University of Wollongong

A Tutorial on Face Recognition

Ce Zhan

Advanced Multimedia Research Lab
University of Wollongong

University of Wollongong

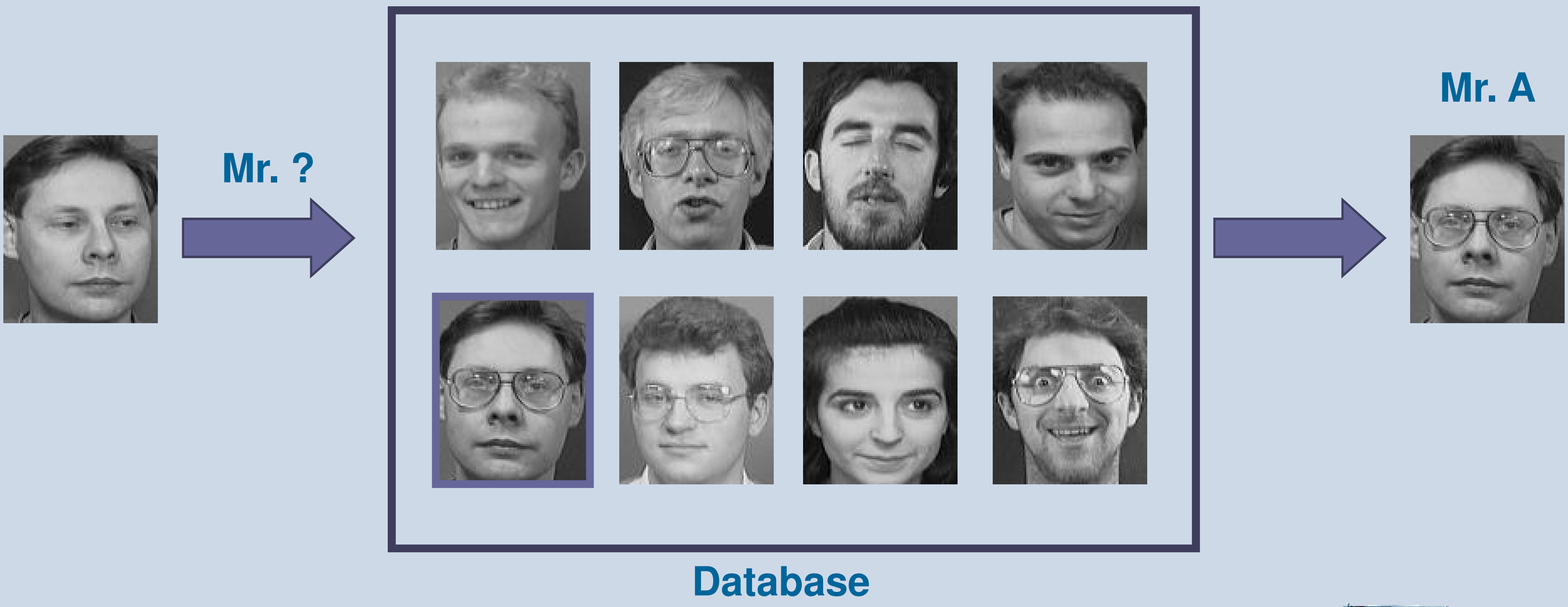
Outline

- Introduction
- How human recognize faces
- How computer recognize faces
  - Eigenfaces
  - State of the art
  - Challenges
- Related work (demos)

University of Wollongong

Introduction

The Task of Face Recognition



The Task of Face Recognition

- Still or video images of a scene
- Focus on still image in this tutorial
- A stored database (sample)
- Identification: determine the identity from the database
- Verification: confirm or reject the claimed identity

Why Face Recognition?

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

- Suitable for numerous applications
  - Access control and security
  - Smart cards
  - Information security
  - Law enforcement
  - Surveillance
  - Effective without the participant’s cooperation or knowledge

Application samples

Samsung Magicgate: A door lock control system using face verification technology

The Smart Gate installation at International Airport

Sony Aibo ERS-210 can recognize owner's face

How Humans Recognize Faces ?

Unconstrained conditions

- Illumination
- Expression
- Occlusion
- Accessories
- Low-resolution
- …

Unconstrained conditions

Individuals shown in order are:
Michael Jordan, Woody Allen, Goldie Hawn, Bill Clinton, Tom Hanks, Saddam Hussein, Elvis Presley, Jay Leno, Dustin Hoffman, Prince Charles, Cher, and Richard Nixon.

