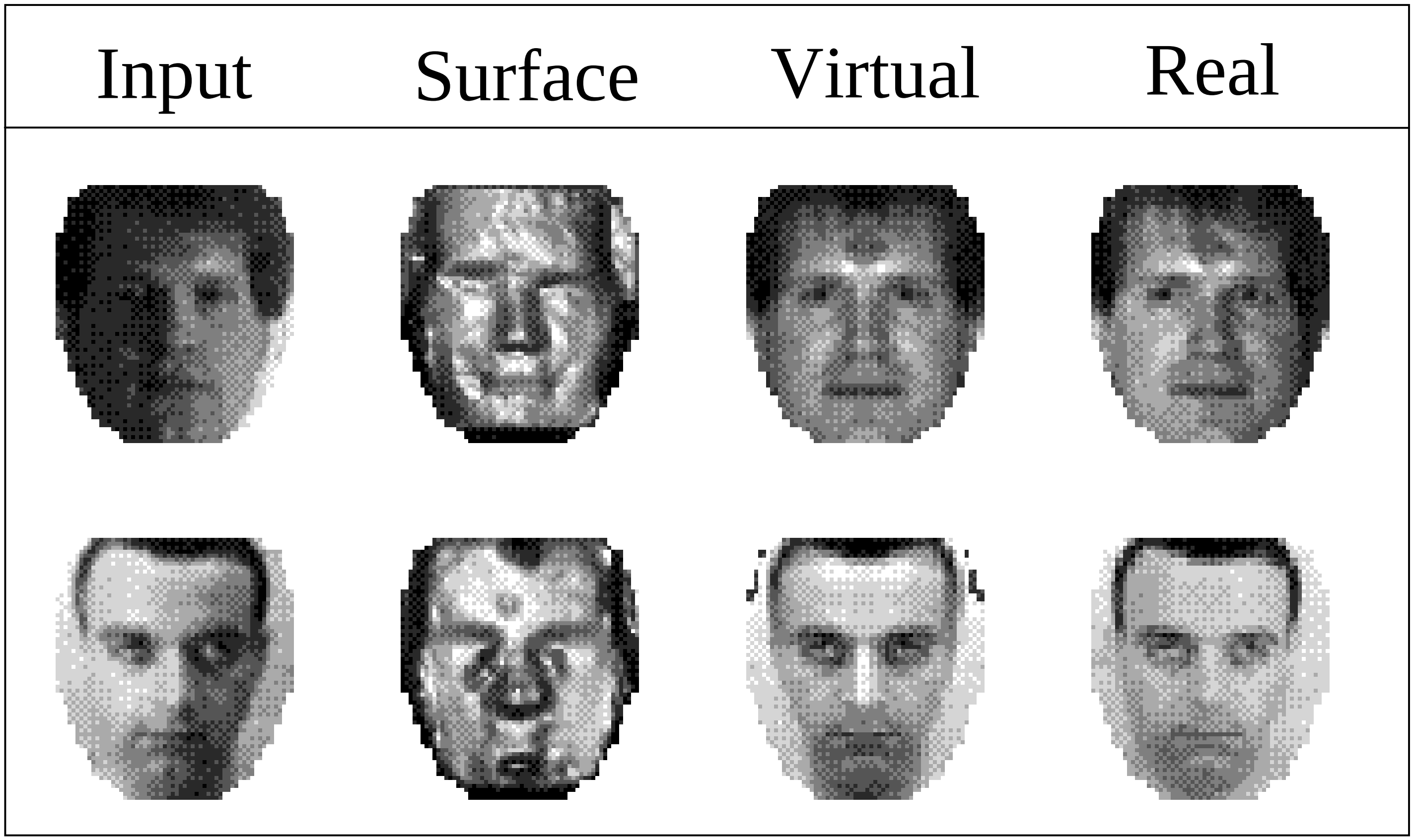
Original Shape-norm Texture-norm



# Advanced Modeling *

# 3D Models from Shading (SFS)

Zhao & Chellappa (2000)



# 3D Morphable Models

Brand (CVPR'01)

