# Face Recognition

**Massimo Tistarelli**
*Computer Vision Laboratory*
*University of Sassari – Italy*
*tista@uniss.it*

# Credits

- From the laboratory staff:

  Linda Brodo
  Marinella Cadoni
  Massimo Gessa
  Enrico Grosso
  Andrea Lagorio
  Ludovica Lorusso
  Gianluca Masala
  Luca Pulina
  Ajita Rattani
  Elif Surer
  Yunlian Sun
  Daksha Yadav (past visiting)
  Yu Guan (past visiting)
  Marcos Ortega Hortas (past visiting)
  Albert Ali Salah (past visiting)

# Credits

- …and other labs:

  Manuele Bicego – University of Verona
  Rama Chellappa – University of Maryland
  Anil Jain – Michigan State University
  Alice O'Toole – University of Texas at Dallas
  Jonathon Phillips – NIST
  Norman Poh – University of Surrey

# Face recognition

I. **What happened in 20+ years of research in face recognition?**

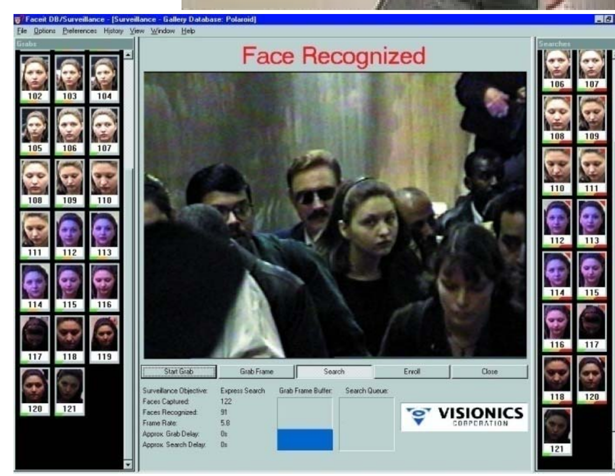
II. **What can we learn?**

III. **What is still to be done?**

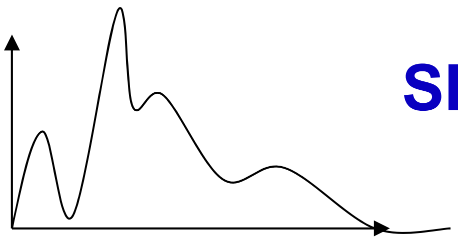
# *Why* face recognition?

- Most natural for humans
- Highly acceptable and non-intrusive
- Highly applicable:
  - Static identity verification
  - Uncontrolled face detection and identification from video


- Medium to Low performances
- Not unique (twins)
- Aging and time effects



# *How* face recognition?



**SIGNAL PROCESSING**

***Information Theory***

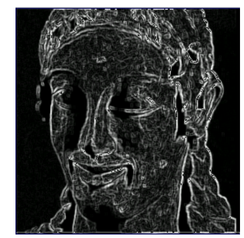
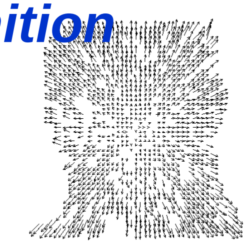
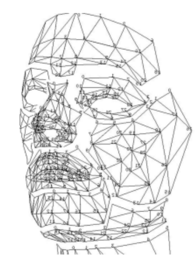


**IMAGE PROCESSING**

***Communication***
***and***
***Computer Science***

***Pattern Recognition***



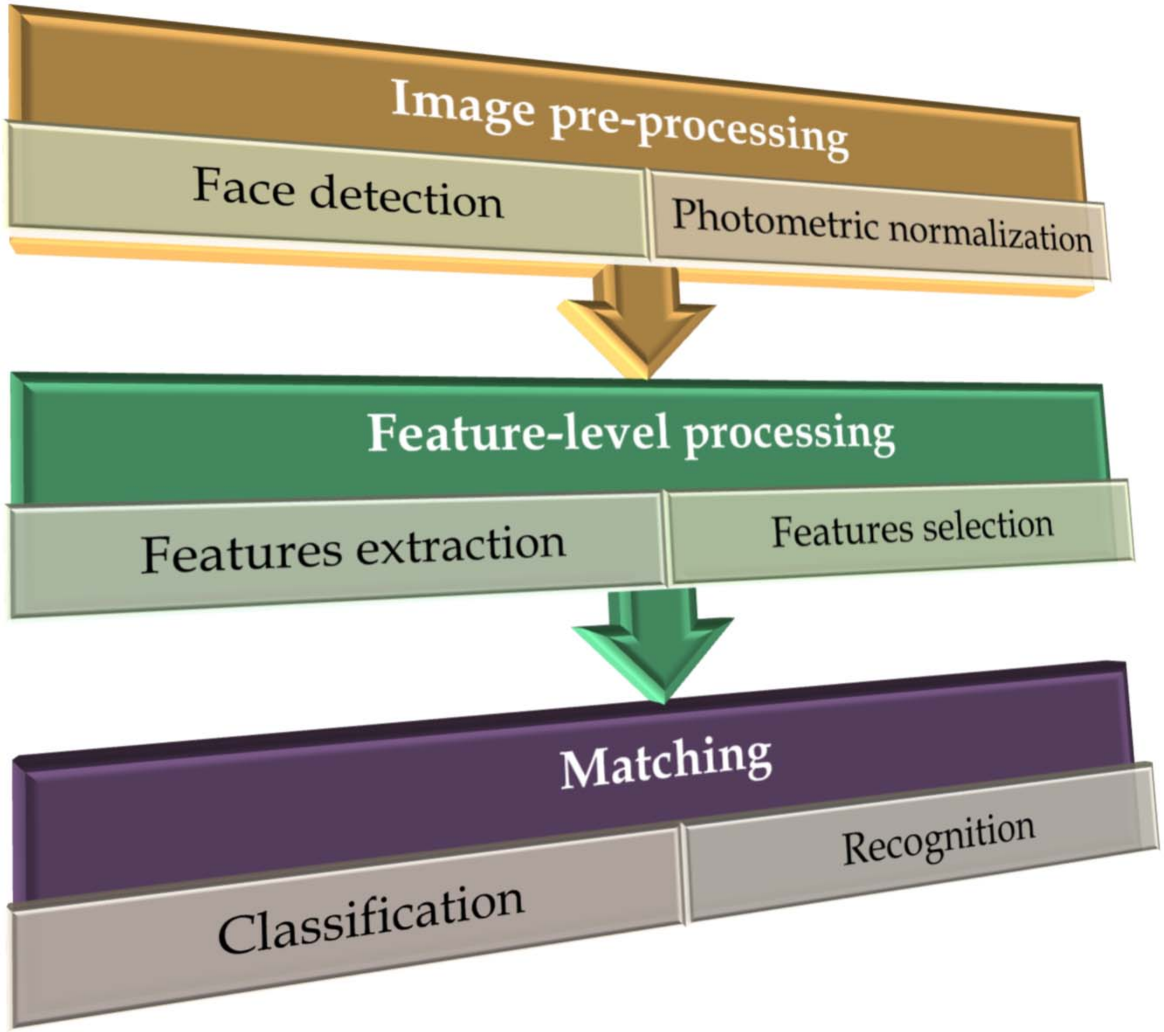


**COMPUTER VISION**

***Machine Learning***

**IMAGE UNDERSTANDING**

# Fundamentals



# Must Read!

- D.H. Ballard and C.M. Brown *Computer Vision*
- W.K. Pratt *Digital Image Processing*
- B.K.P. Horn *Robot Vision*
- A.K. Jain *and* S. Li *Handbook of Face Recognition*
- E. Trucco *and* A. Verri *Introductory Techniques for 3D Computer Vision*
- J. Bigun *Vision with direction*
- M. Tistarelli, R. Chellappa, S. Z. Li *Handbook of Remote Biometrics*
- C. M. Bishop *Pattern Recognition and Machine Learning*
- *Others ...*

# An *almost* «fair» comparison (*from Jain et al 1997.*)

Vision Lab

| BIOMETRICS | Universality | Uniqueness | Permanence | Collectability | Performance | Acceptability | Circumvention |
|---|---|---|---|---|---|---|---|
| Face | High | Low?? | Medium | High | Low?? | High | Low?? |
| Fingerprint | Medium | High | High | Medium | High | Medium | Low |
| Hand Geometry | Medium | Medium | Medium | High | Medium | Medium | Medium |
| Iris | High | High | High | Medium | High | Low | High |
| Retinal Scan | High | High | Medium | Low | High | Low | High |
| Signature | Low | Low | Low | High | Low | High | Low |
| Voice | Medium | Low | Low | Medium | Low | High | Low |
| Facial Thermogram | High | High | Low | High | Medium | High | High |

# Biometrics evaluations…

| Modality | Test Label | Test Parameter | False Reject Rate | False Accept Rate |
|---|---|---|---|---|
| **Fingerprint** | FpVTE 2003 | US Government operational data | 0.1% | 1% |
| **Fingerprint** | FVC 2006 | Heterogeneous population (young, elderly) | 2.2% | 2.2% |
| **Face** | FRGC 2006 | Controlled Illumination, high-resolution images | 0.8-1.6% | 0.1% |
| **Voice** | NIST 2004 | Text independent, multi-lingual | 5-10% | 2-5% |
| **Iris** | ITIRT 2005 | Indoor environment | 0.99% | 0.94% |
| **Iris** | ICE 2006 | Controlled Illumination, broad quality range | 1.1-1.4% | 0.1% |

# Genotypic vs phenotipic traits

**Biometric traits develop:**

1. through genetics: **Genotypic**
2. through random variations in the early phases of an embryo's development: **Phenotypic**
3. through training: **Behavioral**

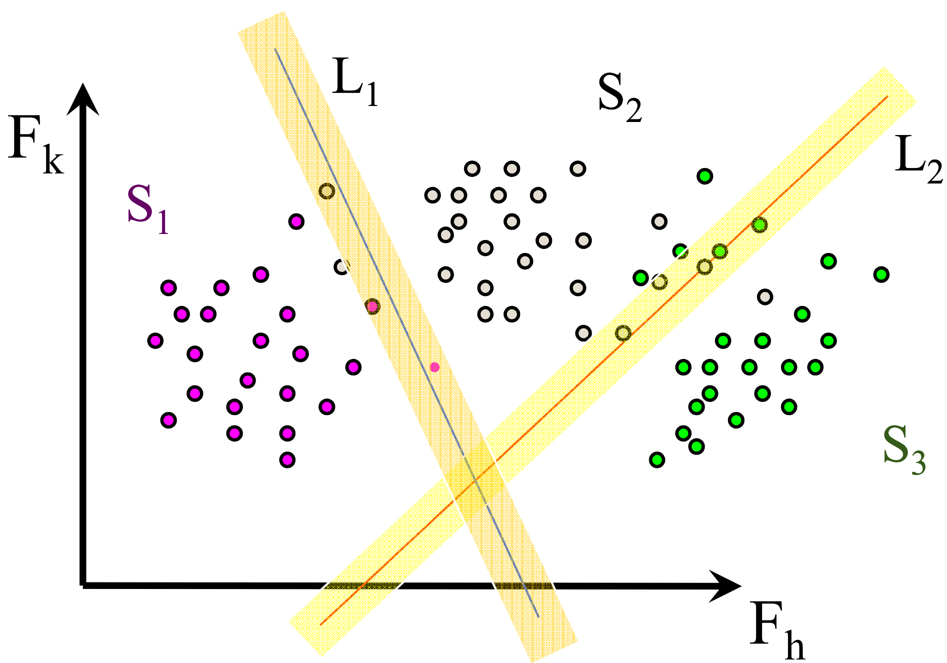
| *Biometric Trait* | *genotypic* | *phenotypic* | *behavioral* |
|---|---|---|---|
| Fingerprint (only minutia) | o | ooo | o |
| Signature (dynamic) | oo | o | ooo |
| **Facial geometry** | ooo | o | o |
| Iris pattern | o | ooo | o |
| Retina (Vein structure) | o | ooo | o |
| Hand geometry | ooo | o | o |
| Finger geometry | ooo | o | o |
| Vein structure of the back of hand | o | ooo | o |
| Ear form | ooo | o | o |
| Voice (Tone) | ooo | o | oo |
| DNA | ooo | o | o |
| Odor | ooo | o | o |
| Keyboard Strokes | o | o | ooo |
| **Comparison: Password** | | | (ooo) |

**Source:** http://www.bromba.com/faq/biofaqe.htm#entstehen

# Identification of human faces

**A class (identity) separation problem:**

- Choice of optimal representation
- Inter-class similarity vs intra-class variability




$= \Psi(F_h, F_k)$

# False match and false non-match

Faces that look similar
**FALSE ACCEPTANCE**

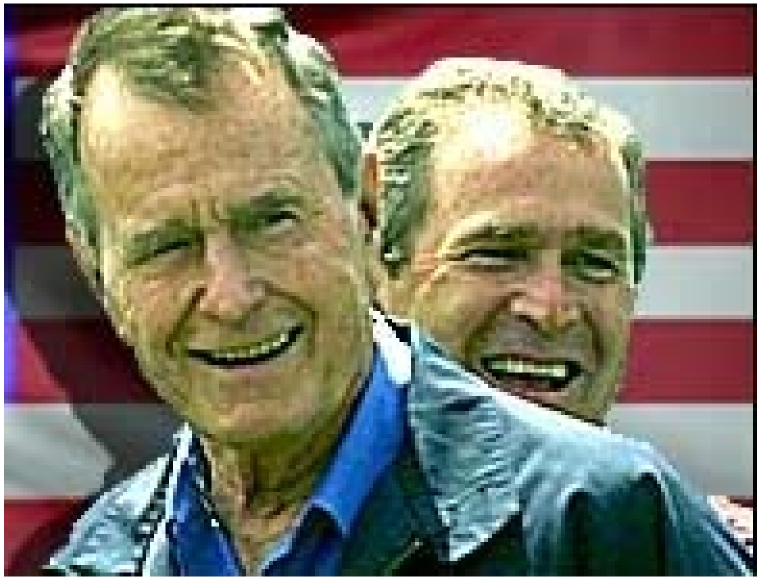
Faces that look different
**FALSE REJECTION**