Local Vs. Holistic

Individual facial components, local feature

Relationships between the components, holistic configuration

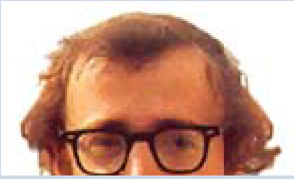
University of Wollongong

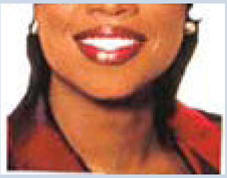
Local Vs. Holistic

Try to name the famous faces depicted in the two halves of the image

Local Vs. Holistic

Now try again



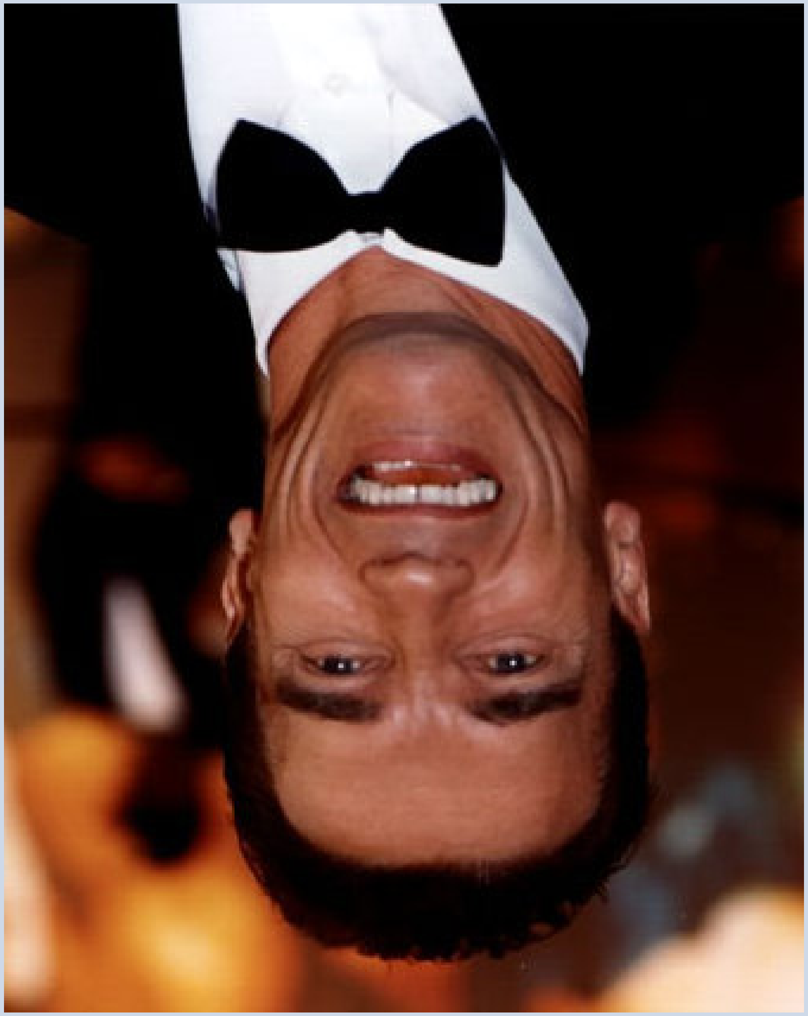


- It much more difficult to perform this task when the halves are aligned compared to misaligned halves

- Presumably because holistic processing interacts (and in this case, interferes) with feature-based processing

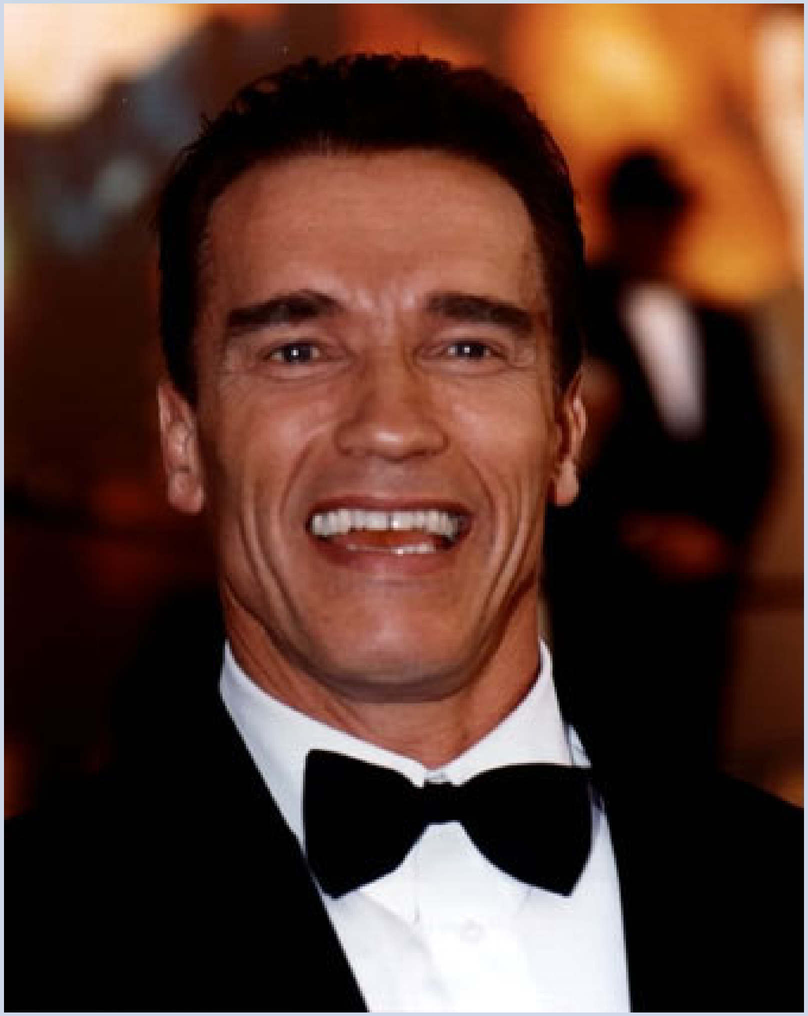
Local Vs. Holistic

Local Vs. Holistic



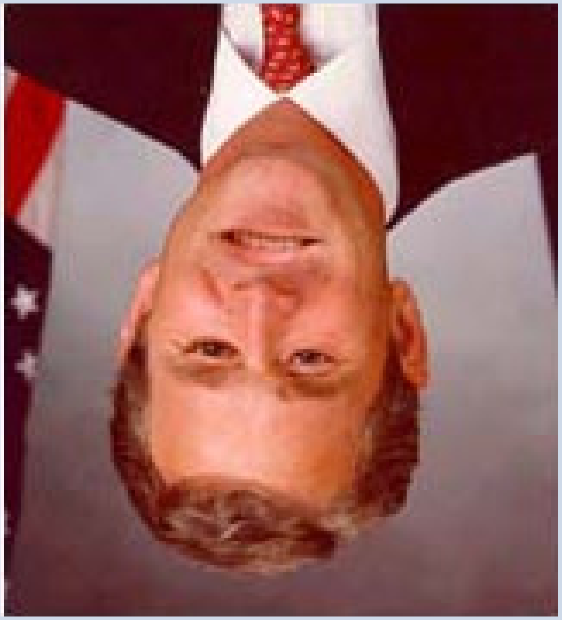
Do you recognize him?

Local Vs. Holistic



Arnold Schwarzenegger

Local Vs. Holistic



- Face inversion disrupts face recognition since holistic information is changed
- Who he is?
- Anything wrong?