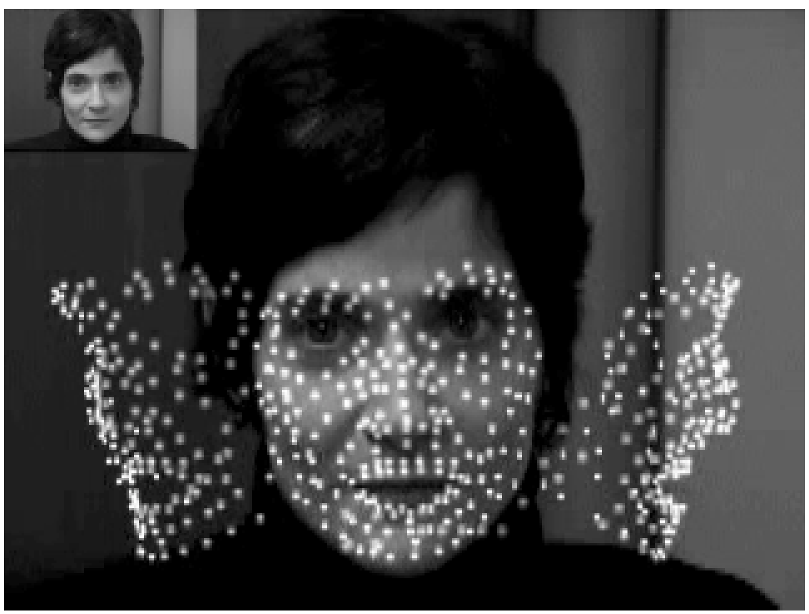
INPUT: Raw Video + 3D Morphable Model

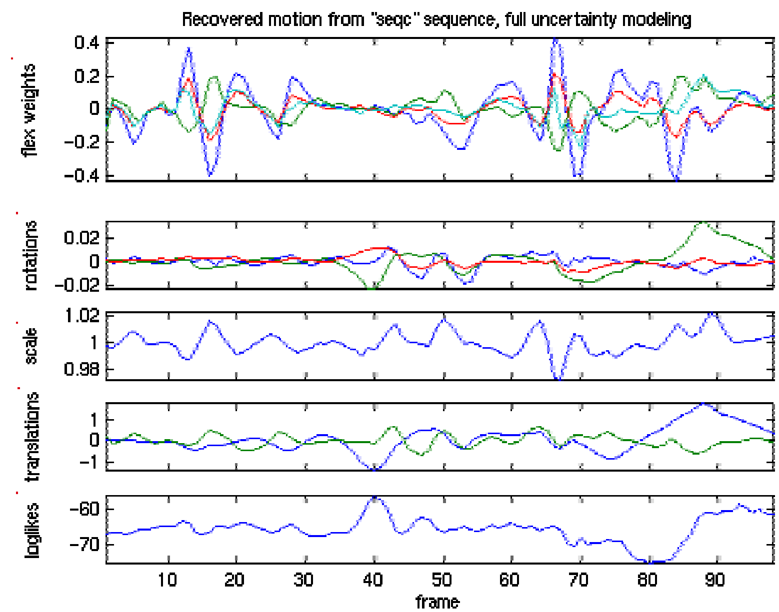
Marker-less Tracking

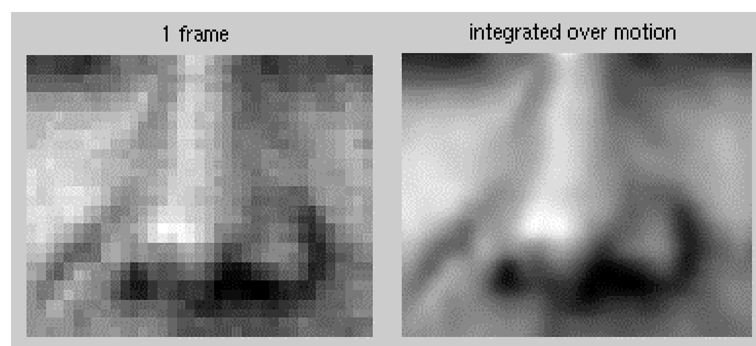
# 3D Morphable Models

Brand (CVPR'01)

OUTPUT : 3D Motions
+ Shape changes
+ Super-Res Texture





# 3D Morphable Models

Brand (CVPR'01)