# Inter-class *similarity*

**Two different people may have very similar appearance**

**FALSE MATCH**

**Twins**



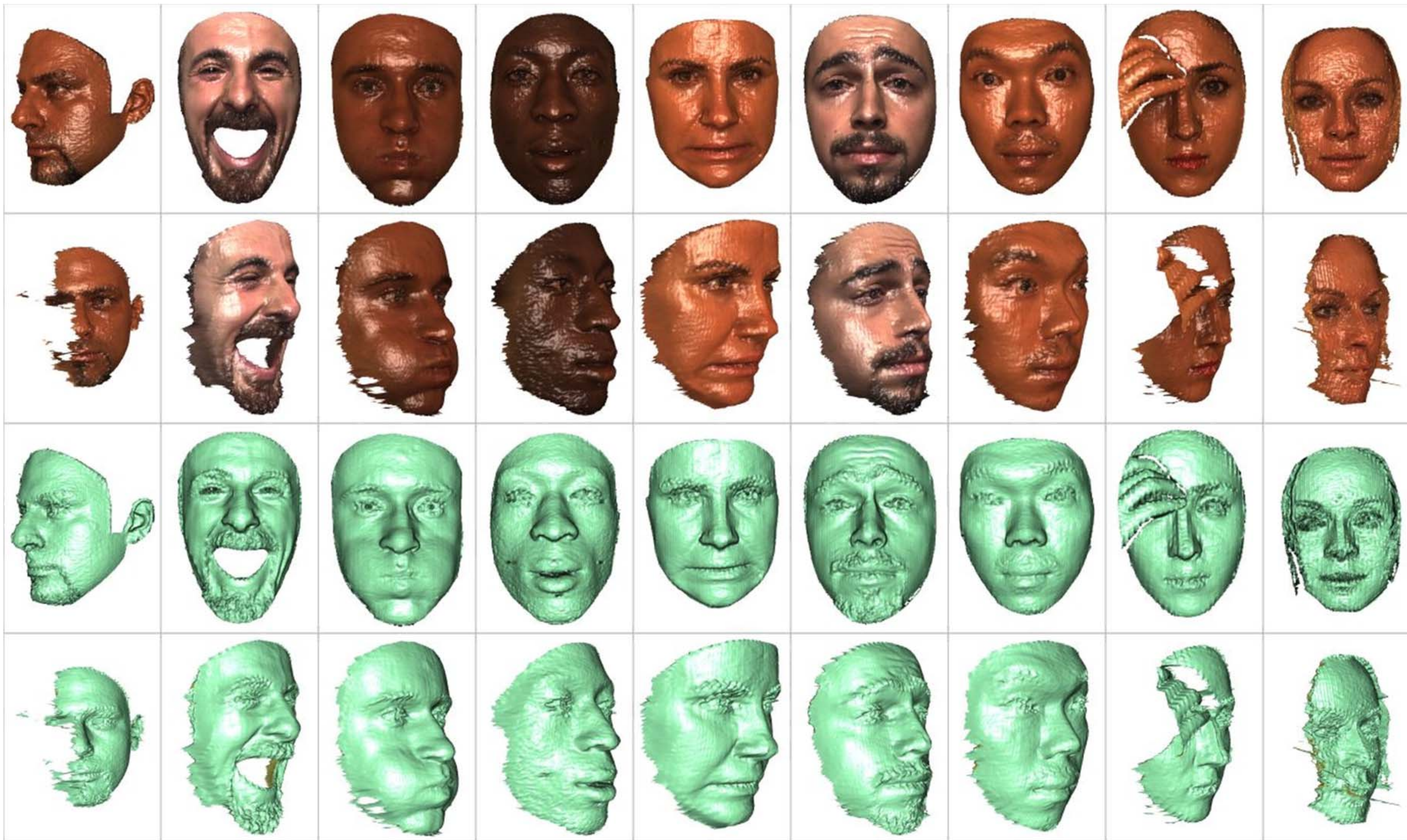
**Father and son**

# Intra-class *variability*

**The same person may present very different biometric samples**

**FALSE NON-MATCH**

# Face shape and texture

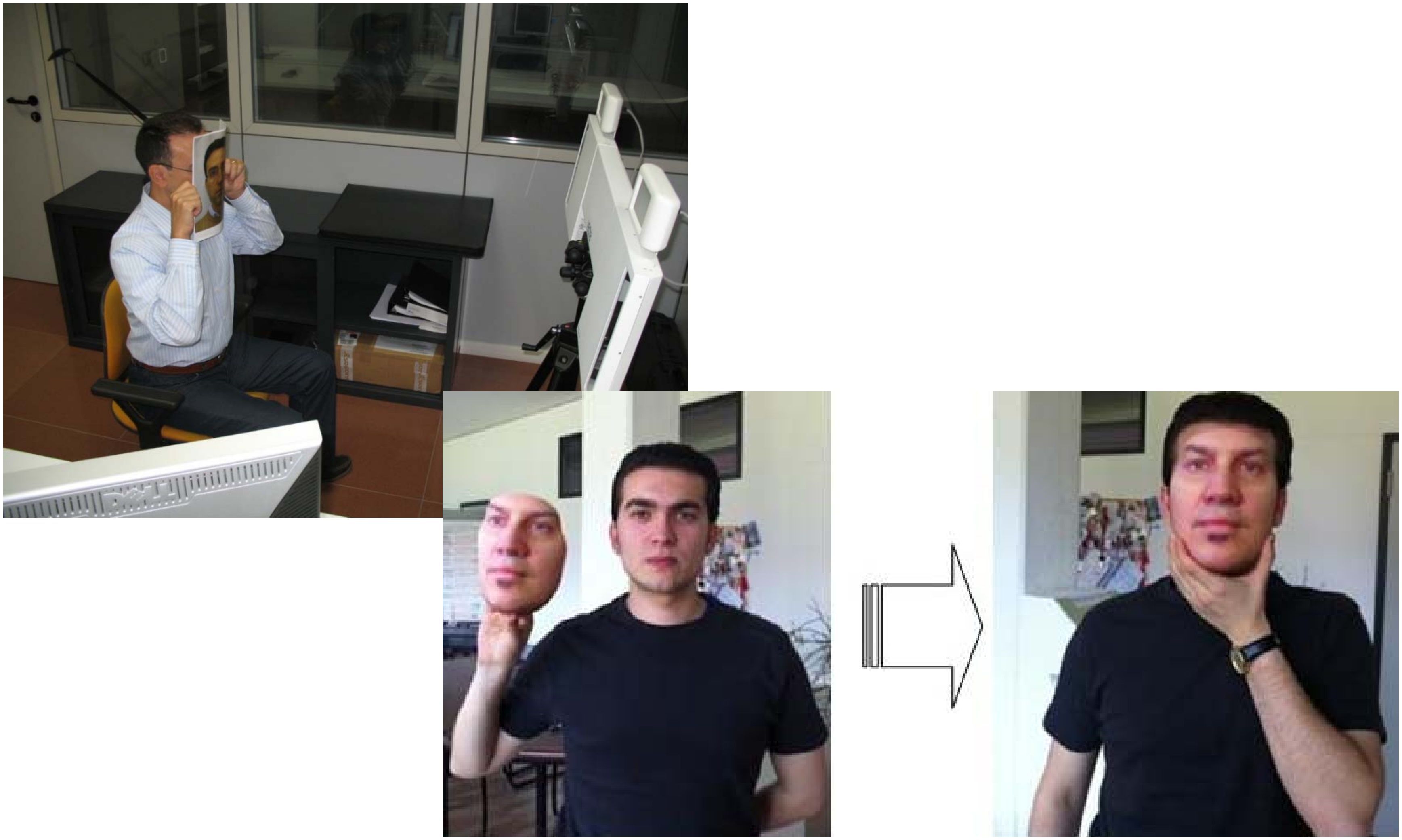


A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, "Bosphorus Database for 3D Face Analysis", The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008) Roskilde University, Denmark, May 2008.

# Face makeup

# Face surgery

# Face spoofing

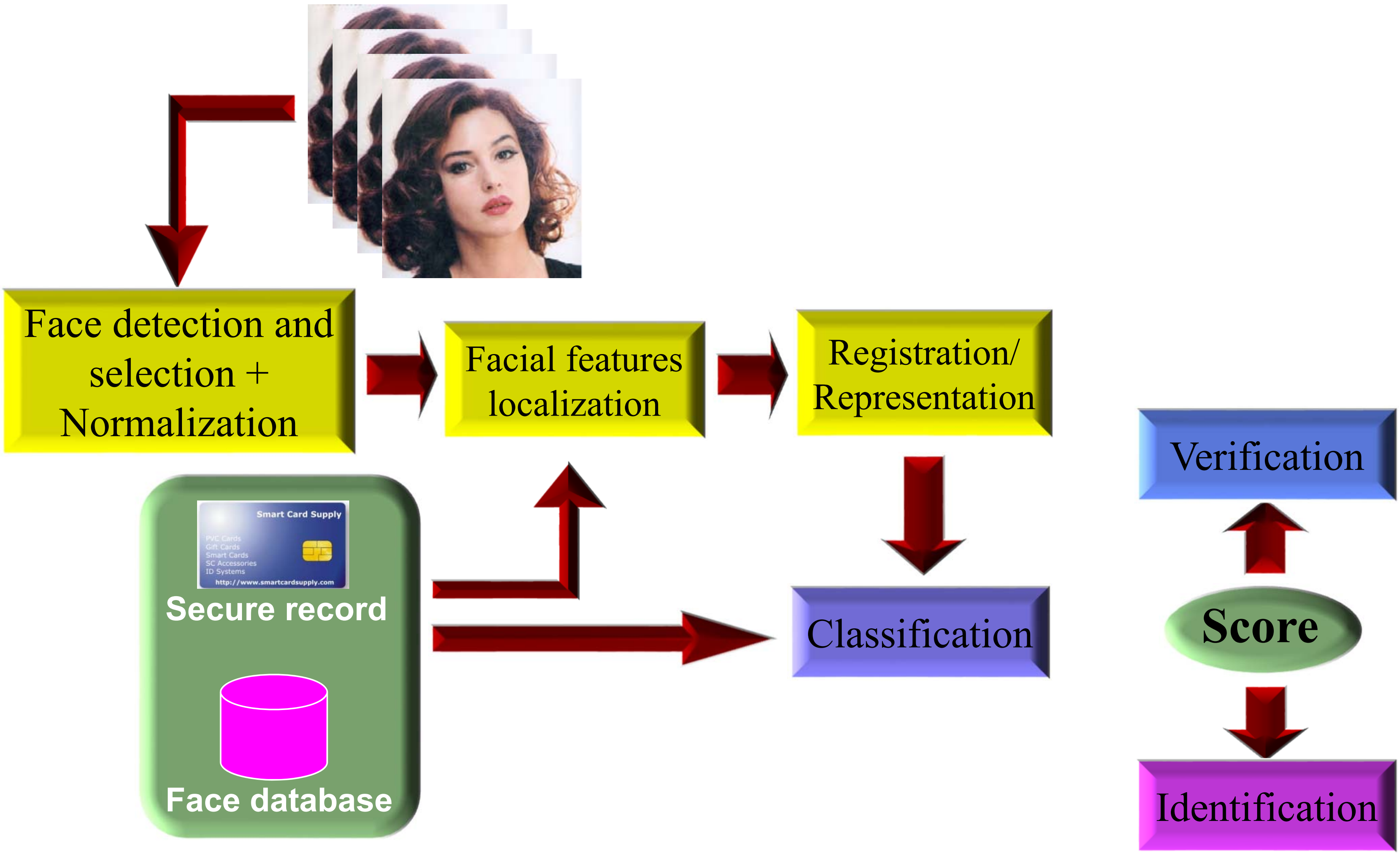


# An ill-posed problem

**Do they look at all similar… ?**

# FACE RECOGNITION TECHNOLOGIES

# Face recognition operation



# Taxonomy of face recognition

Vision Lab

Algorithms

**Pose-dependency**

- Pose-dependent
- Pose-invariant

**Face representation**

- Viewer-centered Images
- Object-centered Models

**Matching features**

**Appearance-based (Holistic)**

PCA, LDA,
ICA, LPP, SRC
Kernel . . .

**Feature-based (Analytic)**

Gabor sets
SIFT, SURF,
LBP-XX . . .

**Hybrid**

LFA
EGBM/JETs
AAM . . .

*Deep Learning...*

- Gordon et al., 1995 3D from 2D
- Lengagne et al., 1996 3D from stereo
- Atick et al., 1996 LFA
- Yan et al., 1996 Geometric modeling
- Blanz & Vetter, 1999 Morph. models
- Zhao et al., 2000 Shape from Shading
- Zhang et al., 2000 3D from video

# Holistic face recognition

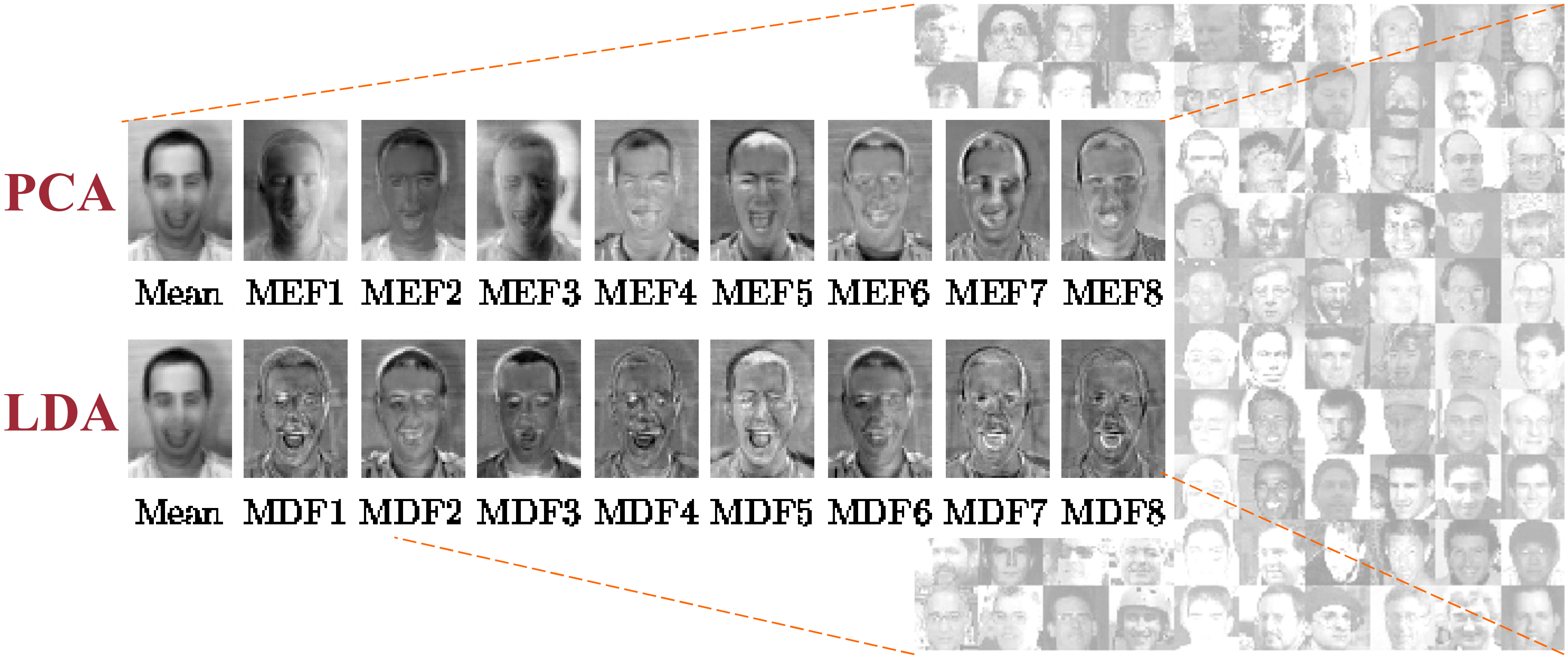
- The basic idea of many similar approaches is to define a basis of vectors to describe any face in the "*universal space*" of all existing faces…

- The basic tool is the ***Singular Values Decomposition***:

$$\mathbf{A} = \mathbf{U} \cdot \Sigma \cdot \mathbf{W}$$

- The eigenvectors (*r* columns of U) of the decomposition define the basis of vectors and the eigenvalues $\sigma_i$ define the "relevance" of each eigenvector (*eigenface)*

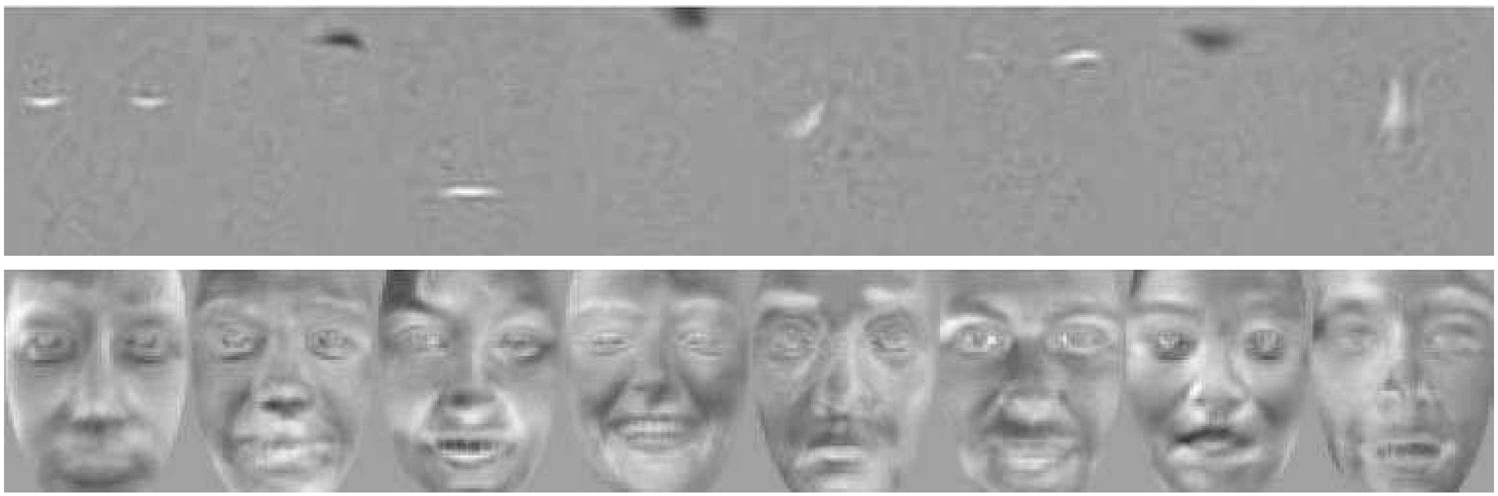
# Holistic face recognition




- Both PCA and LDA produce a set of orthogonal basis images.
- Both provide a compact and global representation of face images.
- LDA explicitly attempts to model the difference between the classes of data.
- PCA does not take into account any difference in class.