University of Wollongong

Local Vs. Holistic

Local Vs. Holistic

Inverting the eyes and mouth of a face makes it look grotesque, but inverting this whole image removes this grotesque appearance

Significance of facial features

- Importance: eyes followed by the mouth and then the nose
- When using profiles, a distinctive nose shape could be more important
- The upper part of the face is more useful
- Eyebrows are as important at least as eyes

President Richard M. Nixon and actress Winona Ryder

What kind of configure information we need?

- Precise measurements of attributes such as inter-eye distance, width of mouth, and length of nose?
- It's still recognizable when the interfeature distance and distance ratios across the x and y dimensions are changed

Ratios of distances within the same dimension stay unchanged

High-Frequency Information Not Enough

- Edges and lines
- Initial representation for visual inputs, capture the most important aspects of images
- Invariant illumination variations
- However it’s not enough
- Do you recognize him?

Viewpoint invariant?

- Some research suggest object recognition is viewpoint invariant
- Some experiments suggest that memory for faces is highly viewpoint-dependent
- Generalization even from one profile viewpoint to another is poor
- Though generalization from one three-quarter view to the other is very good

Guidance from psychology study

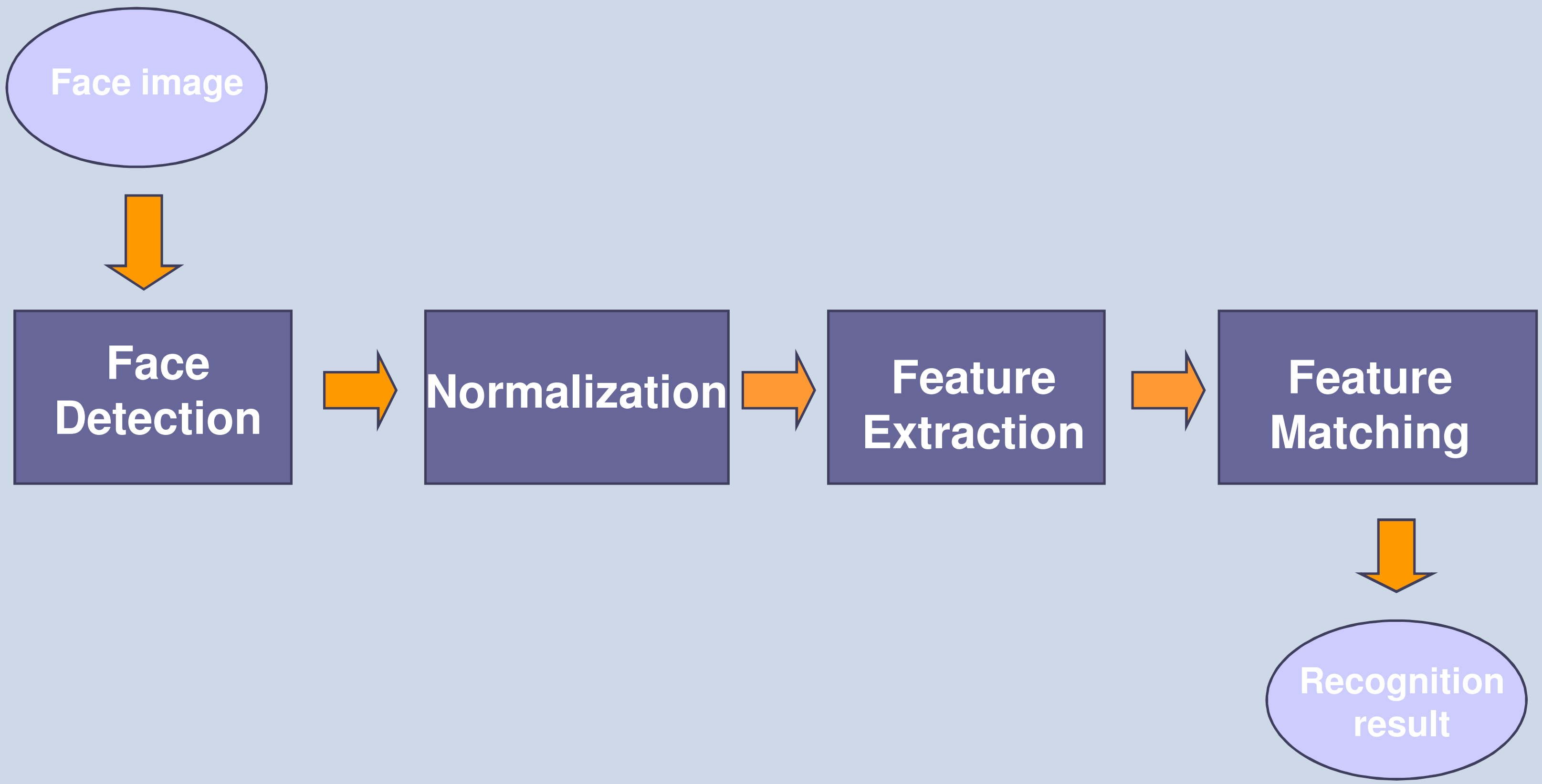
- Is face perception the results of holistic or local feature analysis?
  - Holistic methods
  - Local methods
  - Hybrid methods
- Significance of facial features
  - Different weights for features
- Viewpoint-invariant studies
  - Handle pose problem

How Computer Recognize Faces ?

Categorization

- Follow the guideline suggested by psychological study:
  - Holistic: using whole face region as input
  - Local feature based: using local features (geometric and /or appearance)
  - Hybrid: using both local features and whole face region

Processing Stages



Early Works

- The concept of automated face recognition was brought up in 1960s

- The first system
  - In 1965
  - Semi-automated
  - Little of the work was published

- Early research focused on the feasibility question

Early Works

- Geometric-based methods were widely use
  - geometric characteristics: the distance between two eyes, the thickness of lips…
  - manually extracted from the reference face
  - stored as templates for the later matching
  - provided limited results
  - experiments were usually carried out using datasets consisting of as few as 10 images.