3D animation models & control parameters from casual video

Input video

Video re-write with extracted 3D shape

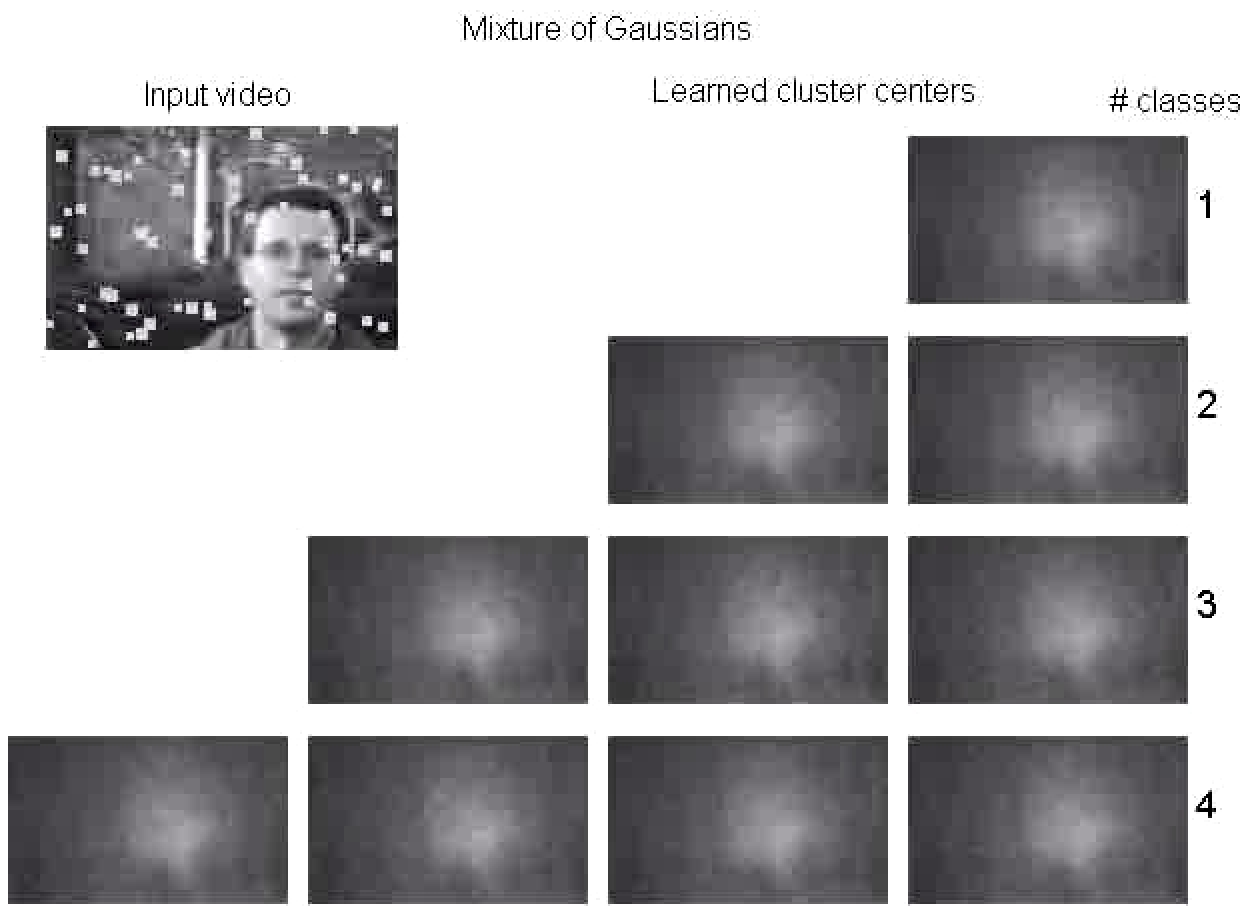
# Transformation-Invariant Clustering

Frey & Jojic, CVPR/NIPS/PAMI 1999-2001, *www.psi.toronto.edu*

Goal: Unsupervised clustering of images/videos that contain objects that are transformed (eg, translated)

Standard clustering fails!