# Holistic face recognition

ICA

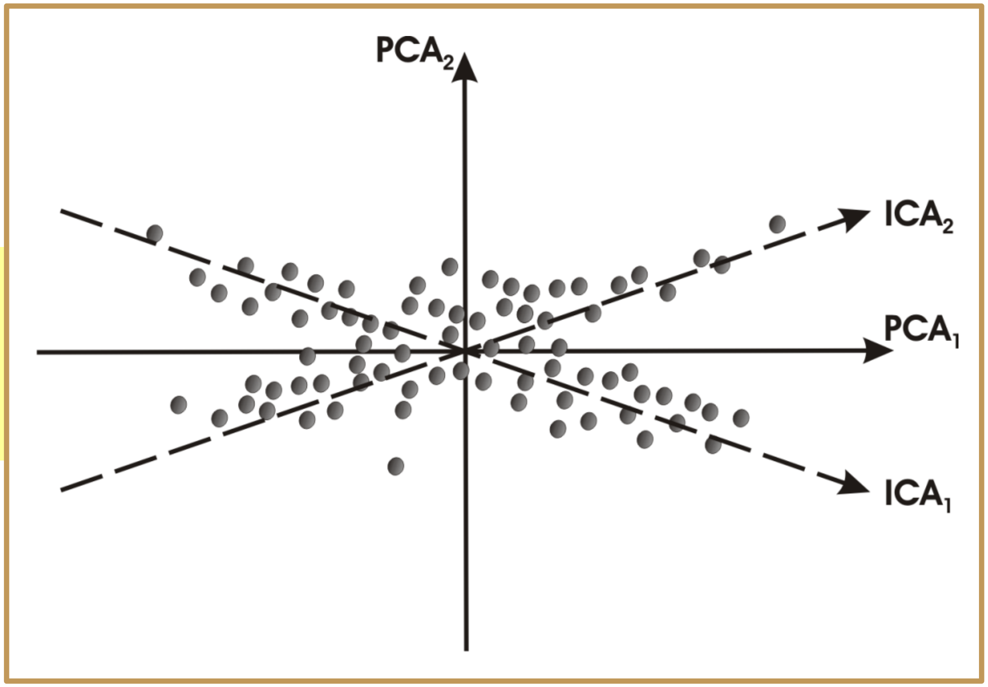


- The Independent Component Analysis (ICA) is based on higher order optimization to find independent (orthonormal) components for the face sub-spaces
- Better description of the inter-class variability

# Holistic face recognition



ICA




**Orthonormality**
**vs**
**Maximal variance**

PCA

# Face recognition via Sparse Representations John Wright et al. PAMI 2009

- Automatic face recognition algorithm robust to occlusion, expressions and disguise.
  - Represent the test face as a *sparse linear combination* of the training faces.
  - Estimate the class of the test image from the sparse coefficients.
- Can identify and reject "non face" images.
- …Performance can be affected by illumination variations and mis-alignment.

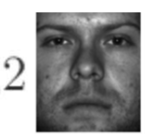
# Formulation

- Let $v_{ij}$ be the j$^{th}$ training image in the i$^{th}$ class.
- **A** is the dictionary of the training faces.

$$A \doteq [A_1, A_2, \ldots, A_k] = [\boldsymbol{v}_{1,1}, \boldsymbol{v}_{1,2}, \ldots, \boldsymbol{v}_{k,n_k}]$$

- The test image $\boldsymbol{y}$ is a linear combination of all instances from the correct face class.
  - if $\boldsymbol{y}$ belongs to the i$^{th}$ class:

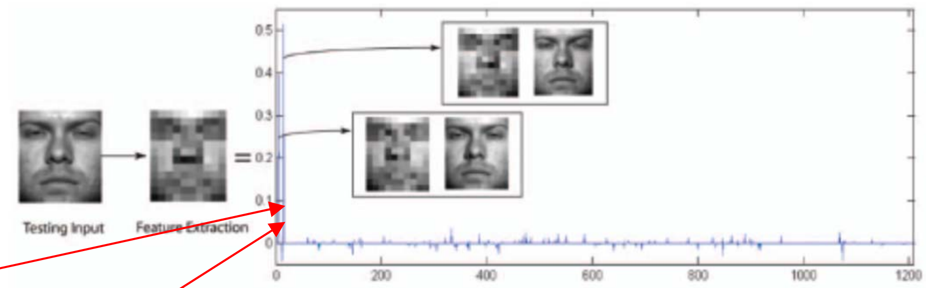
$$\boldsymbol{y} = \alpha_{i,1}\boldsymbol{v}_{i,1} + \alpha_{i,2}\boldsymbol{v}_{i,2} + \cdots + \alpha_{i,n_i}\boldsymbol{v}_{i,n_i}$$

 $= \alpha_{i,1}$  $+\alpha_{i,2}$  $+\cdots+\alpha_{i,n_i}$ 

**Dictionary matrix A**

# Formulation

- This can be rewritten as:




$$y = Ax_0$$

$$x_0 = [0, \cdots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots, 0]^T$$

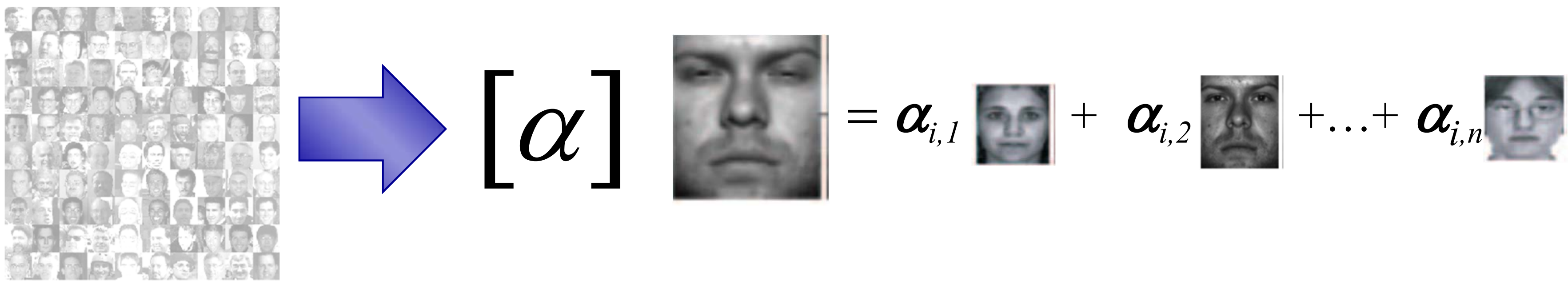
- As the number of classes is high, the coefficient vector is sparse.
- It can be recovered by solving the Basis Pursuit problem:

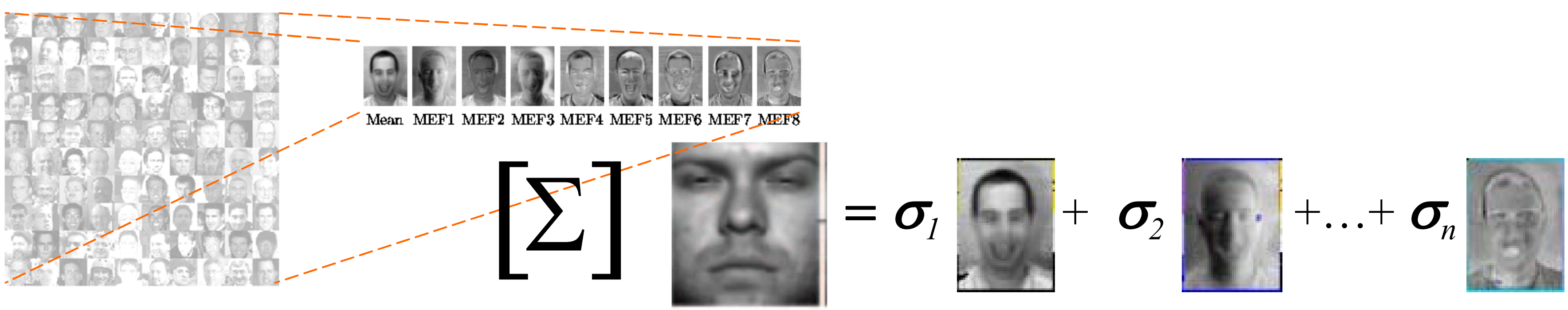
$$\hat{\boldsymbol{x}}_1 = \arg\min \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad A\boldsymbol{x} = \boldsymbol{y}.$$

- The non-zero coefficients in the sparse coefficient vector will correspond to the true class.

# Sparse Representation *vs* Principal Component Analysis

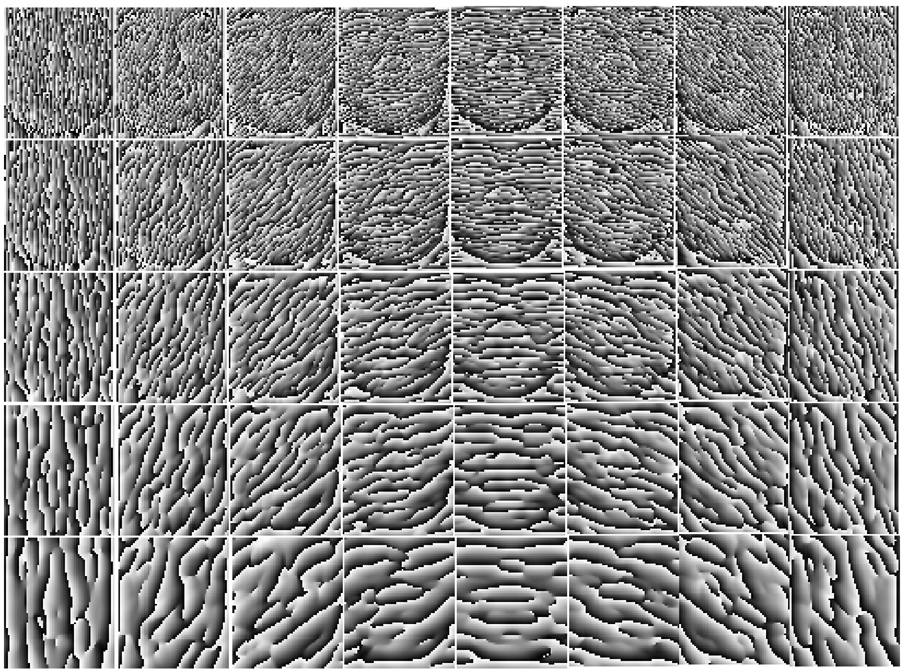
## Similar formulation, different objectives

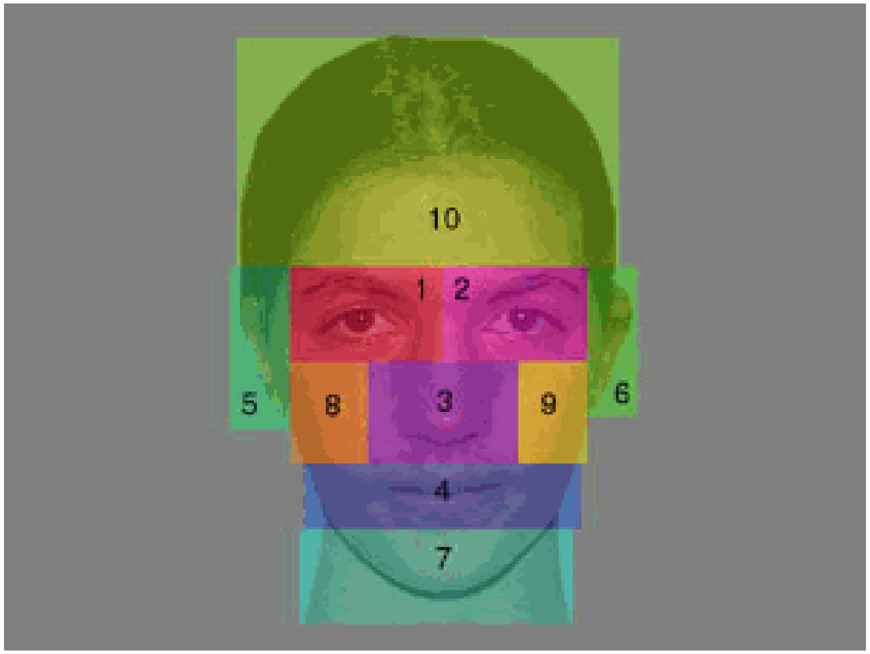


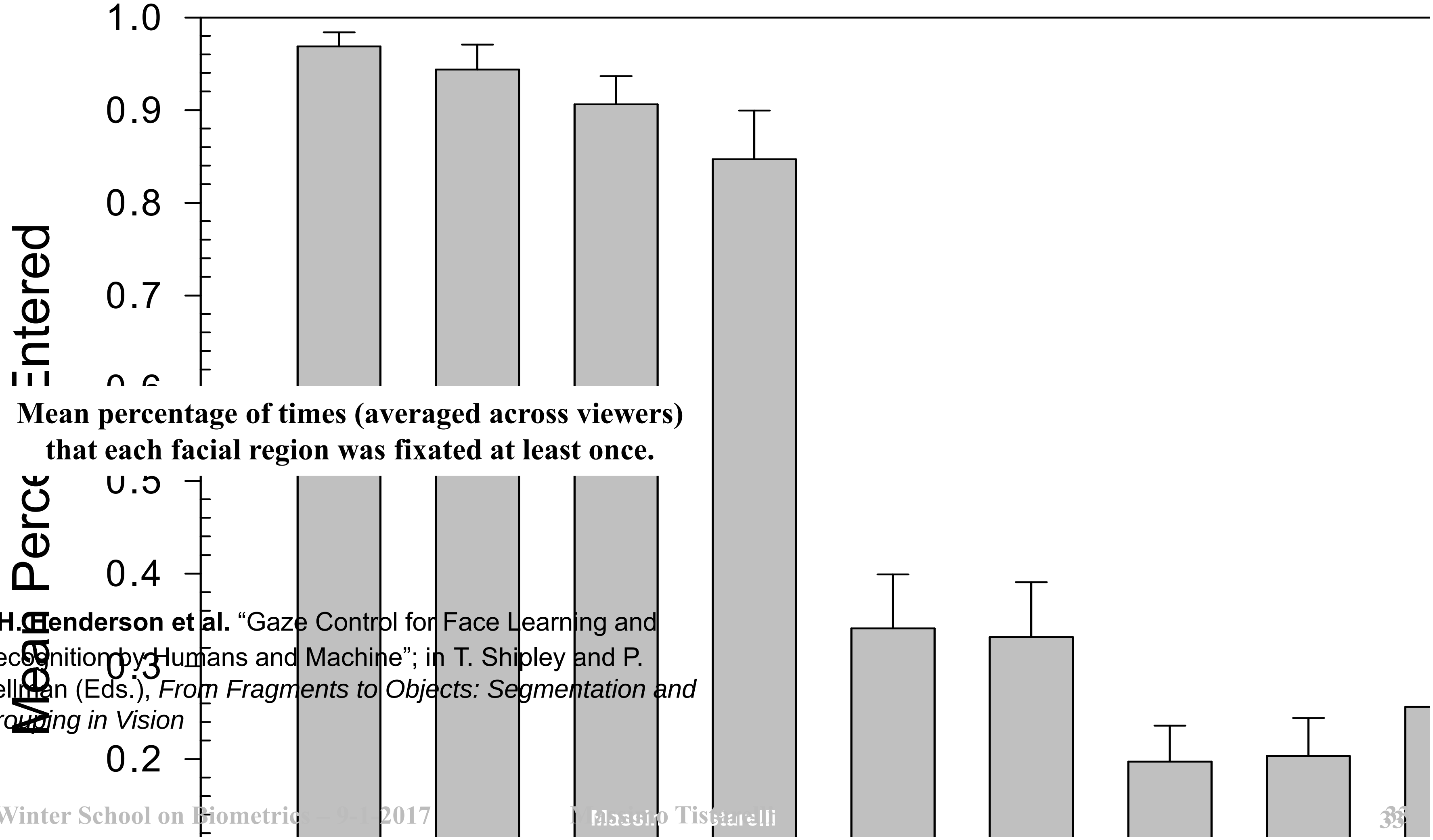
# Understanding *Facial features*

- Gray level oriented patterns/photometric properties
- Physical Landmarks





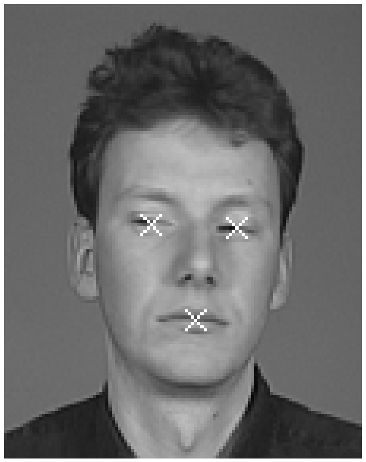
# How to define *facial landmarks*?