One representative work

- By Goldstein in 1971
- Facial features are extracted manually by filling out a questionnaire for each face

Subject No. 77 (Col. 1-3)
Juror No. 6 (Col. 5-6)
Experiment No. (Col. 8-9)
Date 2/4/69

	1	2	3	4	5	Col.
1. HEAD a. Face Shape	Square	Round	Oval	Long		20
2. HAIR a. Coverage	Full	-	Receding	-	Bald	21
b. Length	Short	-	Average	-	Long	22
c. Texture	Straight	-	Wavy	-	Kinky	23
d. Part	Left	Middle	Right	None		24
6. MOUTH a. Lip Thickness UPPER	Thin	-	Medium	-	Thick	36
LOWER	Thin	-	Medium	-	Thick	37
b. Lip Overlap	Upper	Neither	Lower			38

One representative work

- 22 specific subjective features are extracted

Hair	Eyebrows	Eyes	Nose	Mouth	Chin	Ears	Cheeks	Forehead
Coverage	Weight	Opening	Length	Lip Thickness	Profile	Length	Sunken	Receding
full	thin	slit	short	Upper	receding	short	Average	Vertical
receding	medium	medium	medium	thin	straight	medium	Full	Bulging
bald	bushy	wide	long	medium	jutting	long		
Length	Separation	Spacing	Tip	thick		Protrusion		
short	separated	close	upward	Lower		flat		
average	meeting	medium	horizontal	thin		medium		
long		wide	downward	medium		sticking out		
Texture		Shade	Profile	thick		Lobes		
straight		light	concave	Lip Overlap		attached		
wavy		medium	straight	upper		medium		
kinky		dark	hooked	neither		not attached		
Shade				lower				
dark				Width				
medium				small				
light				medium				
gray				large				
white								

- Template matching for recognition

Early Works

- Not affected by illumination variations
- Do not need multiple training samples
- Geometric features not reliable (due to motions and view)
- The measurements and locations need to be manually computed (the biggest weakness)

University of Wollongong

Eigenfaces

Eigenfaces

- Using eigenfaces for recognition was developed by Sirovich and Kirby in 1987
- Used by Turk and Alex Pentland in face Recognition in 1991
- The first successful example of automated facial recognition technology
- Based on Principal Components Analysis (PCA)
- Holistic method
- Fast and Relatively Simple

PCA

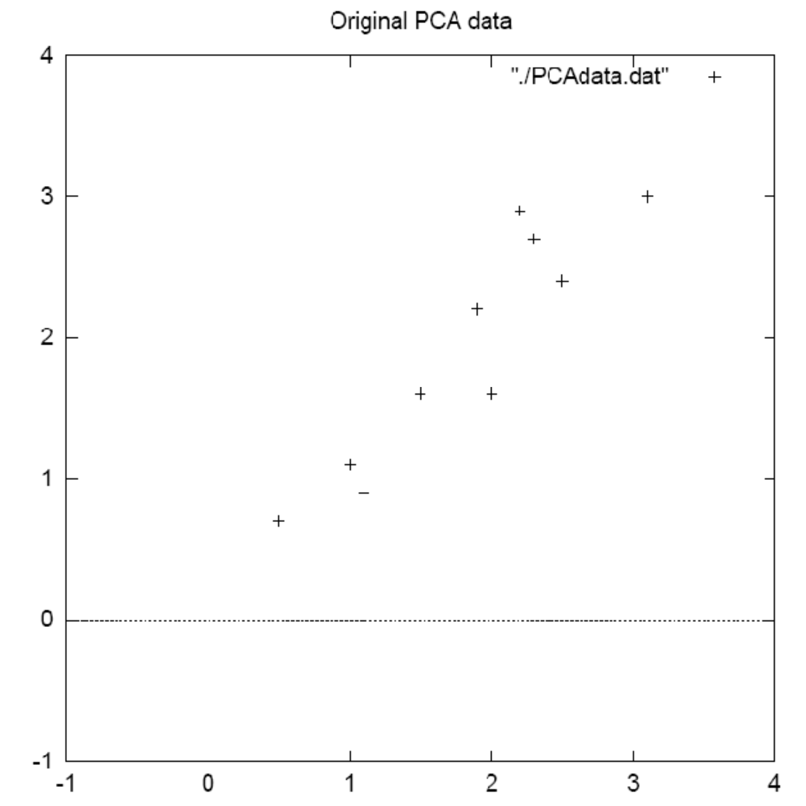
- A way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences
- Seeks directions that are efficient for representing the data
- Reduces the dimension of the data
- Speeds up the computational time

An simple example of PCA

- A 2 dimension dataset

Data =

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9



An simple example of PCA

➢ Subtract the mean

Data =

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

DataAdjust =

x	y
.69	.49
-1.31	-1.21
.39	.99
.09	.29
1.29	1.09
.49	.79
.19	-.31
-.81	-.81
-.31	-.31
-.71	-1.01