Model size




1 class

2 classes

3 classes

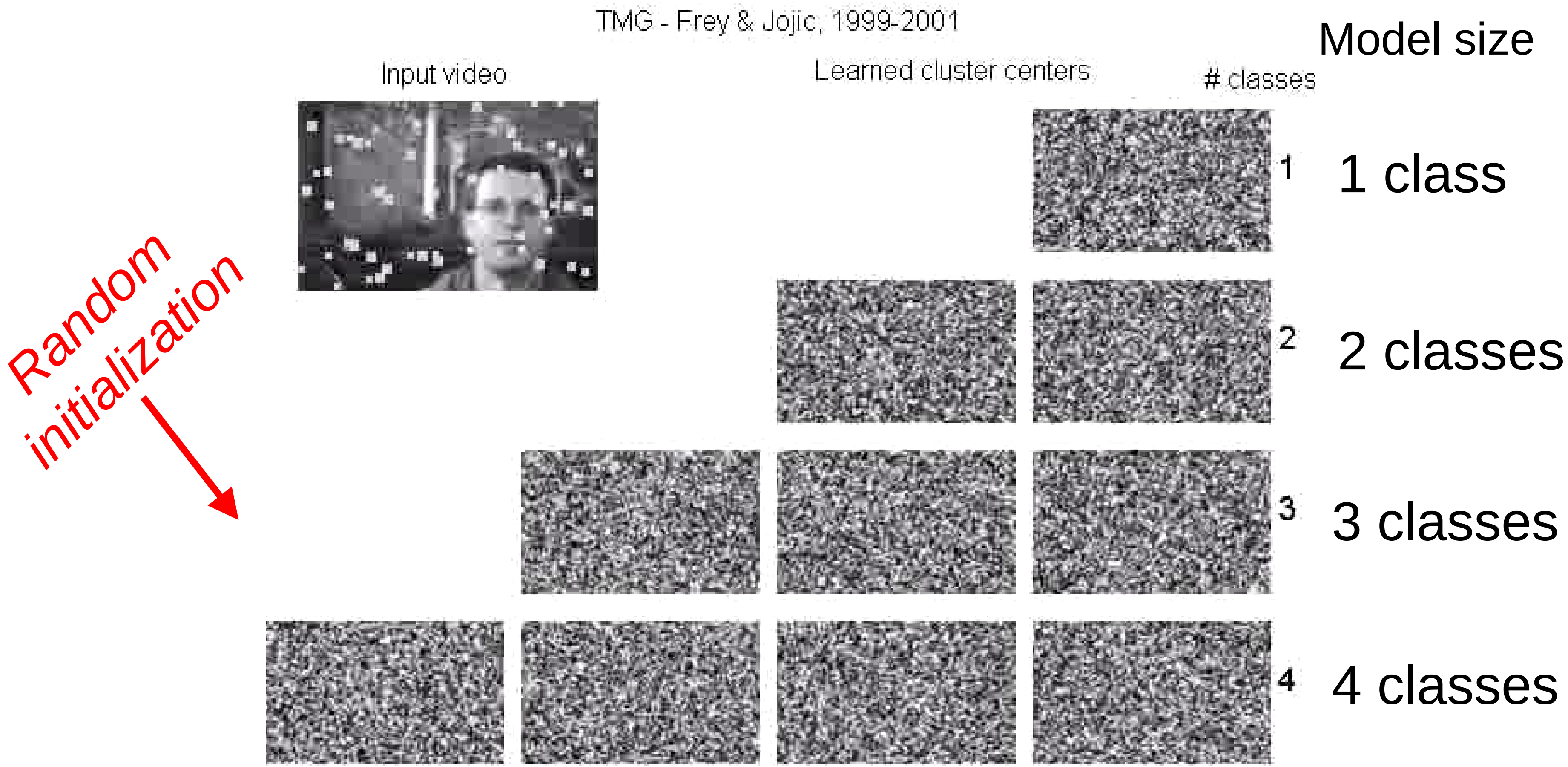
4 classes

# Transformation-Invariant Clustering

Frey & Jojic, CVPR/NIPS/PAMI 1999-2001, *www.psi.toronto.edu*

TMG: Transformation-invariant Mixture of Gaussians

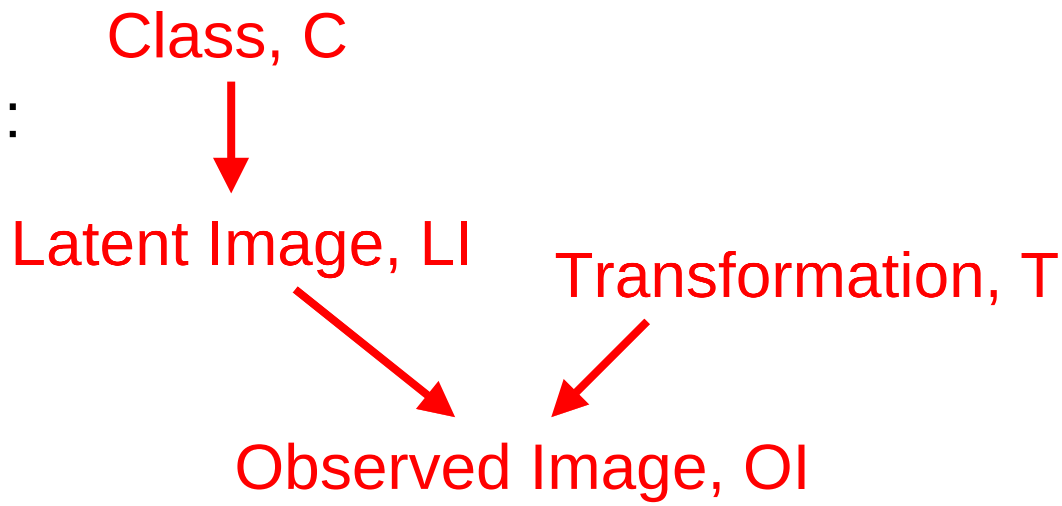
- Works for video and unordered images
- User specifies *input video* and *number of classes*