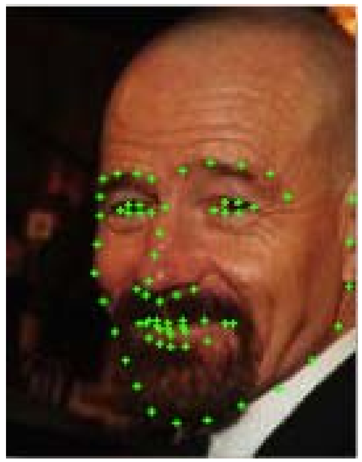


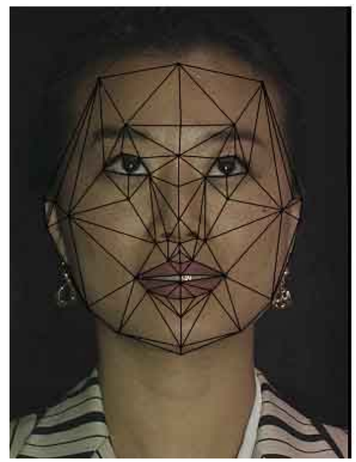

**Mean percentage of times (averaged across viewers) that each facial region was fixated at least once.**

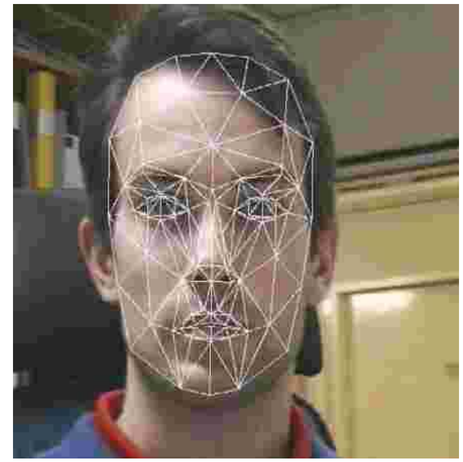
**J.H. Henderson et al.** "Gaze Control for Face Learning and Recognition by Humans and Machine"; in T. Shipley and P. Kellman (Eds.), *From Fragments to Objects: Segmentation and Grouping in Vision*

# Facial features as 2D/3D *landmarks*



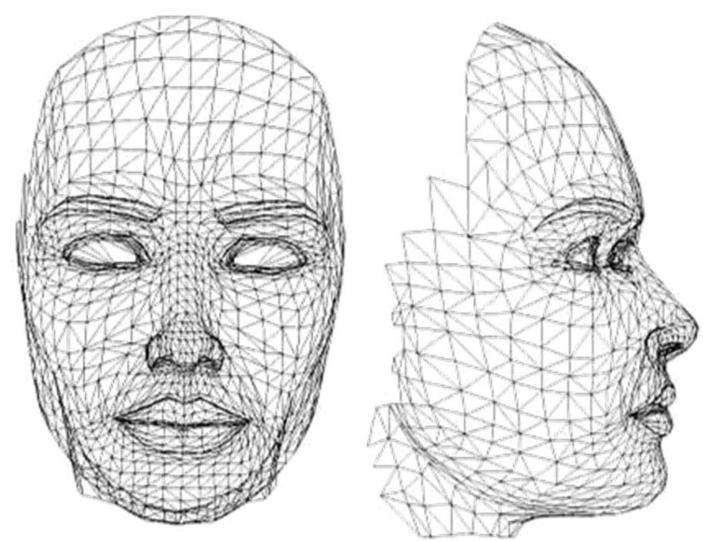




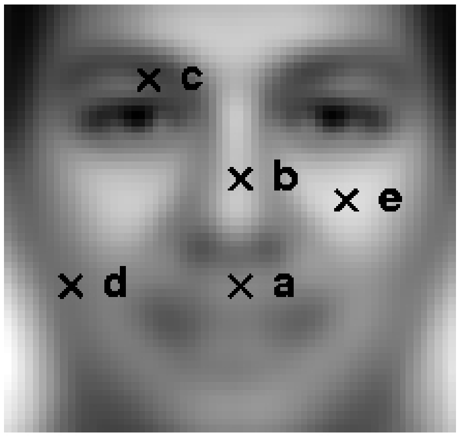


- 2D landmarks can be defined and tracked on face images

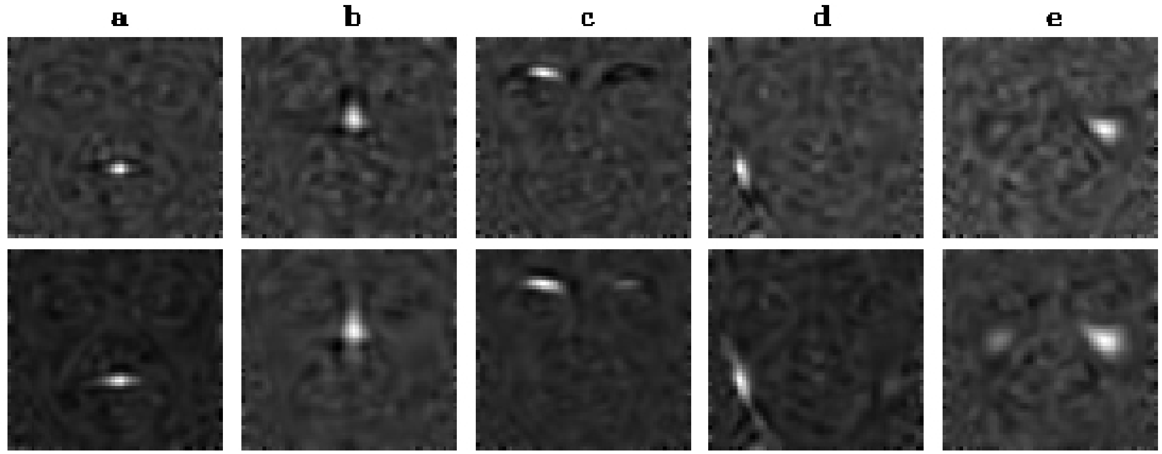
- Simple 2D vs complex 3D representations



# Facial features as 2D/3D *landmarks*




Marked average face image




- Five topographic kernels are shown in the top row
- Five corresponding residual correlations (response) in the bottom row.

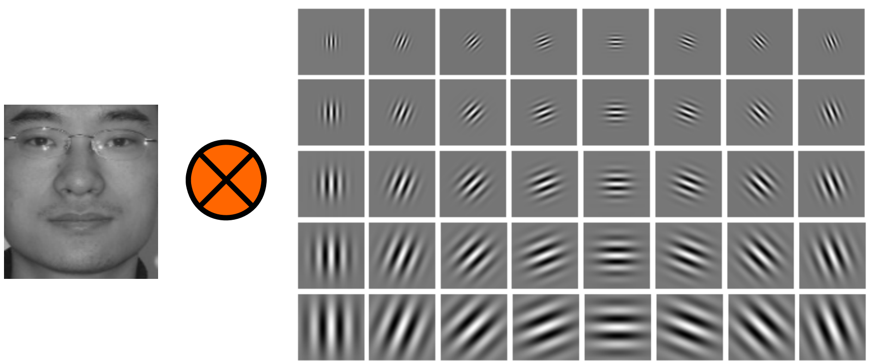
The LFA-based approach (*Local Feature Analyisis*) uses localized kernels, which are constructed from PCA-based eigenvectors, for extracting topographic facial features (e.g., eyebrows, cheek, mouth, etc.)

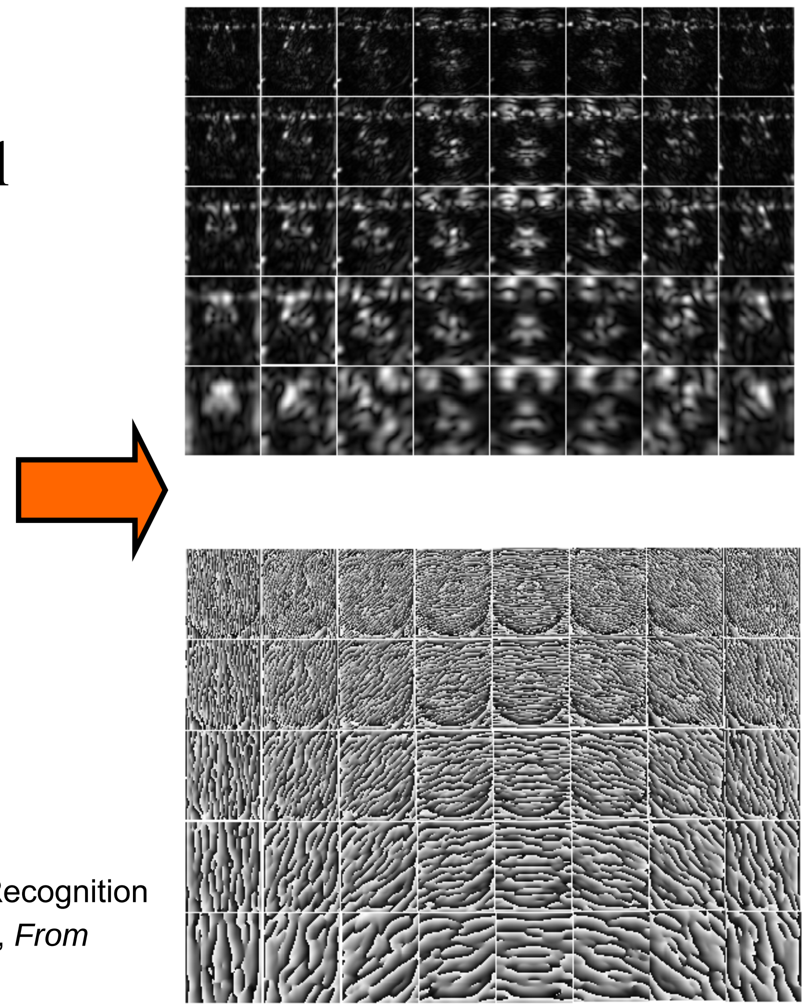

**Arca, Stefano, Paola Campadelli, and Raffaella Lanzarotti.** "A face recognition system based on local feature analysis." In International Conference on Audio-and Video-based Biometric Person Authentication, pp. 182-189. Springer, 2003.

# Facial features as 2D *patterns*

## Gabor wavelets

- Provide a description of the local structure of the facial patterns



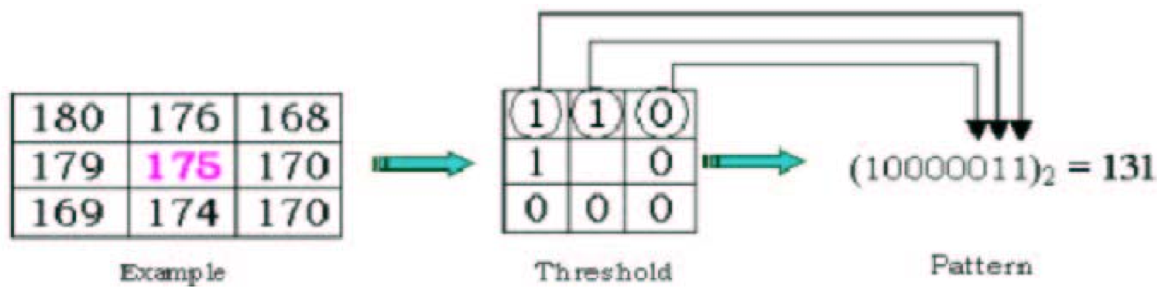


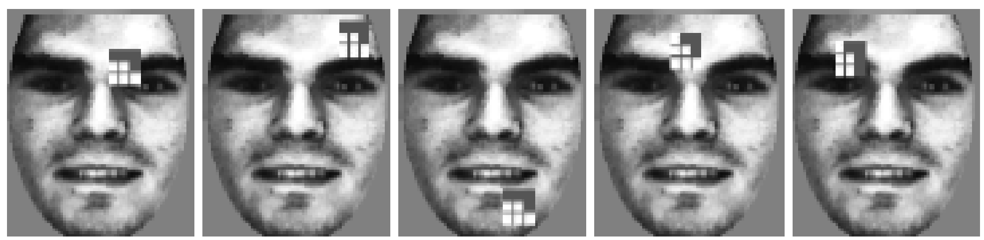
- Convolution with a bank of frequency-tuned filters

**J.H. Henderson et al.** "Gaze Control for Face Learning and Recognition by Humans and Machine"; in T. Shipley and P. Kellman (Eds.), *From Fragments to Objects: Segmentation and Grouping in Vision*

# Facial features as 2D *patterns*

## Local Binary Patterns (LBP)



Pixels are labeled by thresholding the 3x3neighbourhood with the center value and considering the result as a binary number.

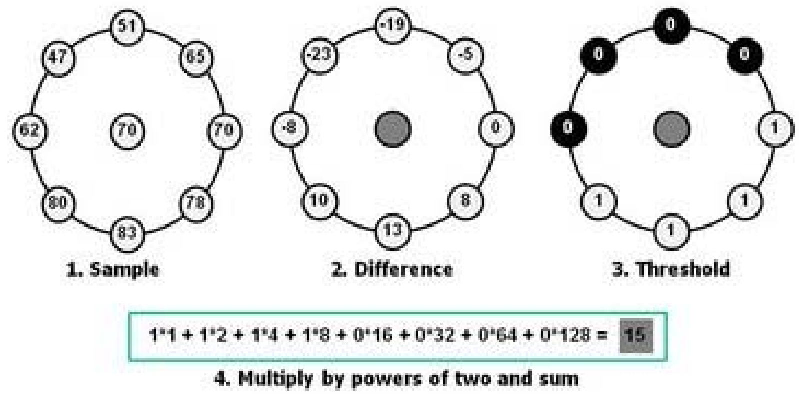
The histogram of the labels is used as a texture descriptor.

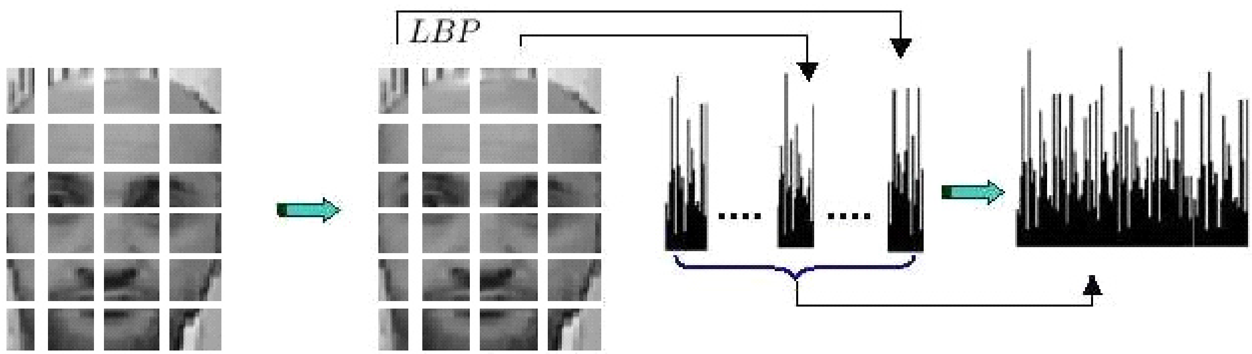
**T. Ahonen et al.** “Face Description with Local Binary Patterns: Application to Face Recognition”; in IEEE Trans. On PAMI 28(12):2037-2041.

# Facial features as 2D *patterns*

The value of the LBP code of a pixel $(x_c, y_c)$ is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad s(x) = \begin{cases} 1, if\ x \geq 0; \\ 0, otherwise. \end{cases}$$




A face image (144x112 pixels)