An simple example of PCA

- Calculate the covariance matrix

$$cov = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

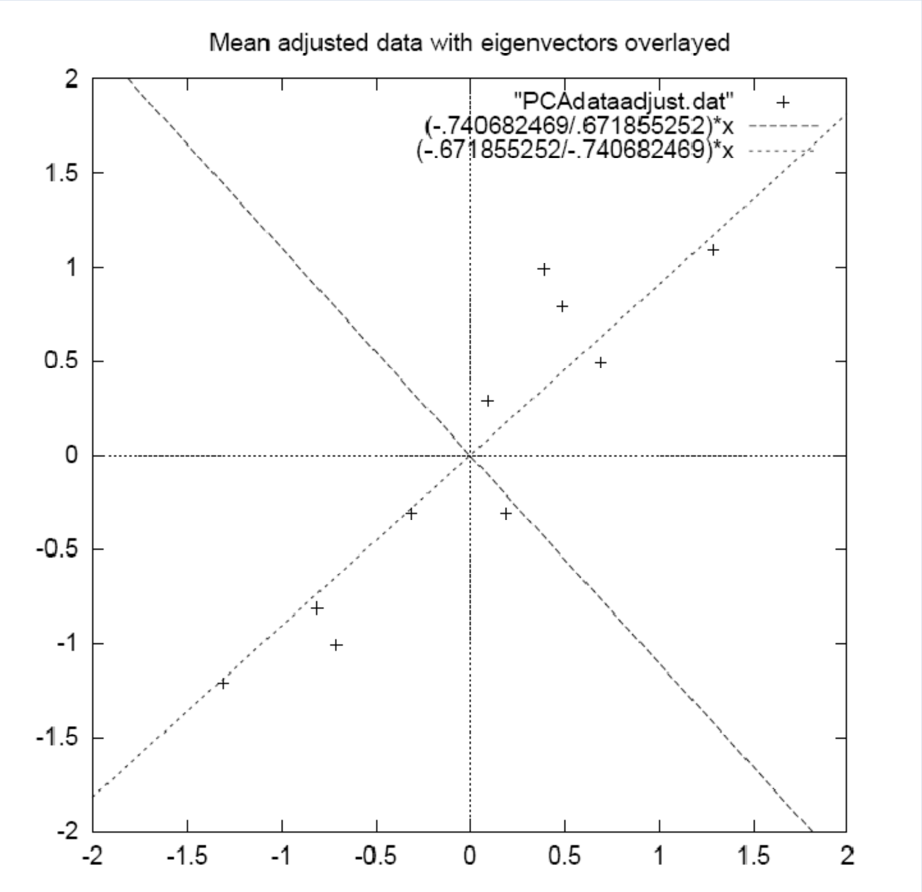
- Calculate the eigenvectors and eigenvalues of the covariance matrix

$$eigenvalues = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

$$eigenvectors = \begin{pmatrix} -.735178656 & -.677873399 \\ .677873399 & -.735178656 \end{pmatrix}$$

An simple example of PCA

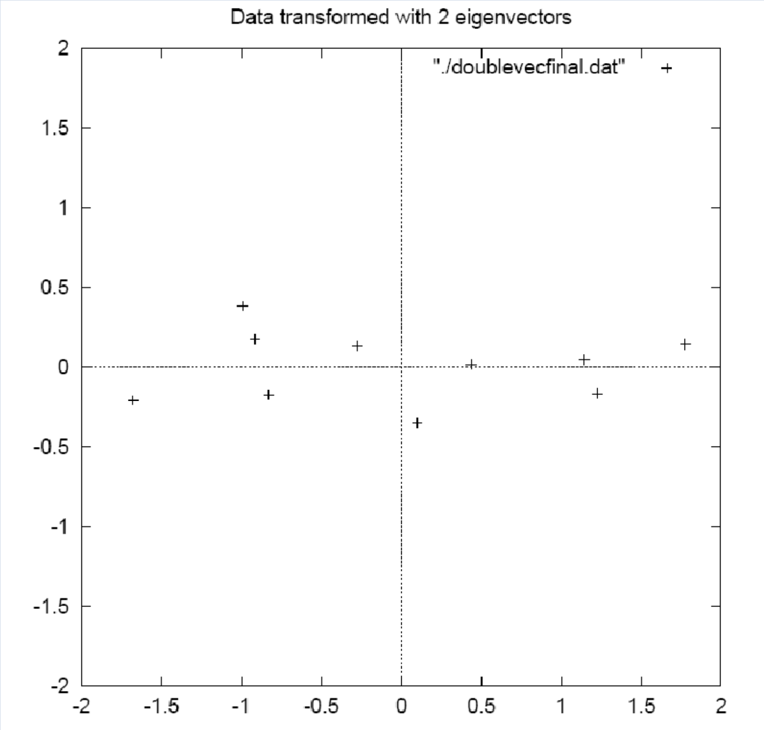
- One of the eigenvectors goes through the middle of the points, like drawing a line of best fit
- That eigenvector is showing us how these two data sets are related along that line.
- The second eigenvector gives us the other, less important, pattern in the data, that all the points follow the main line, but are off to the side of the main line by some amount.



University of Wollongong

An simple example of PCA

- We can either transform the data with both of the eigenvectors



This is basically the original data, rotated so that the eigenvectors are the axes

We have now classified our data point as a combination of the contributions from each of those lines.

lost no information in this decomposition

University of Wollongong

An simple example of PCA

- Or taking only the eigenvector with the largest eigenvalue

the contribution due to the smaller eigenvector is removed

Basically we have transformed our data so that is expressed in terms of the patterns between them

the patterns are the lines that most closely describe the relationships between the data

Transformed Data (Single eigenvector)

x
-.827970186
1.77758033
-.992197494
-.274210416
-1.67580142
-.912949103
.0991094375
1.14457216
.438046137
1.22382056

University of Wollongong

The algorithm

- Assumptions:
  - Images with $N \times N = N^2$
  - M is the number of images in the database
  - P is the number of persons in the database

University of Wollongong

The algorithm

- Assume we have following training database, M=8:

$$= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{N^2} \end{pmatrix}$$

$$= \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{N^2} \end{pmatrix}$$

$$= \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{N^2} \end{pmatrix}$$

$$= \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_{N^2} \end{pmatrix}$$

$$= \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_{N^2} \end{pmatrix}$$

$$= \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N^2} \end{pmatrix}$$

$$= \begin{pmatrix} g_1 \\ g_1 \\ \vdots \\ g_{N^2} \end{pmatrix}$$

$$= \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_{N^2} \end{pmatrix}$$

The algorithm

- Compute the average face:

$$\vec{m} = \frac{1}{M}\begin{pmatrix} a_1 + b_1 + \cdots + h_1 \\ a_2 + b_2 + \cdots + h_2 \\ \vdots \\ a_{N^2} + b_{N^2} + \cdots + h_{N^2} \end{pmatrix}$$

The algorithm

- Compute the average face:

The algorithm

- Subtract the average face from the training faces

$$\vec{a} = \begin{pmatrix} a_1 - m_1 \\ a_2 - m_2 \\ \vdots \quad \vdots \\ a_{N^2} - m_{N^2} \end{pmatrix} \vec{b} = \begin{pmatrix} b_1 - m_1 \\ b_2 - m_2 \\ \vdots \quad \vdots \\ b_{N^2} - m_{N^2} \end{pmatrix} \vec{c} = \begin{pmatrix} c_1 - m_1 \\ c_2 - m_2 \\ \vdots \quad \vdots \\ c_{N^2} - m_{N^2} \end{pmatrix} \cdots \vec{h} = \begin{pmatrix} h_1 - m_1 \\ h_2 - m_2 \\ \vdots \quad \vdots \\ h_{N^2} - m_{N^2} \end{pmatrix}$$

The algorithm

- Now we build A matrix which is N^2 by M

$$A = \left[\vec{a}_m \, \vec{b}_m \, \vec{c}_m \, \vec{d}_m \, \vec{e}_m \, \vec{f}_m \, \vec{g}_m \, \vec{h}_m \right]$$

- The covariance matrix is N^2 by N^2

$$Cov = AA^T$$

The algorithm

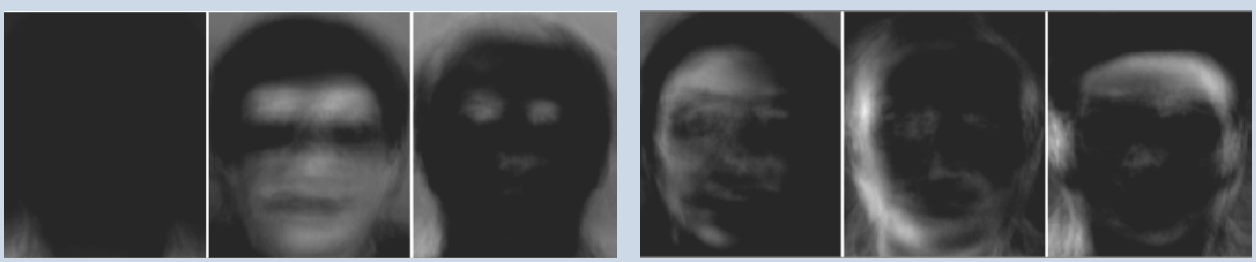
- Find eigenvalues of the covariance matrix
  - The matrix is very large
  - The computational effort is very big
- We are interested in at most M eigenvalues
  - Reduce the dimension of the matrix
  - A trick from linear algebra

The algorithm

- Compute another matrix which is M by M

$$L = A^T A$$

- Find the M eigenvalues and eigenvectors
  - Eigenvectors of Cov and L are equivalent
  - Each eigenvecotor is called a eigenface of the training data (AA^T)



Eigenfaces

The algorithm

- Build matrix V from the eigenvectors of L
- Eigenvectors of Cov are linear combination of image space with the eigenvectors of L

$$U = AV$$

- Eigenvectors represent the variation in the faces

The algorithm

- Compute for each face its projection onto the face space

$$\Omega_1 = U^T(\vec{a}_m), \Omega_2 = U^T(\vec{b}_m), \cdots \Omega_8 = U^T(\vec{h}_m),$$

- Compute a threshold

$$\theta = \frac{1}{2}\max\{\|\Omega_i - \Omega_j\|\} \quad for \quad i, j = 1 \ldots M$$

University of Wollongong

The algorithm

- To recognize a test face

$$= \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{N^2} \end{pmatrix}$$

- Subtract the average face from it

$$\vec{r} = \begin{pmatrix} r_1 - m_1 \\ r_2 - m_2 \\ \vdots \quad\quad \vdots \\ r_{N^2} - m_{N^2} \end{pmatrix}$$

University of Wollongong

The algorithm

- Compute its projection onto the face space

$$\Omega = U^T(\vec{r}_m)$$

- Compute the distance in the face space between the test face and all known faces

$$\varepsilon_i^2 = \left\|\Omega - \Omega_i\right\|^2 \quad for \quad i = 1 \ldots M$$

University of Wollongong

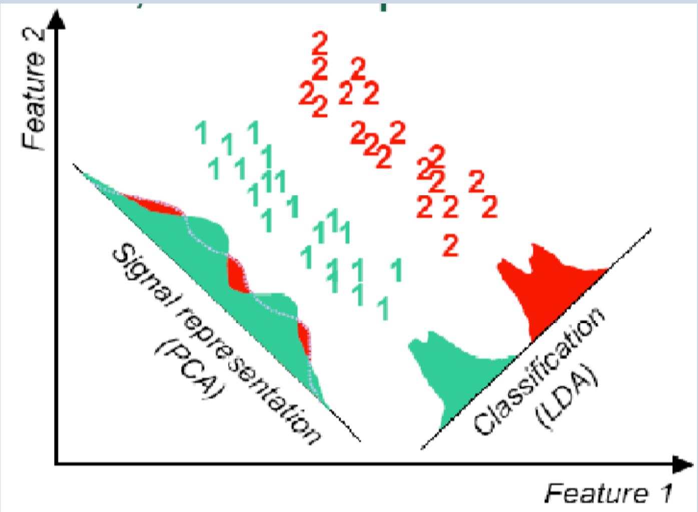
The algorithm

- For identification
  - The face will be indentified as the nearest known face in the face space

- For verification
  - $if \quad \min\{\varepsilon_i\} < \theta \quad (i = 1 \ldots M)$ it is a known face

Linear Discriminant Analysis (LDA)

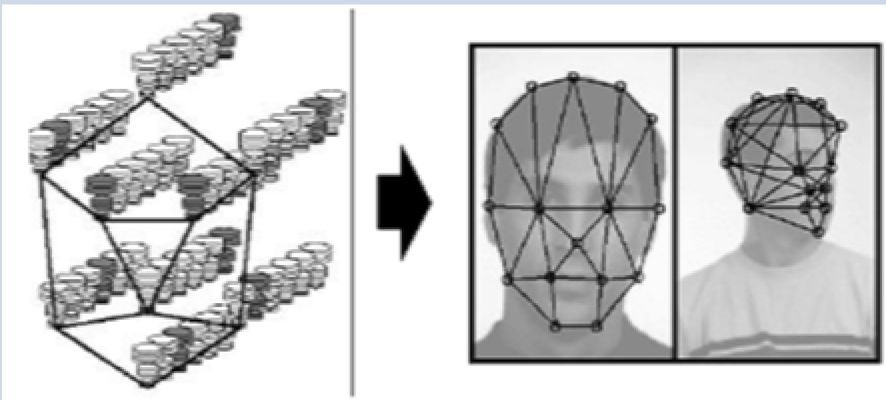
- Extension of eigenface
- Seeks to find a linear transformation by maximising the between-class variance and minimising the within-class variance
- Improve classification performance when more than one samples are available per class