# Transformation-Invariant Clustering

Frey & Jojic, CVPR/NIPS/PAMI 1999-2001, *www.psi.toronto.edu*

Generative model:




$P(C, LI, T, OI \mid \theta) = P(C)\, P(LI \mid C)\, P(T)\, P(OI \mid LI, T)$

$\theta$ = model parameters: mean and cov of each class, prob of class, prob of transformation, cov of observation noise

Inference: Compute $P(C, LI, T \mid OI, \theta)$

Learning, E-step: Compute suff stats using $P(C,LI,T \mid OI, \theta)$

Learning, M-step: Modify $\theta$

# Transformation-Invariant Clustering

Frey & Jojic, CVPR/NIPS/PAMI 1999-2001, *www.psi.toronto.edu*

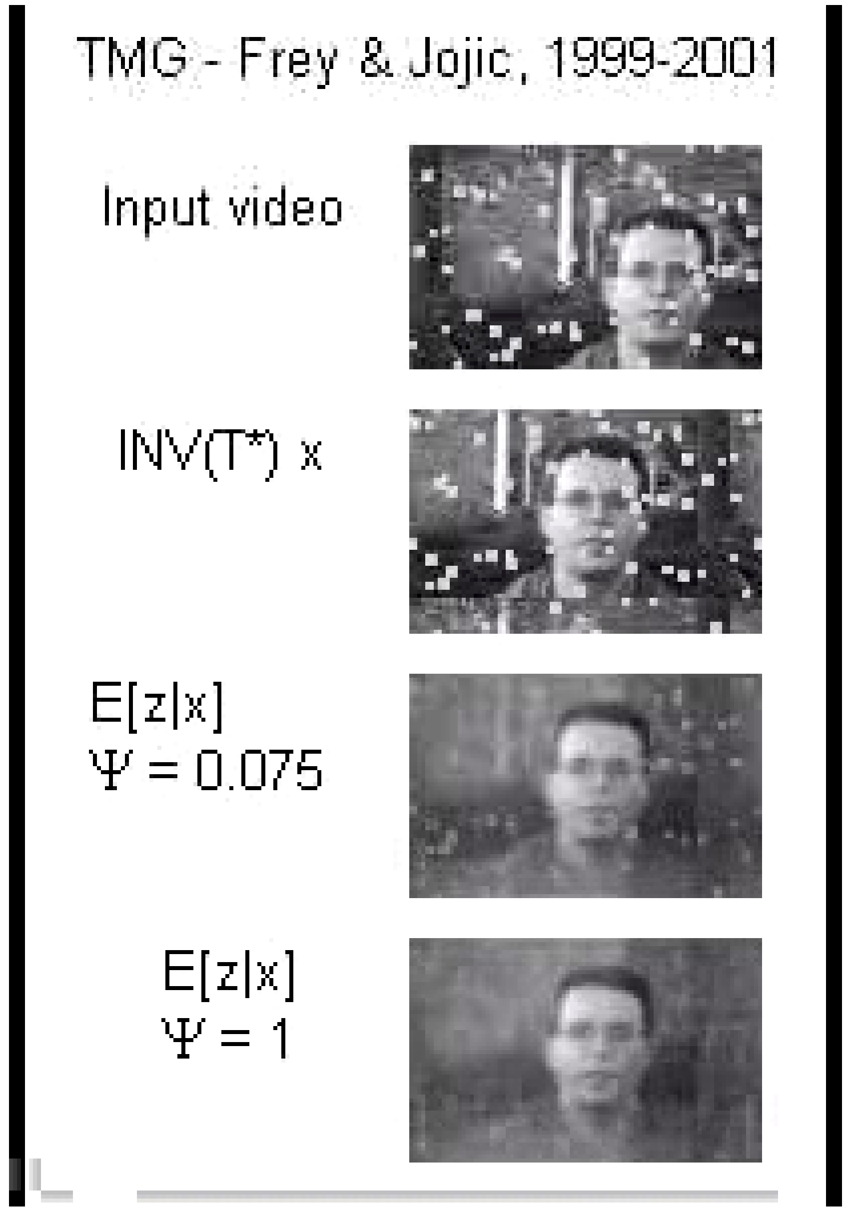
After training, TMG can be used to process video