The image is divided into 24 blocks of 24*28 pixels

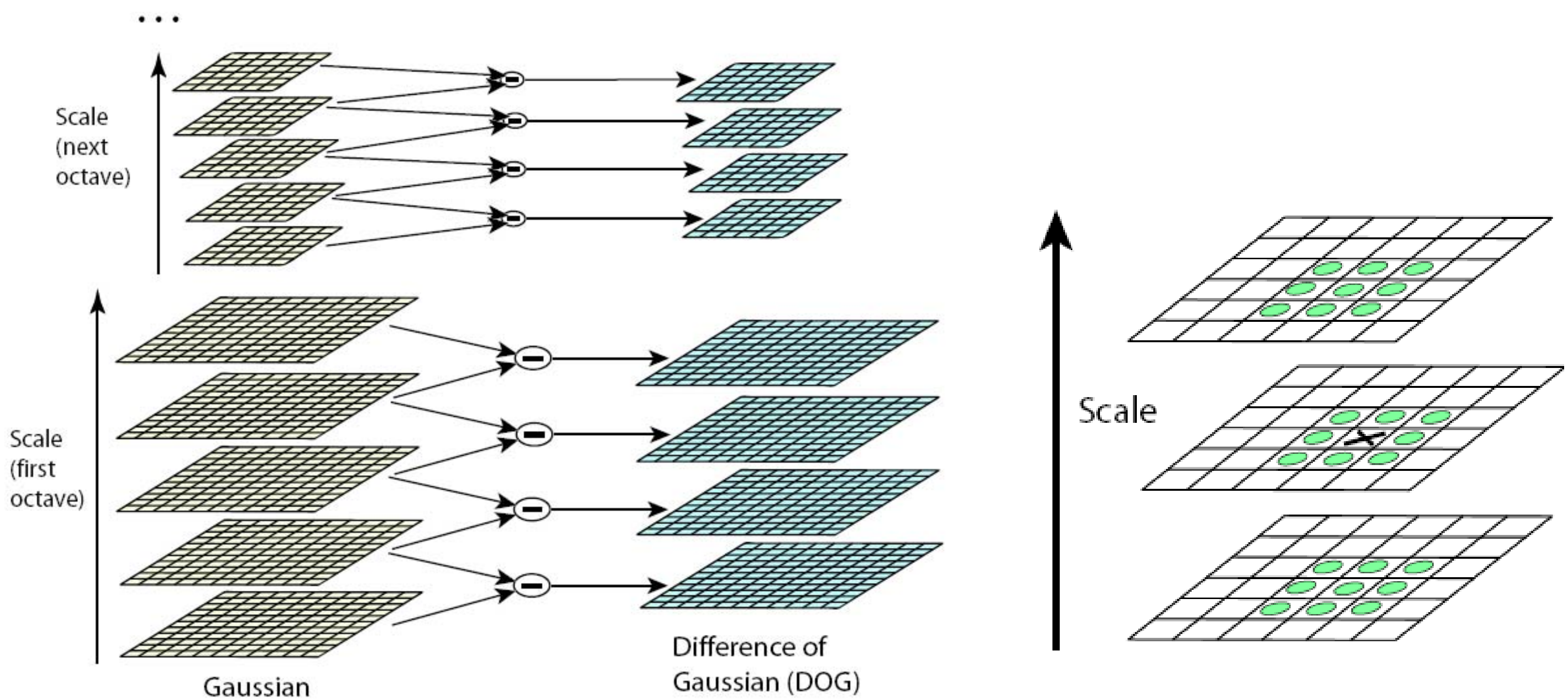
LBP histogram from each block

Feature histogram

# Scale Invariant Features

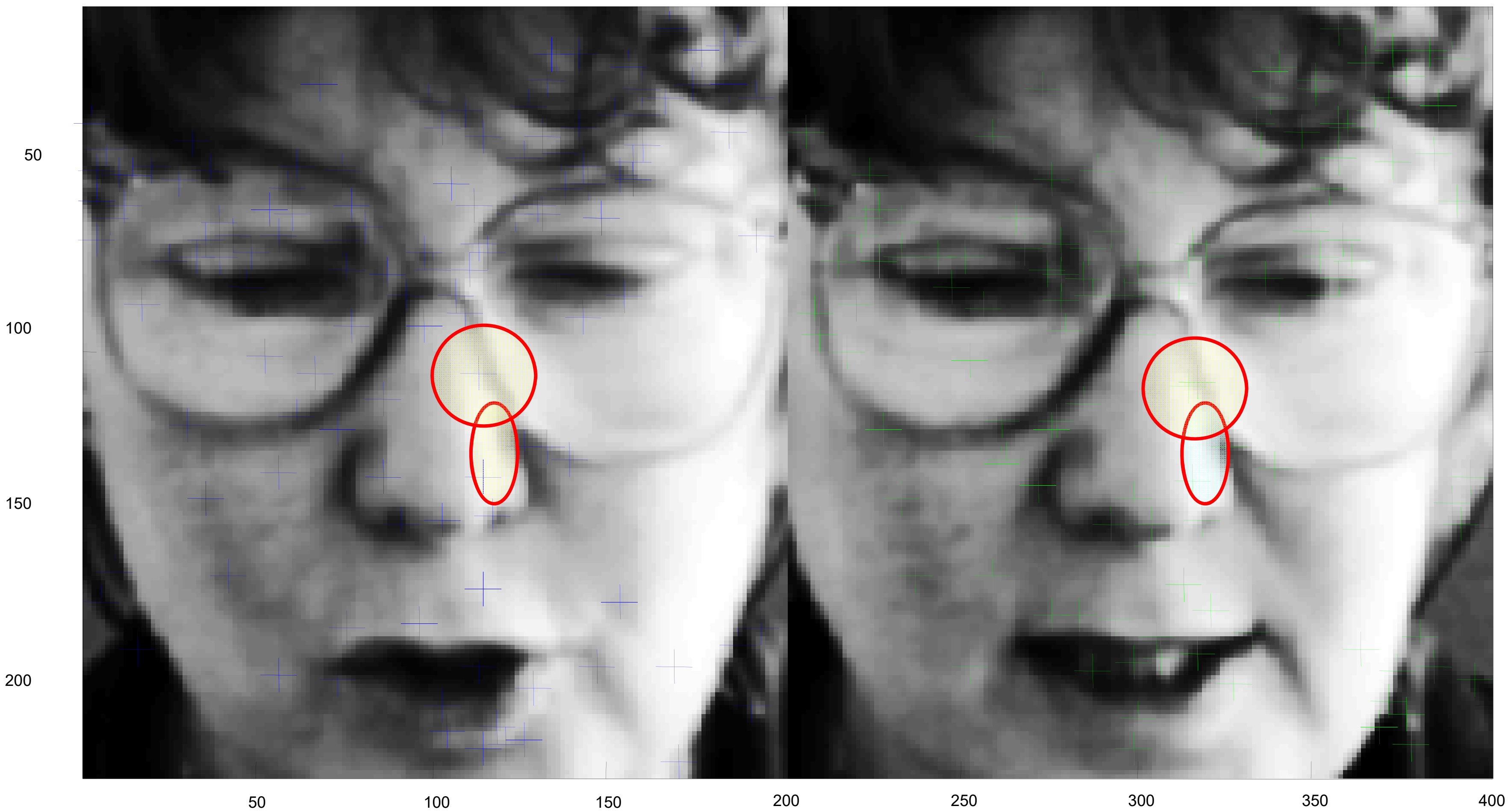
$$\mathbf{D}(x, y, \sigma, k) = (\mathbf{G}(x, y, k\sigma) - \mathbf{G}(x, y, \sigma)) * \mathbf{I}(x, y)$$

$$\mathbf{D}(x, y, \sigma, k) = \mathbf{L}(x, y, k\sigma) - \mathbf{L}(x, y, \sigma)$$

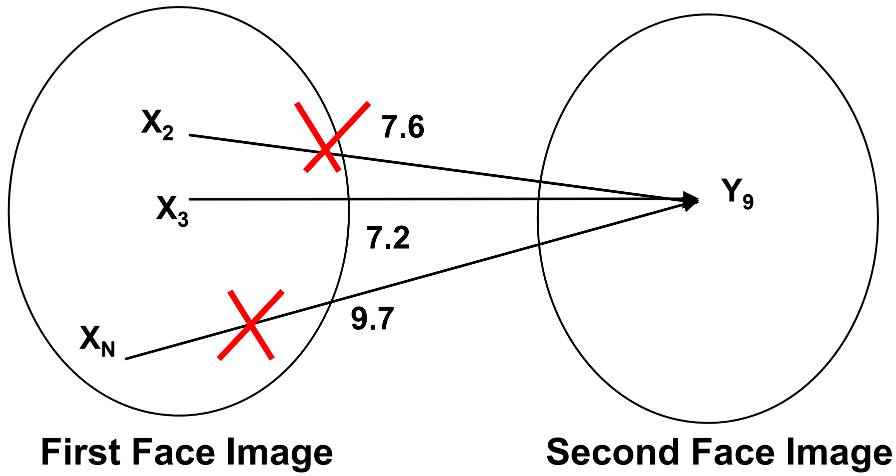


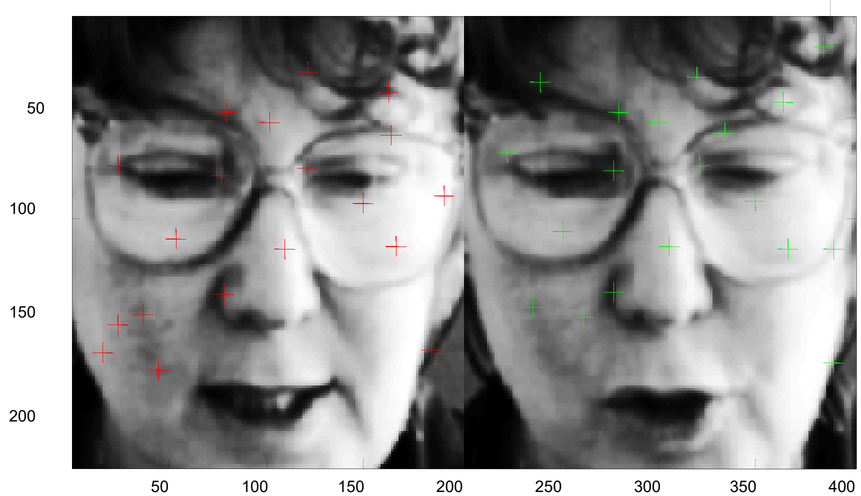

***G. Lowe**, "Object recognition from local scale invariant features", International Conference on Computer Vision , 1999.*

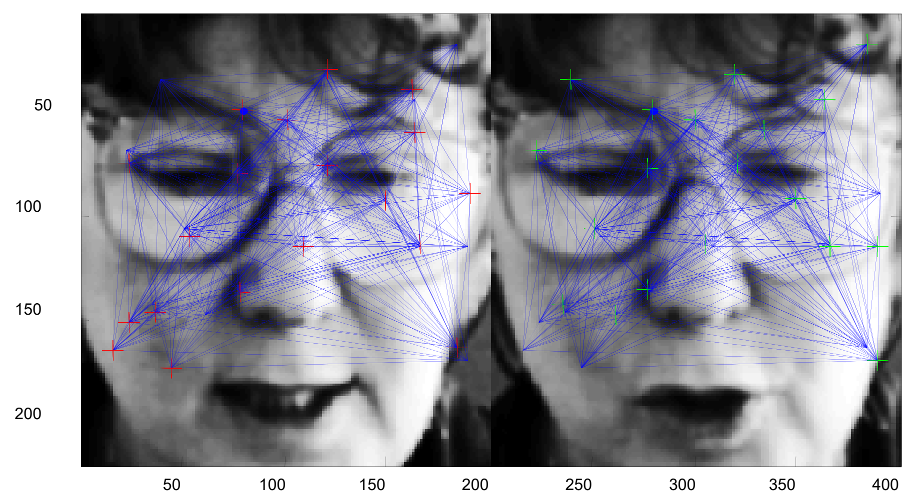
# Scale Invariant Features



# Scale Invariant Features



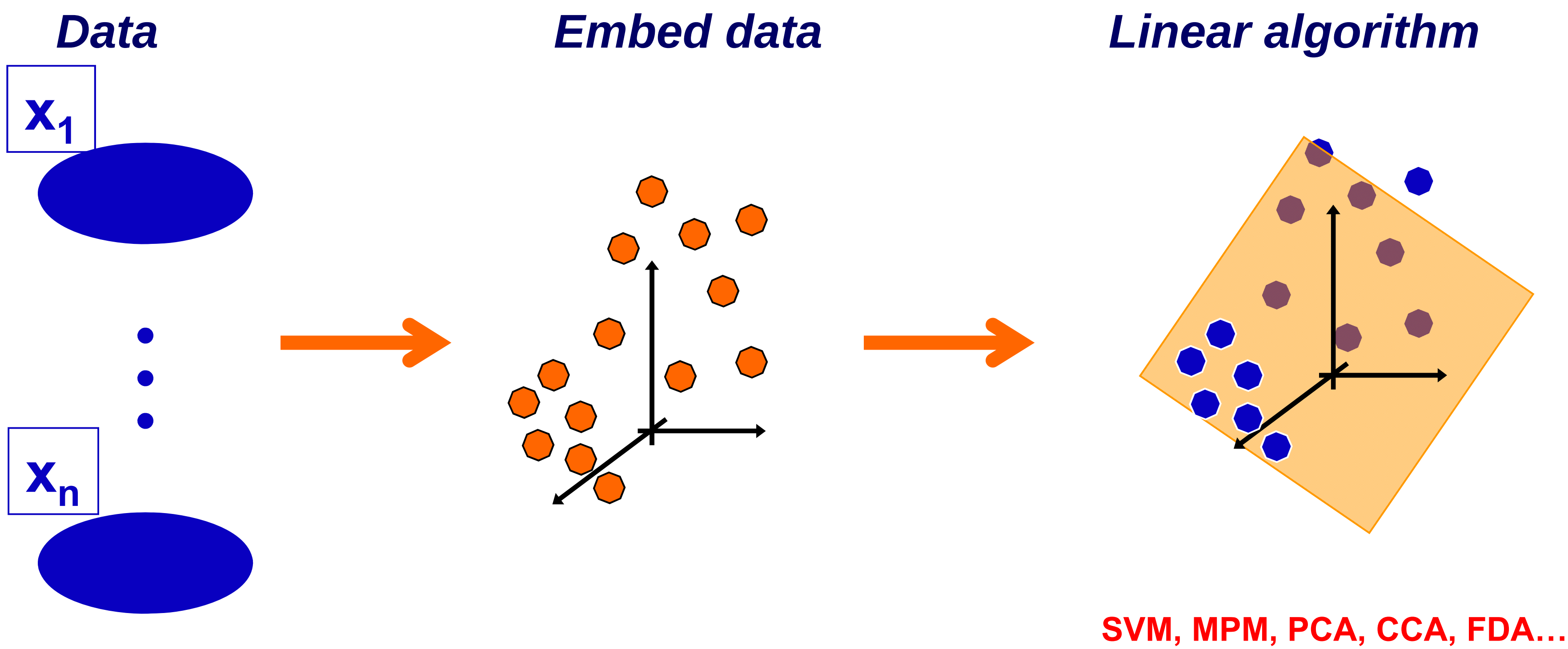



# Kernel methods

- K-PCA; K-ICA; K-LDA… (B. Schölkopf et al. 1998)
- Are all variations of existing face-space representations. The transformation to the lower space is mediated by a kernel function such as Gaussian, polinomial, sigmoid and Radial Basis Functions
- More robust to noise and discretization
- Better separation of classes
- General *Learning Theory*

# Kernel methods

Vision Lab




**If data is described by numerical vectors: embedding ~ (non-linear) transformation**

# Support Vector Machines

## Support Vector Machines are <u>binary</u> classifiers

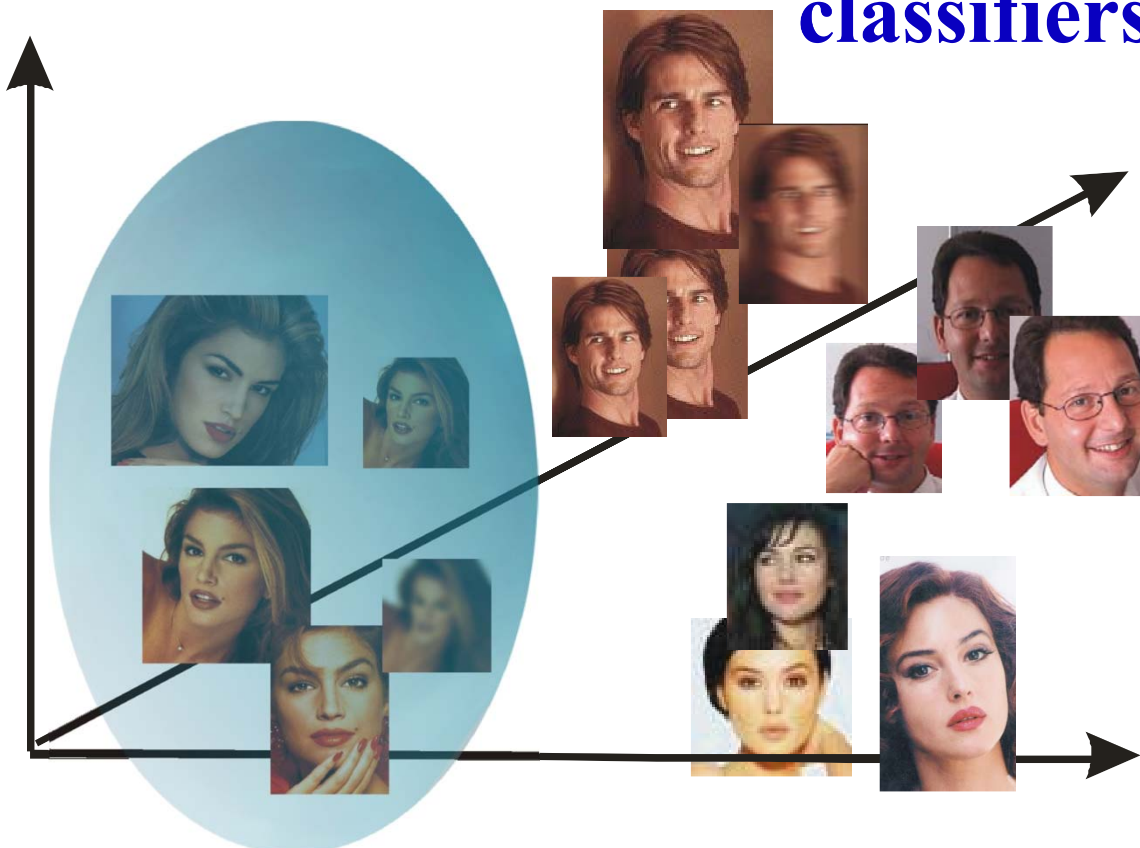
Class 1

Class 2

**V. Vapnik, S.E. Golowich, A.J. Smola**: Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. Neural Information Processing Systems 1996: 281-287