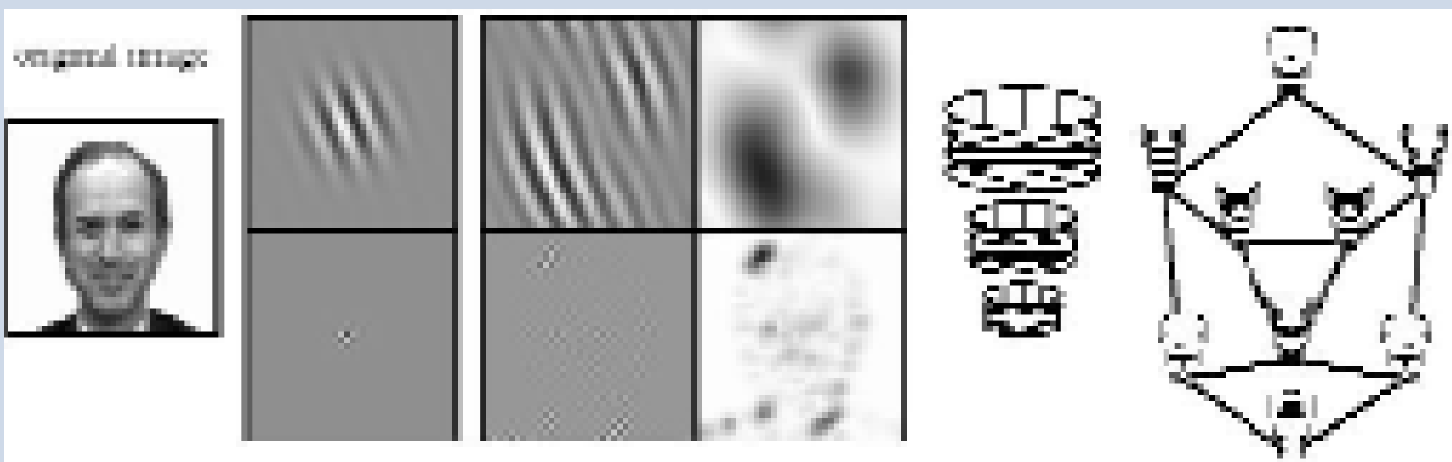


Elastic Bunch Graph Matching (EBGM)

- Local feature based method
- Faces are represented as graphs, with nodes positioned at fiducial points and edges labeled with distance vectors
- Each node contains a set of Gabor wavelet coefficients
- The identification of a new face consists of determining among the constructed graphs, the one which maximises the graph similarity function

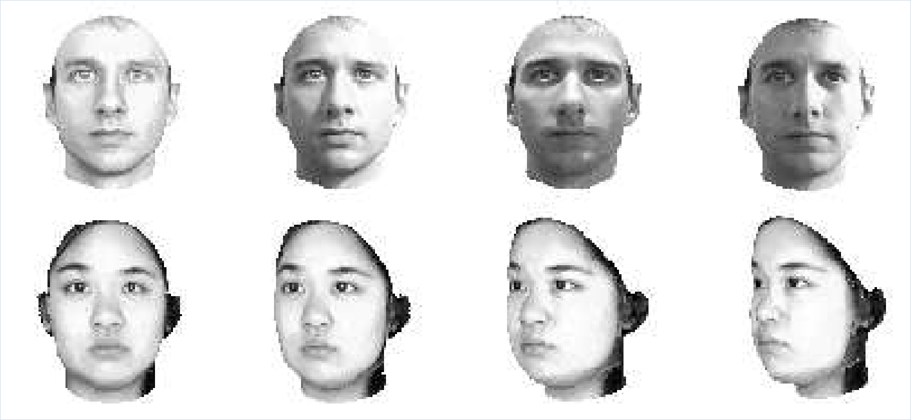
Elastic Bunch Graph Matching (EBGM)



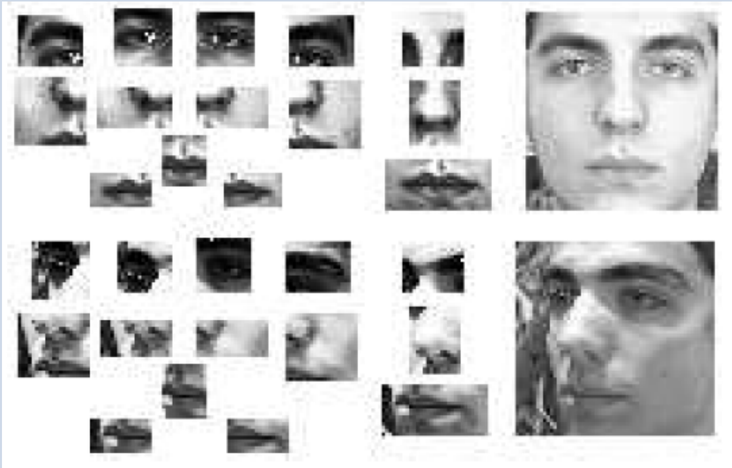


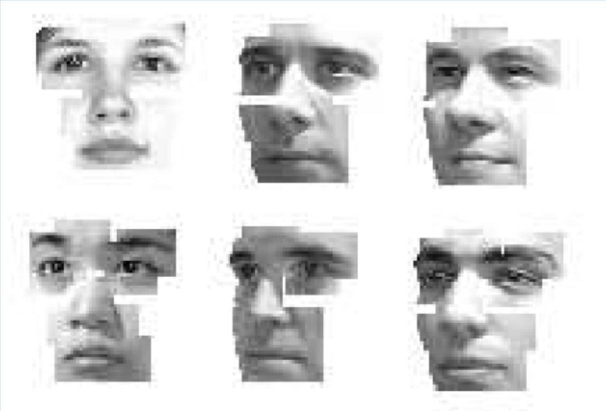
Component-based face recognition with morphable models [6]

- A hybrid method
- Synthetic training samples from morphable models under different lightings and poses
- Solving one sample and profile problem



Component-based face recognition with morphable models [6]





State of the Art

- In 2005, the performance of the latest face recognition algorithms were evaluated in [5]
  - New algorithms are 10 times more accurate than the face recognition algorithms of 2002
  - 100 times more accurate than those of 1995.
- Reached a significant level but still far away from the capability of human perception
- Take advantage of domain knowledge

Questions to be answered

- How big should be the face image size?
- Is face recognition unique compared to other object recognition?
- How to compare (evaluate) different systems?