P(C,LI,T|OI,θ) is used to

stabilize the input

remove snow from face

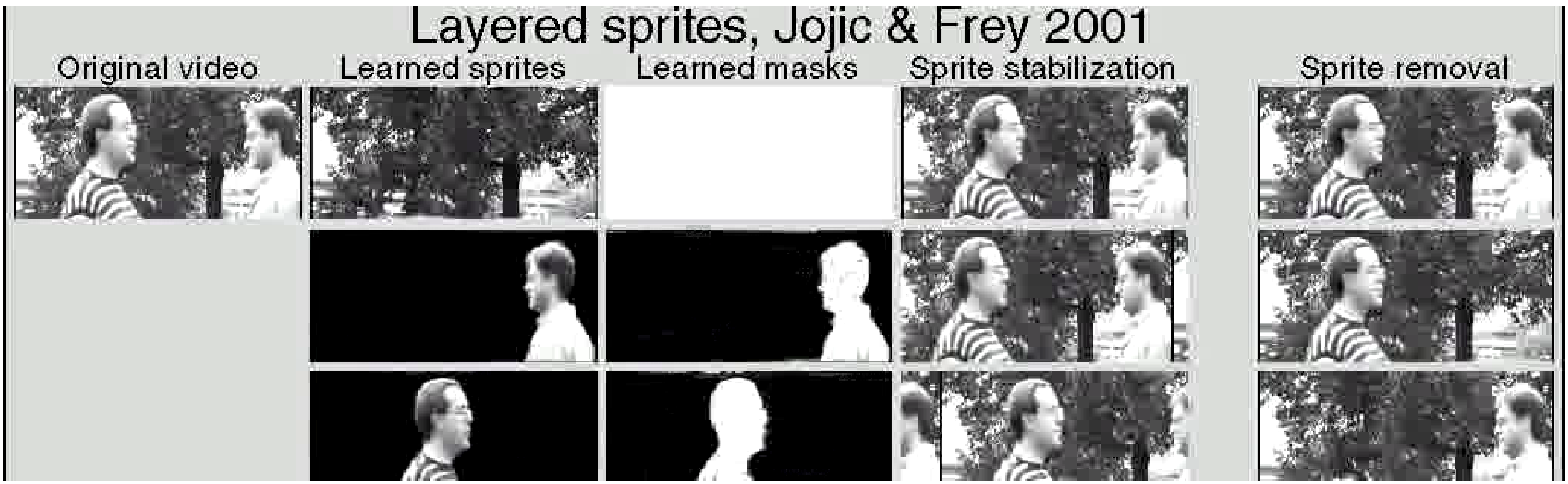
remove background



# Teaser: Flexible, Layered Sprites
# Locating and "filling in" occluded faces

Jojic & Frey, CVPR 2001, *www.psi.toronto.edu*

- User specifies # classes, # layers, input video/images




- Algorithm learns a *flexible appearance model* and a *flexible transparency map* for each class
- After training, the algorithm can infer the position, depth, appearance, and opacity of each subject

PAUSE

# Future Directions

- Better manifold models — Tennenbaum's "**Isomap**"
- Decision-theoretic methods
  - entropy, probability — Frey, Penev, Moghaddam

- Pose 2D/3D — Heisele, Poggio, Vetter
- Illumination — Belhumeur & Kriegman
- Anatomical models — Essa, Waters, Terzopoulis

***Surveillance*** **in <u>Unconstrained</u> Environments!**

# Face Recognition Resources

Face Recognition Home Page:

* http://www.cs.rug.nl/~peterkr/FACE/face.html

PAMI Special Issue on Face & Gesture (July '97)

FERET

* http://www.dodcounterdrug.com/facialrecognition/Feret/feret.htm

Face-Recognition Vendor Test (FRVT 2000)

* http://www.dodcounterdrug.com/facialrecognition/FRVT2000/frvt2000.htm

Biometrics Consortium

* http://www.biometrics.org