# One-Class
# Support Vector Machines

## One-Class Support Vector Machines are <u>unary</u> classifiers



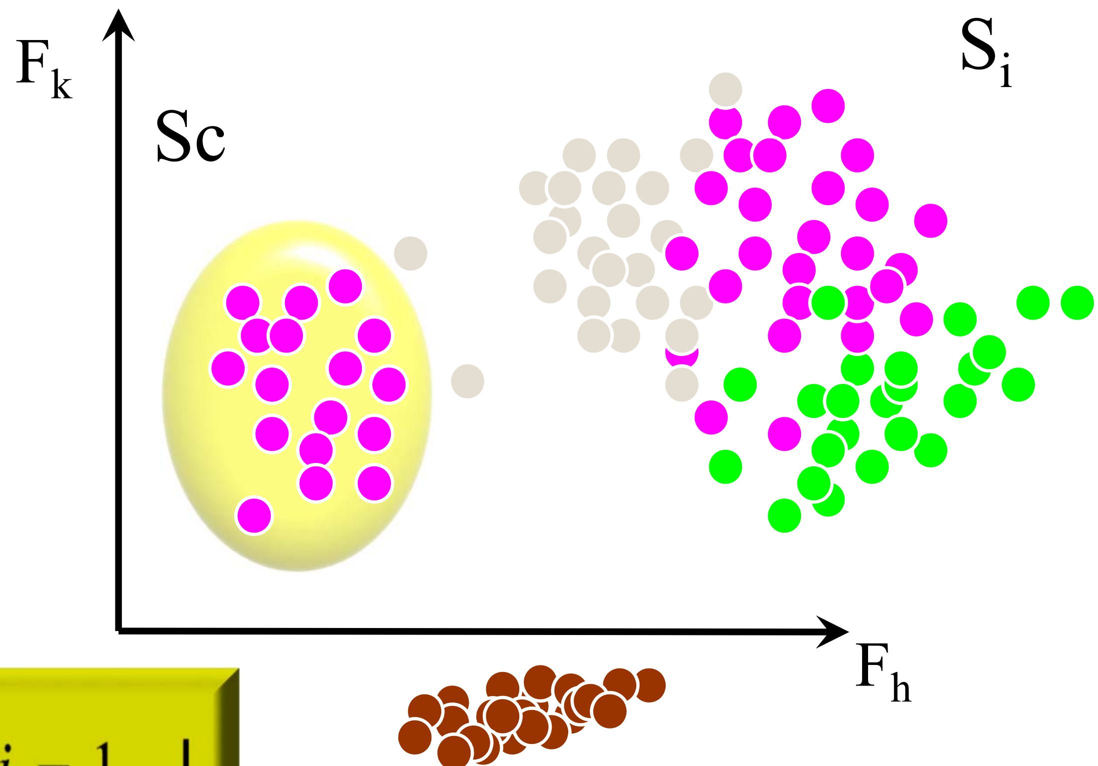
Class 1

Impostors

**Ben-Hur, A., Horn, D., Siegelmann, H., , Vapnik, V.:** « Support vector clustering ». Journal of Machine Learning Research **2** (2001) 125–137

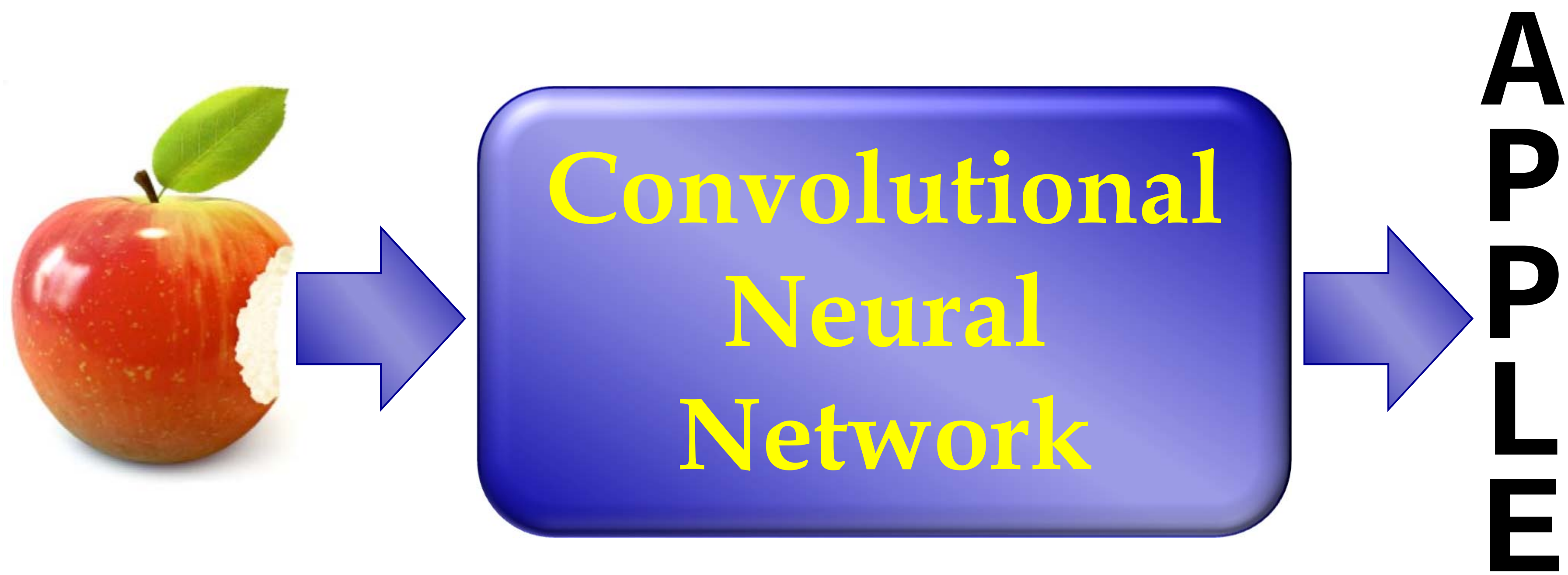
# One-Class Support Vector Machines

- The separating surface is a hyperspehere
- Selectivity can be adjusted by two parameters
- No need for direct “impostor” training

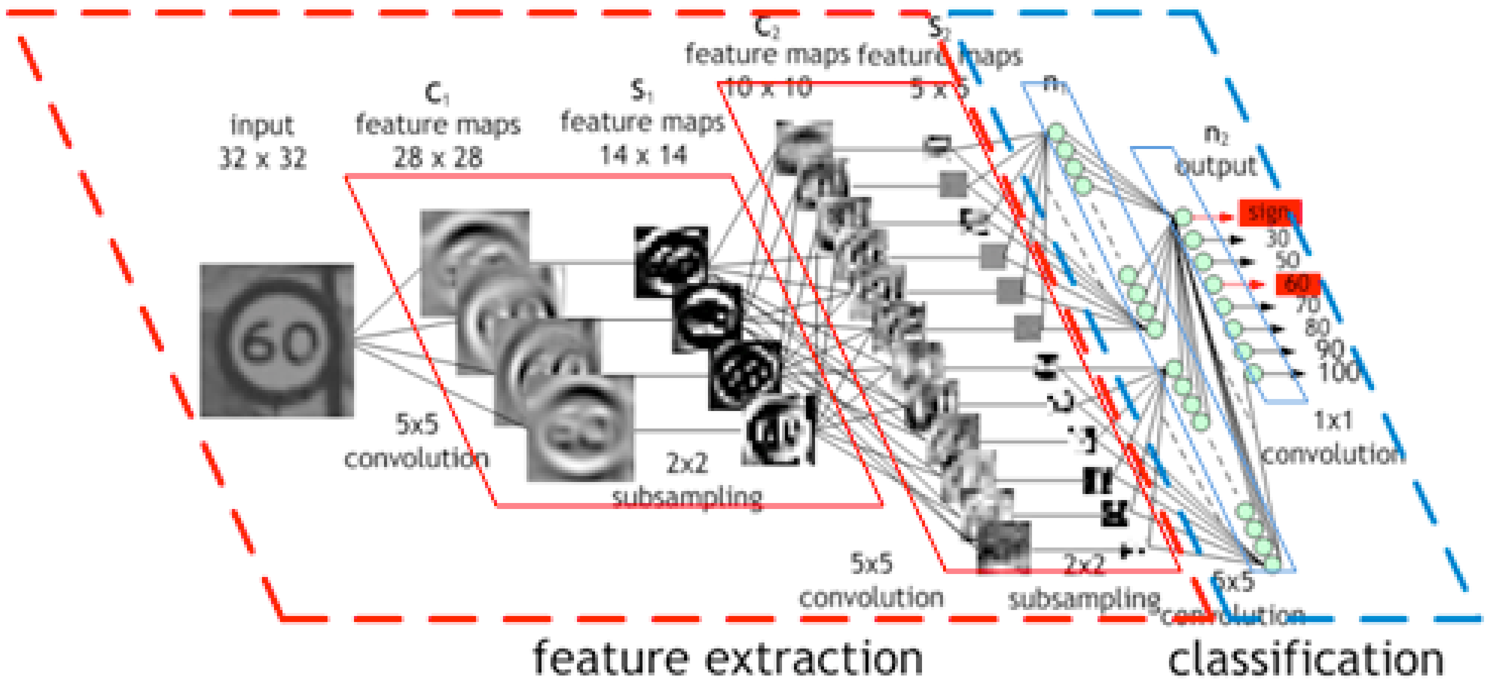



$$\|x_i - a\|^2 \leq R^2 \qquad i = 1 \ldots l$$

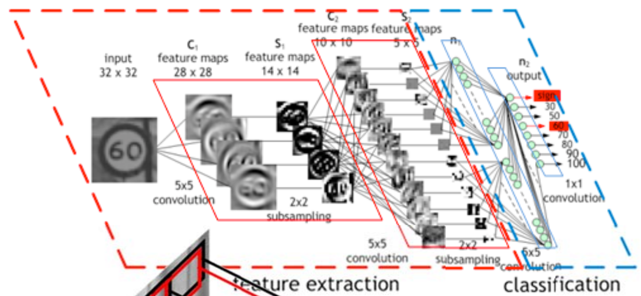
# Convolutional Neural Networks

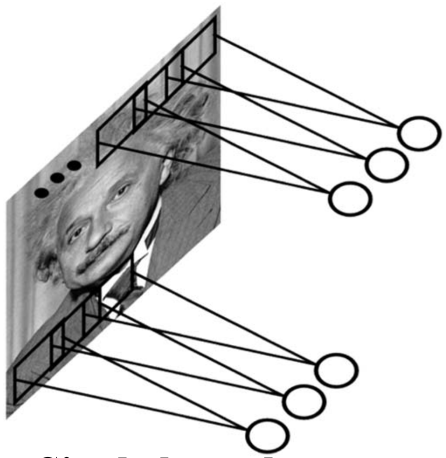


# Convolutional Neural Networks

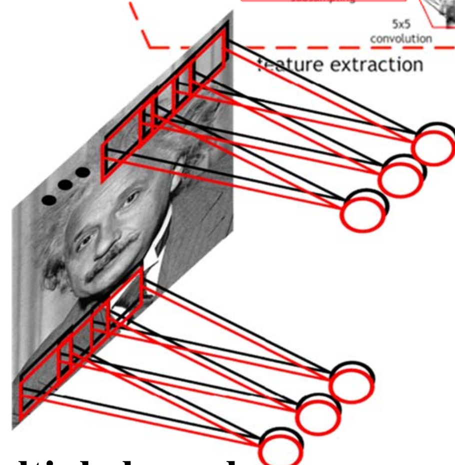


# Convolutional Neural Networks

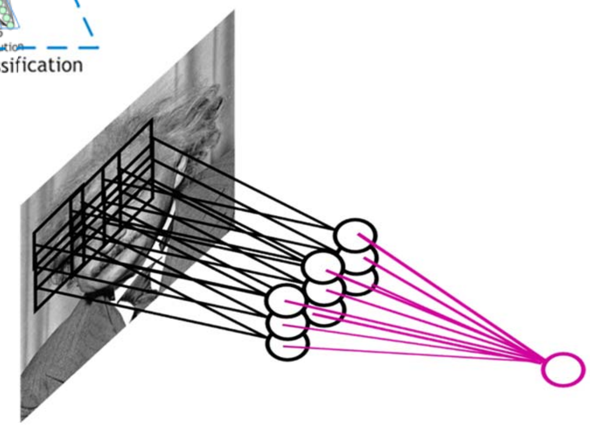


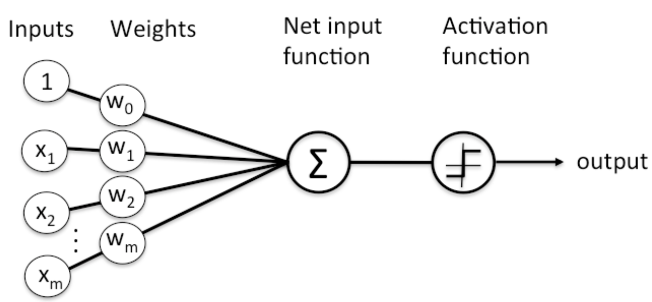
**Single kernel Convolution**



**Multiple kernels Convolution**



**Spatial Pooling**

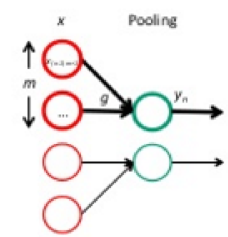



Let $m$ be the size of pooling region, $x$ be the input, and $y$ be the output of the pooling layer. subsample($f$, $g$)[$n$] denotes the $n$-th element of subsample($f$, $g$).

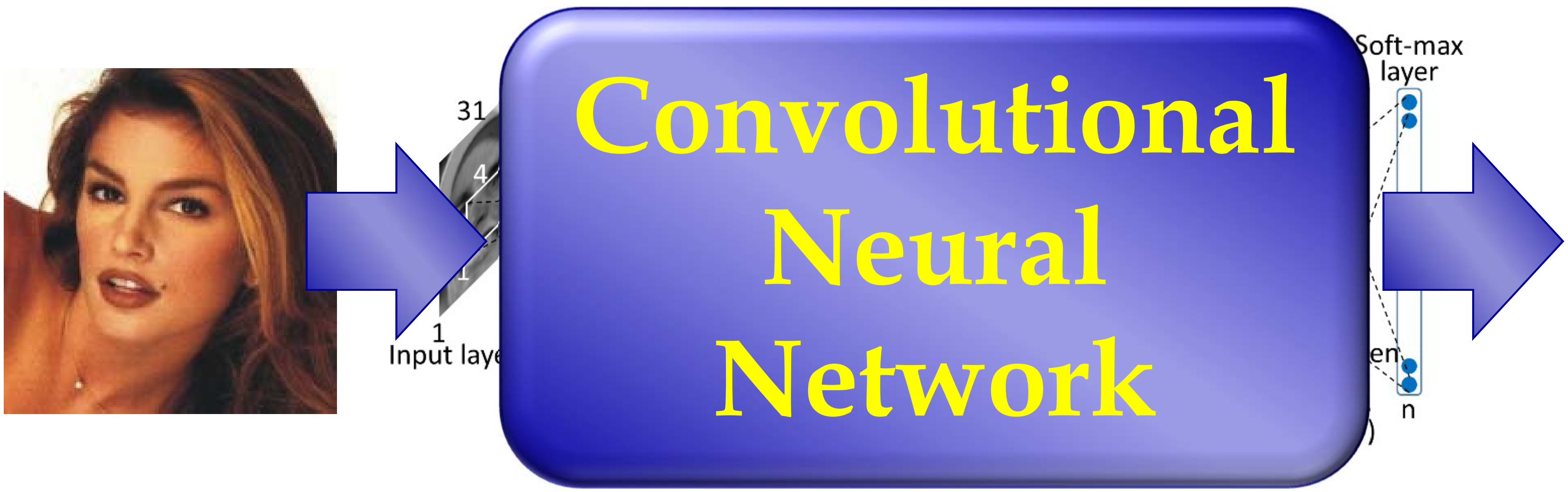
$$y_n = \text{subsample}(x,g)[n] = g\left(x_{(n-1)m+1:nm}\right)$$

$$y = \text{subsample}(x,g) = [y_n]$$

$$g(x) = \begin{cases} \dfrac{\sum_{k=1}^{m} x_k}{m}, \ \dfrac{\partial g}{\partial x} = \dfrac{1}{m} & \text{mean pooling} \\ \max(x), \ \dfrac{\partial g}{\partial x_i} = \begin{cases} 1 \text{ if } x_i = \max(x) \\ 0 \text{ otherwise} \end{cases} & \text{max pooling} \\ \|x\|_p = \left(\sum_{k=1}^{m} |x_k|^p\right)^{1/p}, \ \dfrac{\partial g}{\partial x_i} = \left(\sum_{k=1}^{m} |x_k|^p\right)^{1/p-1} |x_i|^{p-1} & \text{L}^p \text{ pooling} \\ \text{or any other differentiable } \mathbf{R}^m \to \mathbf{R} \text{ functions} \end{cases}$$