Challenges

- Recognition being too sensitive to inaccurate facial feature localization

- Robustly recognizing faces
  - Small and/or noisy images
  - Unconstrained variation: lighting, pose, expression
  - Images acquired years apart

- Limited training samples

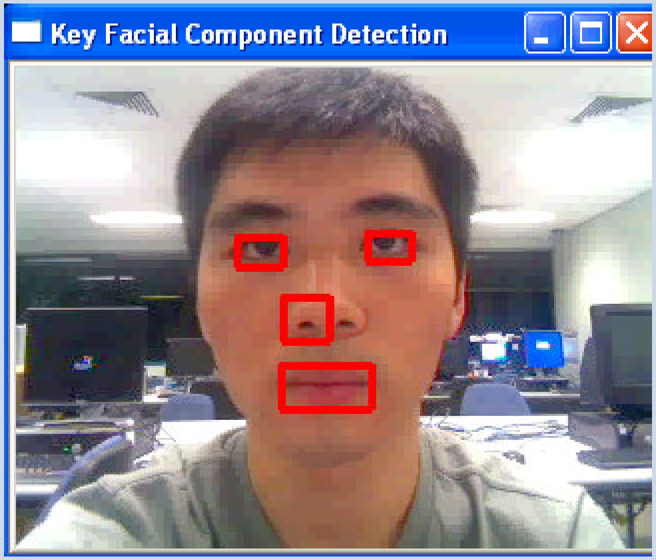
- What's the principal and optimal way to arbitrate and combine local features and global features

University of Wollongong

Related Works

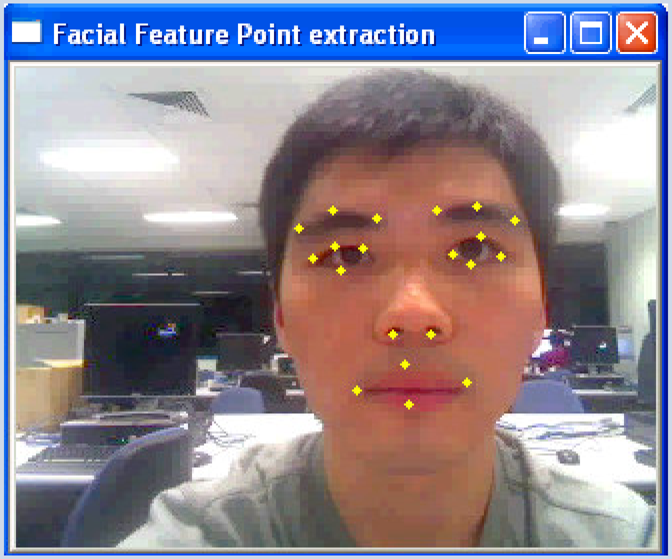
Key Facial Component Detection

❖ A Real Time Demo



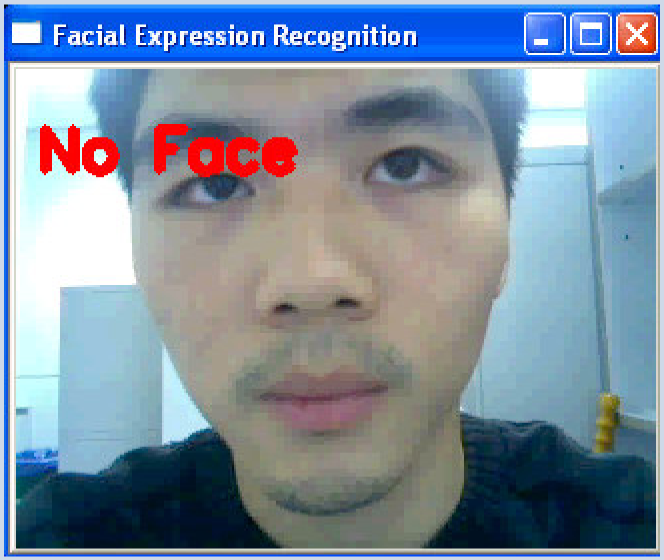
Facial Landmark Localization

- A Real Time Demo



Facial Expression Recognition

❖ A real time demo



References

[1] "Face Recognition by Humans: 19 Results all Computer Vision Researchers Should Know About", Pawan Sinha, Benjamin Balas, Yuri Ostrovsky, Richard Russell, Proceedings of the IEEE, Vol. 94, No. 11, November 2006, pp. 1948-1962

[2] "Face Recognition: A Literature Survey", W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, ACM Computing Surveys, 2003, pp. 399-458

[3] "Eigenfaces for Recognition", M. Turk, A. Pentland, Journal of Cognitive Neurosicence, Vol. 3, No. 1, 1991, pp. 71-86

[4] "Identification of human faces", A.J. Goldstein, L.D. Harmon and A.B. Lesk, Proceedings of the IEEE Volume 59, Issue 5, May 1971, pp. 748 – 760

[5] "Overview of the face recognition grand challenge", P.J. Phillips et.al, CVPR 2005, Volume: 1, pp. 947- 954

[6] "Component-Based Face Recognition with 3D Morphable Models" B. Weyrauch, B. Heisele, J. Huang, V. Blanz, CVPRW04, pp.85

Thanks !