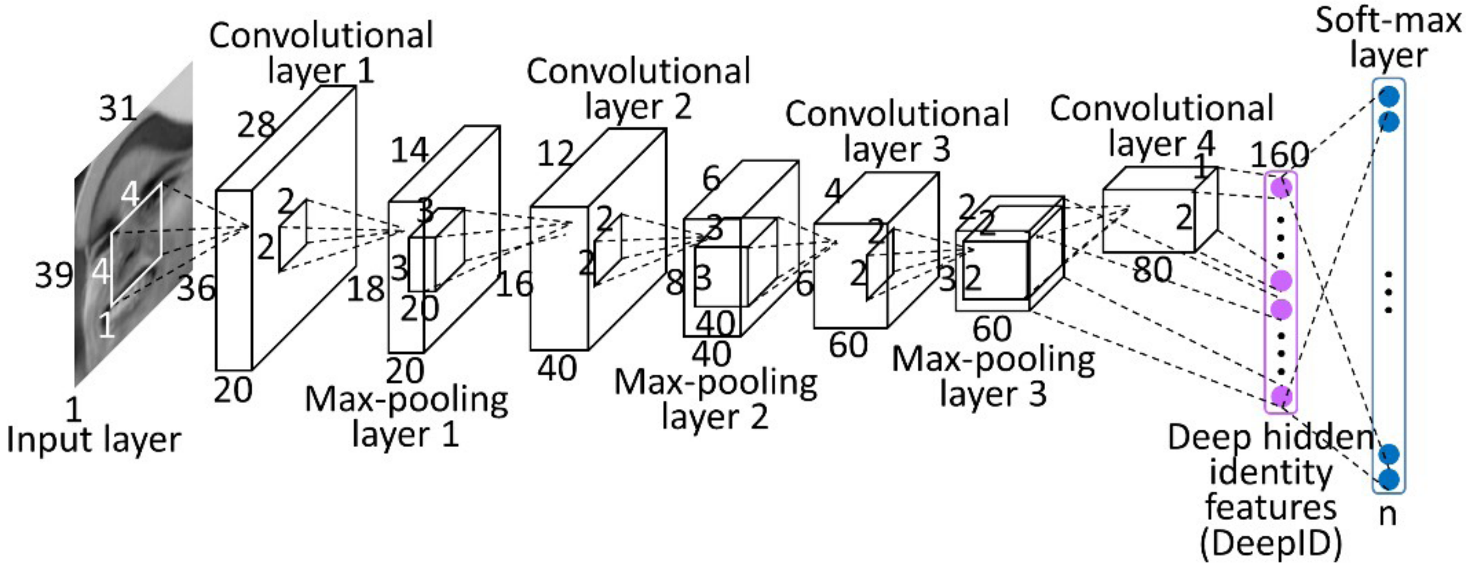


# Convolutional Neural Networks




Cindy crawford

# Convolutional Neural Networks



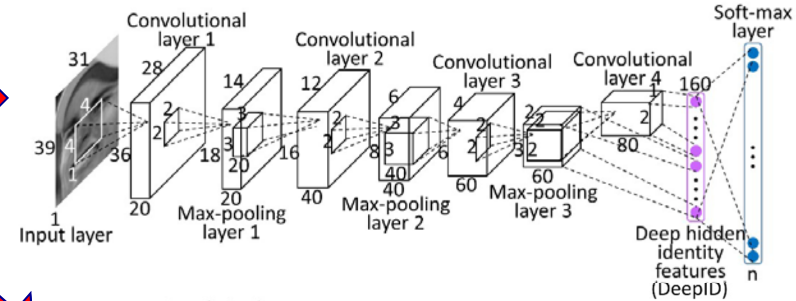

A deep CNN is used to extract a feature vector with relatively high dimension. The network can be supervised by multiclass loss and verification loss

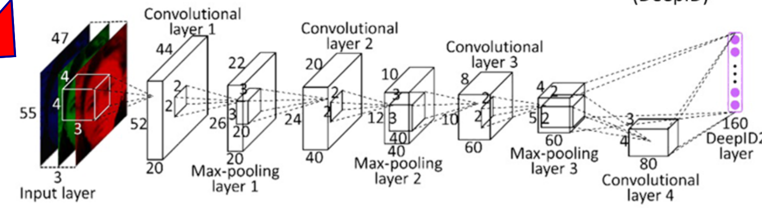
PCA, Joint Bayesian or metric-learning methods are used to learn a more efficient low dimensional representation

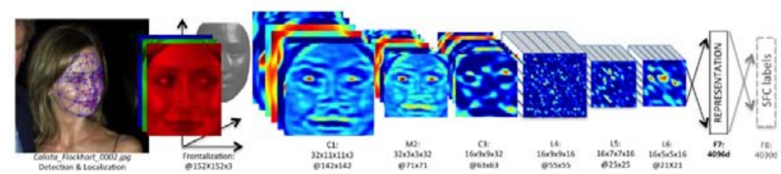
The amount of training data can range from **100K** up to **260M**

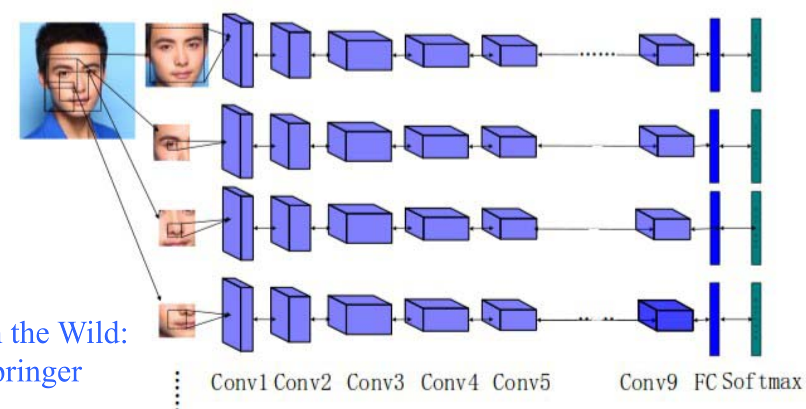
# Convolutional Neural Networks

- DeepID (Y. Sun, X.Wang, X. Tang – CVPR 2014)
- DeepID2 (Y. Sun, X.Wang, X. Tang - NIPS 2014)
- DeepID2+
- DeepID3
- DeepFace (Y. Taigman, M. Yang, M. Ranzato, L. Wolf – CVPR 2015)
- Face++
- FaceNet
- Baidu (J.Liu, Y.Deng, T.Bai, Z.Wei, C.Huang CVPR 2015)
- …**What's next?**



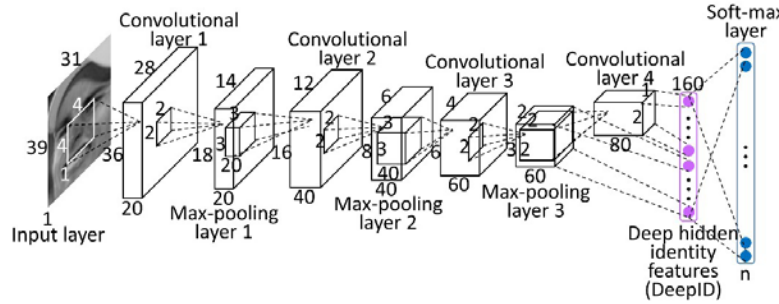






E. Learned-Miller, G. Huang, A. RoyChowdhury, H. Li, G. Hua, "Labeled Faces in the Wild: A Survey", Advances in Face Detection and Facial Image Analysis, pp 189-248, Springer 2016.

# Convolutional Neural Networks



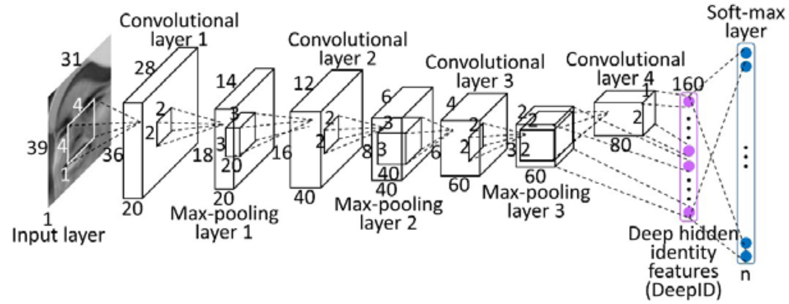

| Method | Net. Loss | Outside data | # models | Aligned | Verif. metric | Layers | Accu. |
|---|---|---|---|---|---|---|---|
| DeepFace [97] | ident. | 4M | 4 | 3D | wt. chi-sq. | 8 | 97.35±0.25 |
| Canon. view CNN [115] | ident. | 203K | 60 | 2D | Jt. Bayes | 7 | 96.45±0.25 |
| DeepID [92] | ident. | 203K | 60 | 2D | Jt. Bayes | 7 | 97.45±0.26 |
| DeepID2 [88] | ident. + verif. | 203K | 25 | 2D | Jt. Bayes | 7 | 99.15±0.13 |
| DeepID2+ [93] | ident. + verif. | 290K | 25 | 2D | Jt. Bayes | 7 | 99.47±0.12 |
| DeepID3 [89] | ident. + verif. | 290K | 25 | 2D | Jt. Bayes | 10-15 | 99.53±0.10 |
| Face++ [113] | ident. | 5M | 1 | 2D | L2 | 10 | 99.50±0.36 |
| FaceNet [82] | verif. (triplet) | 260M | 1 | no | L2 | 22 | 99.60±0.09 |
| Tencent [8] | - | 1M | 20 | yes | Jt. Bayes | 12 | 99.65±0.25 |

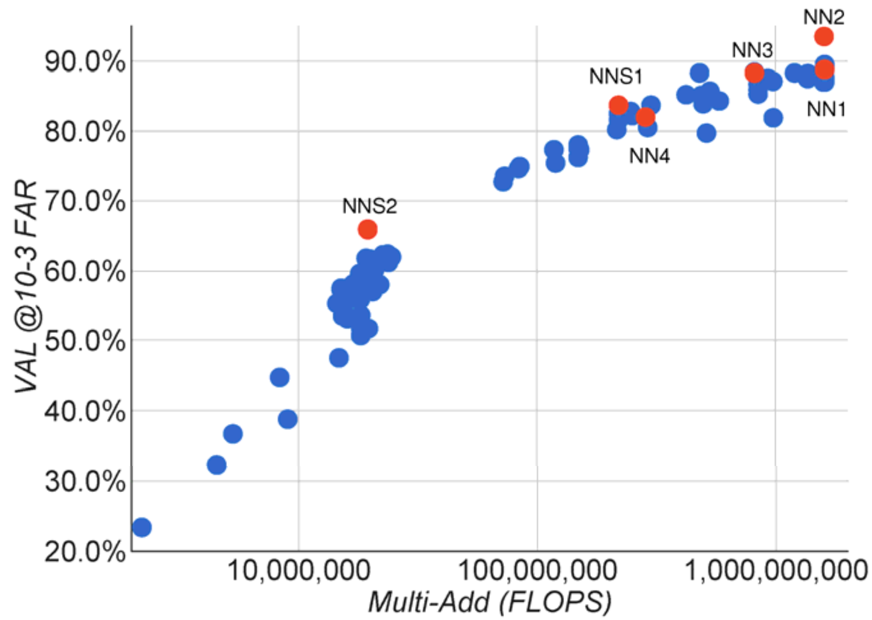
Figure 2. **Outline of the *DeepFace* architecture**. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. **The net includes more than 120 million parameters**, where more than 95% come from the local and fully connected layers.

# Convolutional Neural Networks

| layer | size-in | size-out | kernel | param | FLPS |
|---|---|---|---|---|---|
| conv1 | 220×220×3 | 110×110×64 | 7×7×3, 2 | 9K | 115M |
| pool1 | 110×110×64 | 55×55×64 | 3×3×64, 2 | 0 | |
| rnorm1 | 55×55×64 | 55×55×64 | | 0 | |
| conv2a | 55×55×64 | 55×55×64 | 1×1×64, 1 | 4K | 13M |
| conv2 | 55×55×64 | 55×55×192 | 3×3×64, 1 | 111K | 335M |
| rnorm2 | 55×55×192 | 55×55×192 | | 0 | |
| pool2 | 55×55×192 | 28×28×192 | 3×3×192, 2 | 0 | |
| conv3a | 28×28×192 | 28×28×192 | 1×1×192, 1 | 37K | 29M |
| conv3 | 28×28×192 | 28×28×384 | 3×3×192, 1 | 664K | 521M |
| pool3 | 28×28×384 | 14×14×384 | 3×3×384, 2 | 0 | |
| conv4a | 14×14×384 | 14×14×384 | 1×1×384, 1 | 148K | 29M |
| conv4 | 14×14×384 | 14×14×256 | 3×3×384, 1 | 885K | 173M |
| conv5a | 14×14×256 | 14×14×256 | 1×1×256, 1 | 66K | 13M |
| conv5 | 14×14×256 | 14×14×256 | 3×3×256, 1 | 590K | 116M |
| conv6a | 14×14×256 | 14×14×256 | 1×1×256, 1 | 66K | 13M |
| conv6 | 14×14×256 | 14×14×256 | 3×3×256, 1 | 590K | 116M |
| pool4 | 14×14×256 | 7×7×256 | 3×3×256, 2 | 0 | |
| concat | 7×7×256 | 7×7×256 | | 0 | |
| fc1 | 7×7×256 | 1×32×128 | maxout p=2 | 103M | 103M |
| fc2 | 1×32×128 | 1×32×128 | maxout p=2 | 34M | 34M |
| fc7128 | 1×32×128 | 1×1×128 | | 524K | 0.5M |
| L2 | 1×1×128 | 1×1×128 | | 0 | |
| total | | | | 140M | 1.6B |

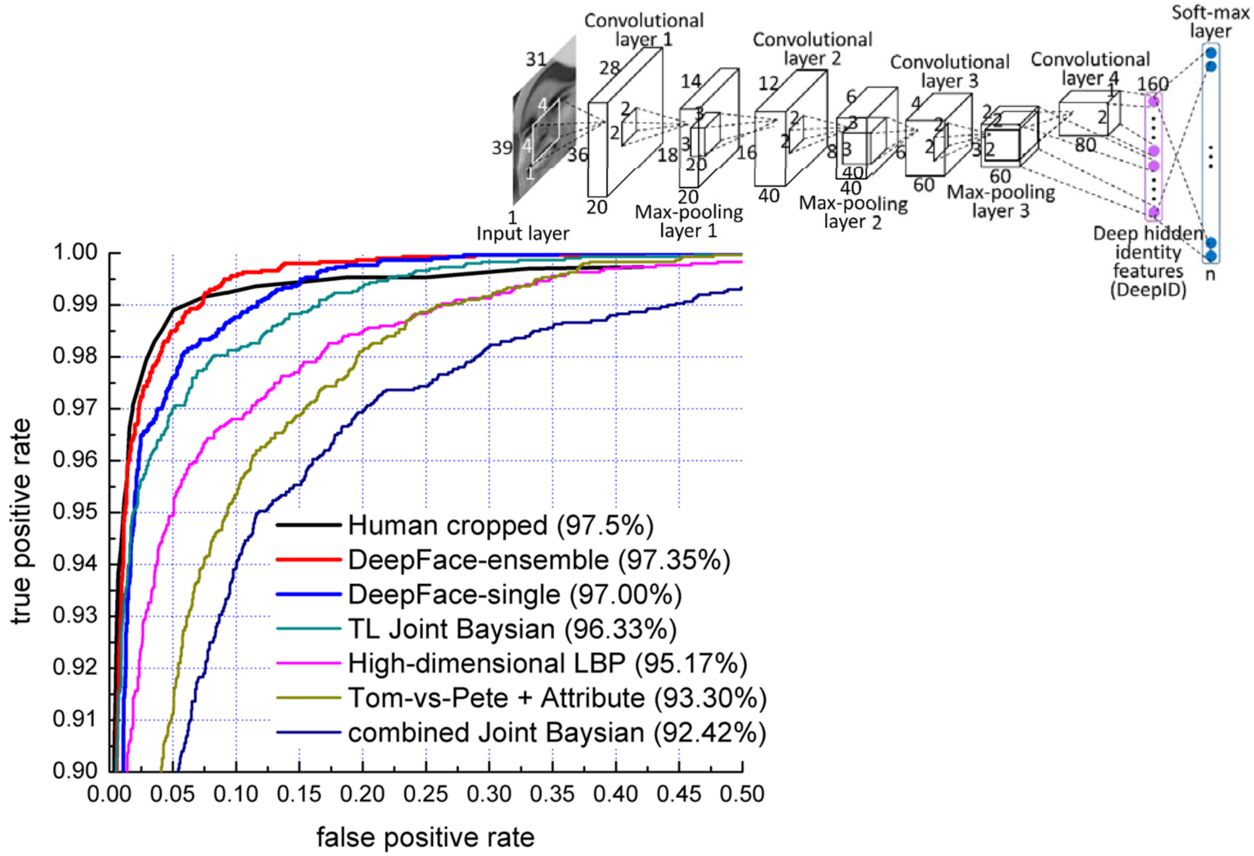
F. Schroff, D. Kalenichenko, J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering", CVPR 2015.




**FLOPS vs. Accuracy trade-off**. Shown is the trade-off between FLOPS and accuracy for a wide range of different model sizes and architectures. Highlighted are the four models that we focus on in our experiments.

# Convolutional Neural Networks

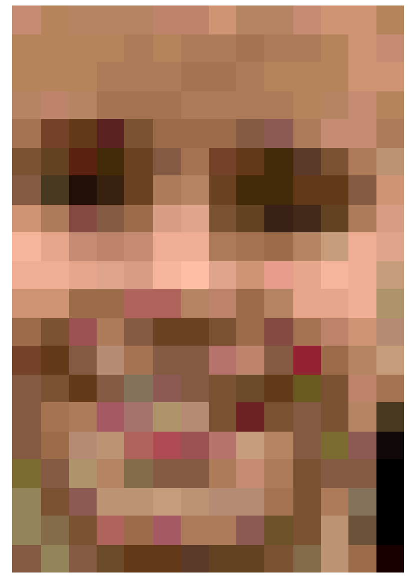



***“The performance of these systems is ironically matched by our present ignorance of why they work as well as they do.”***

F. Anselmi, L. Rosasco, C. Tan and T. Poggio - **Deep Convolutional Networks are Hierarchical Kernel Machines**

# Data dimensionality

## How many pixels to detect a face?

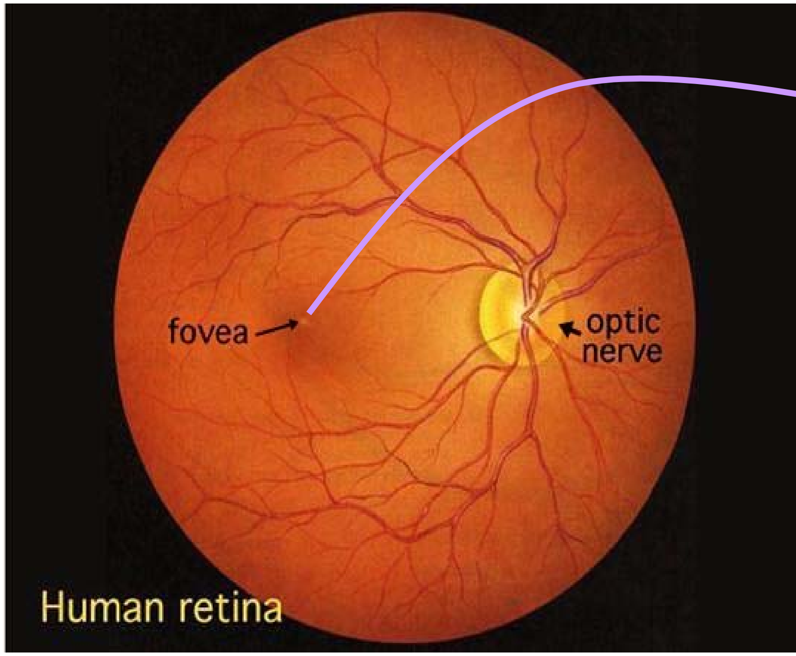


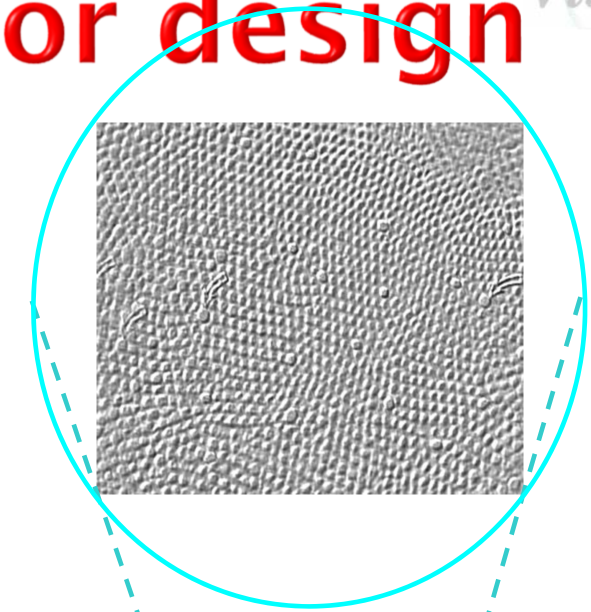
... Not many ... *(20x14)*

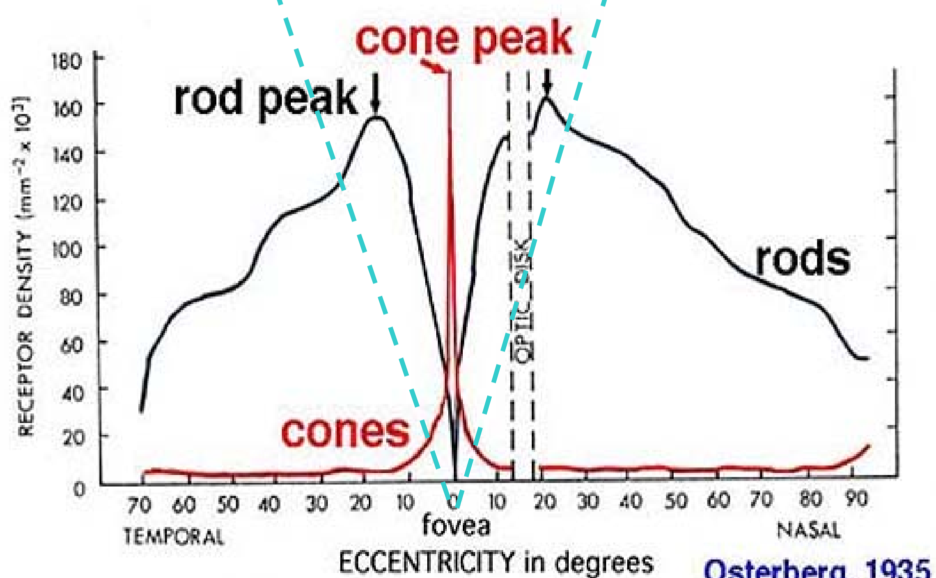
It's more a question of spatial distribution and …

proper frequency tuning

# A unique sensor design

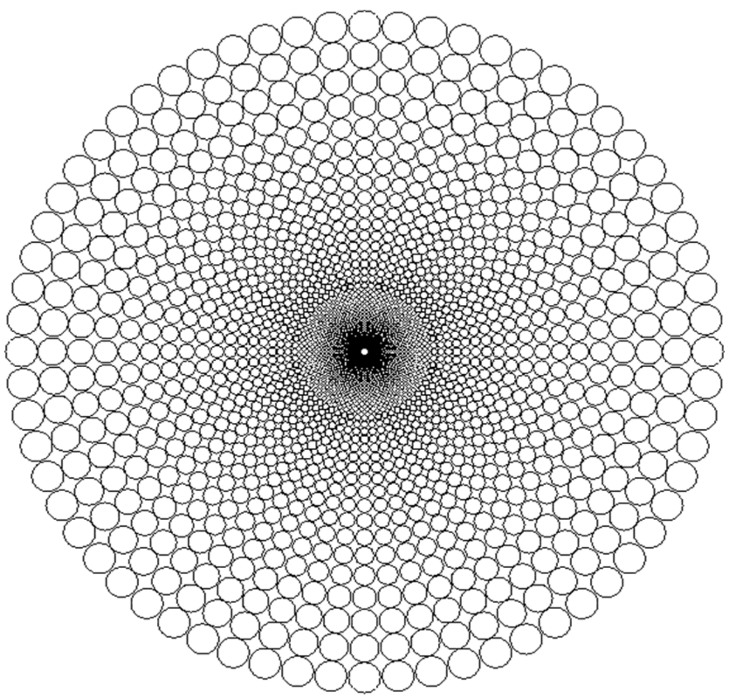


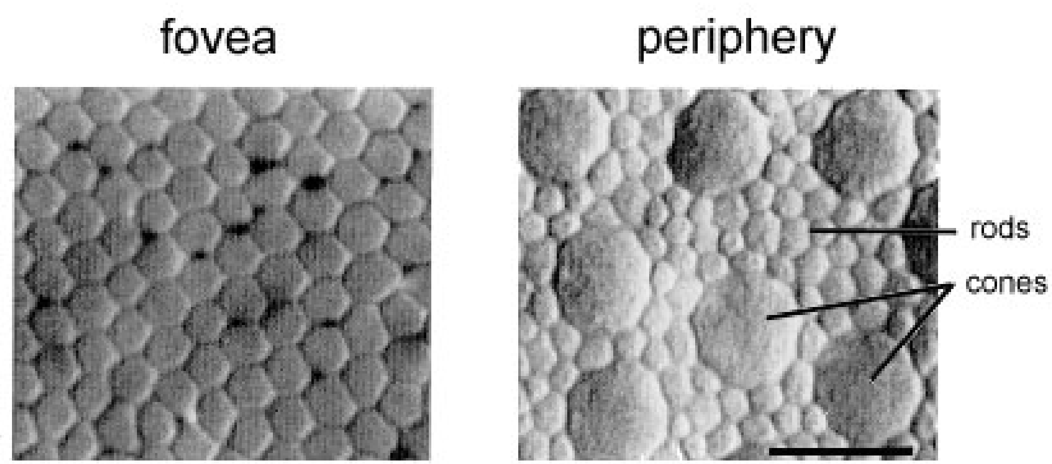





# The retina layout

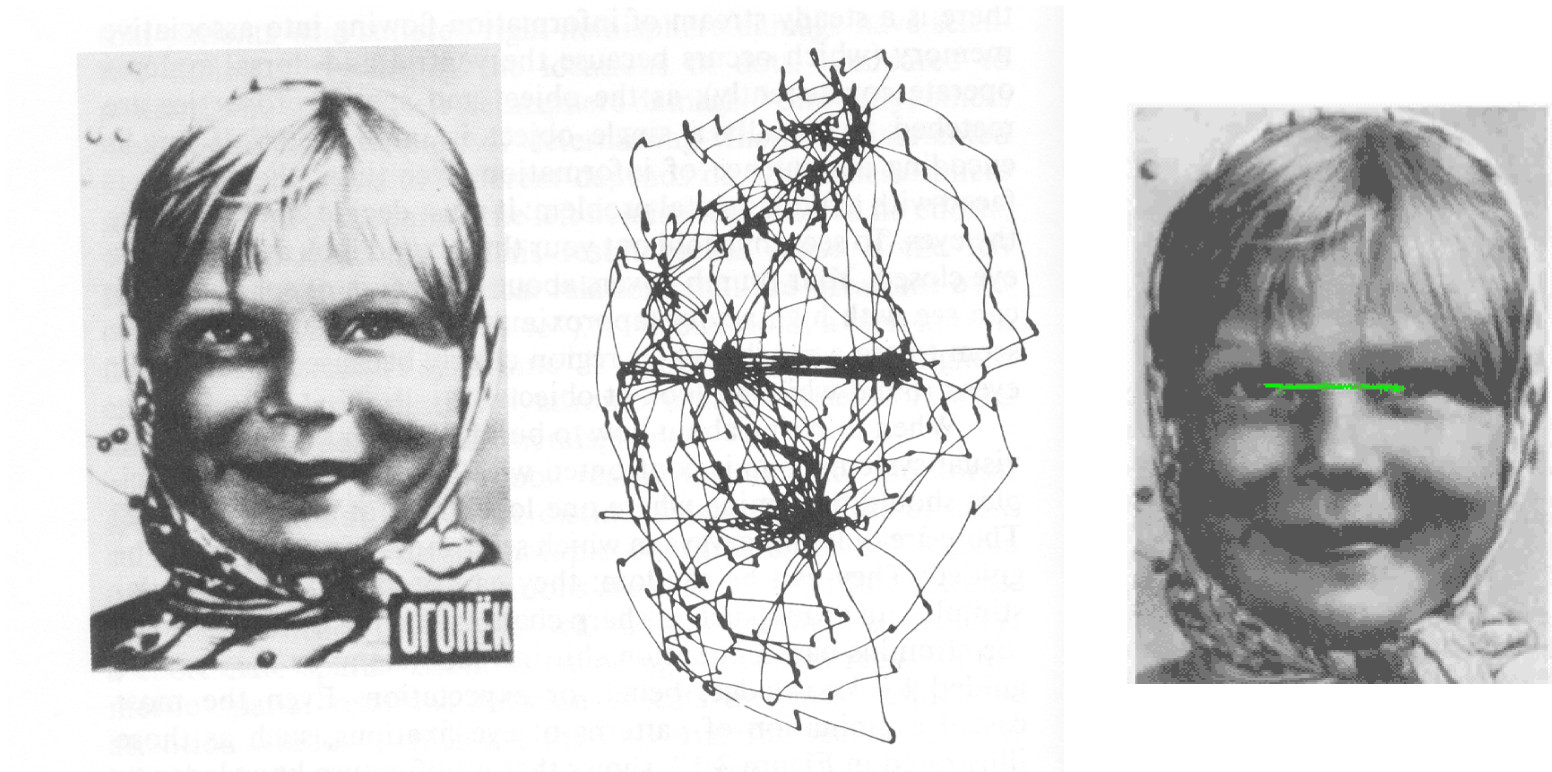
A good approximation of the cones density over the retina is given by the complex log-polar transform





# *Context* analysis: Visual attention

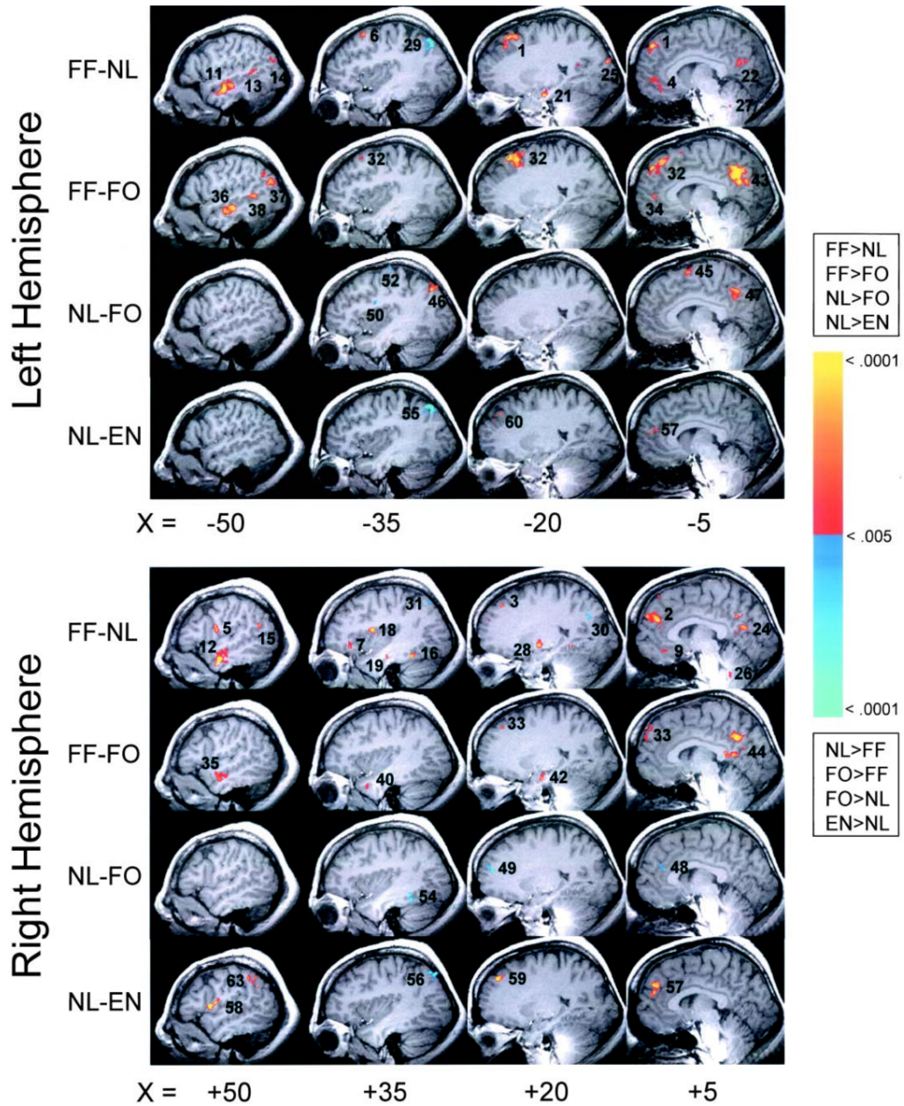
Vision Lab



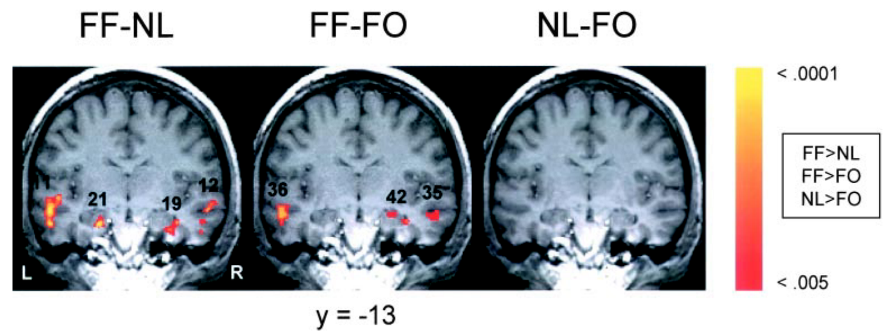

Eye movements while watching a girl’s face
(A.L. Yarbus, “Eye Movements and Vision”, Plenum Press, 1967)

# Functional Magnetic Resonance Imaging


# Brain activation – fMRI maps




Recognition of 50 Familiar Faces (**FF**) vs 50 Newly Learned Faces (**NL**) and compared to rejection of 50 Foil (**FO** -False Objective) faces. Encoding (**EN**) session for learning new faces.




C. L. Leveroni et al. **"Neural Systems Underlying the Recognition of Familiar and Newly Learned Faces",** The Journal of Neuroscience, January 15, 2000, *20*(2):878–886

*Figure 2.* Areas of significantly increased (*red–yellow* scale) and decreased (*blue–cyan* scale) MR signal intensity from *t* tests ($p < 0.005$) comparing the three conditions: FF minus NL, FF minus FO, and NL minus FO. Numbers below each image represent millimeters from the interhemispheric fissure (−, left; +, right). Numbers adjacent to activated foci correspond to location numbers (first column) of Tables 1, 2, and 3.

# Brain activation

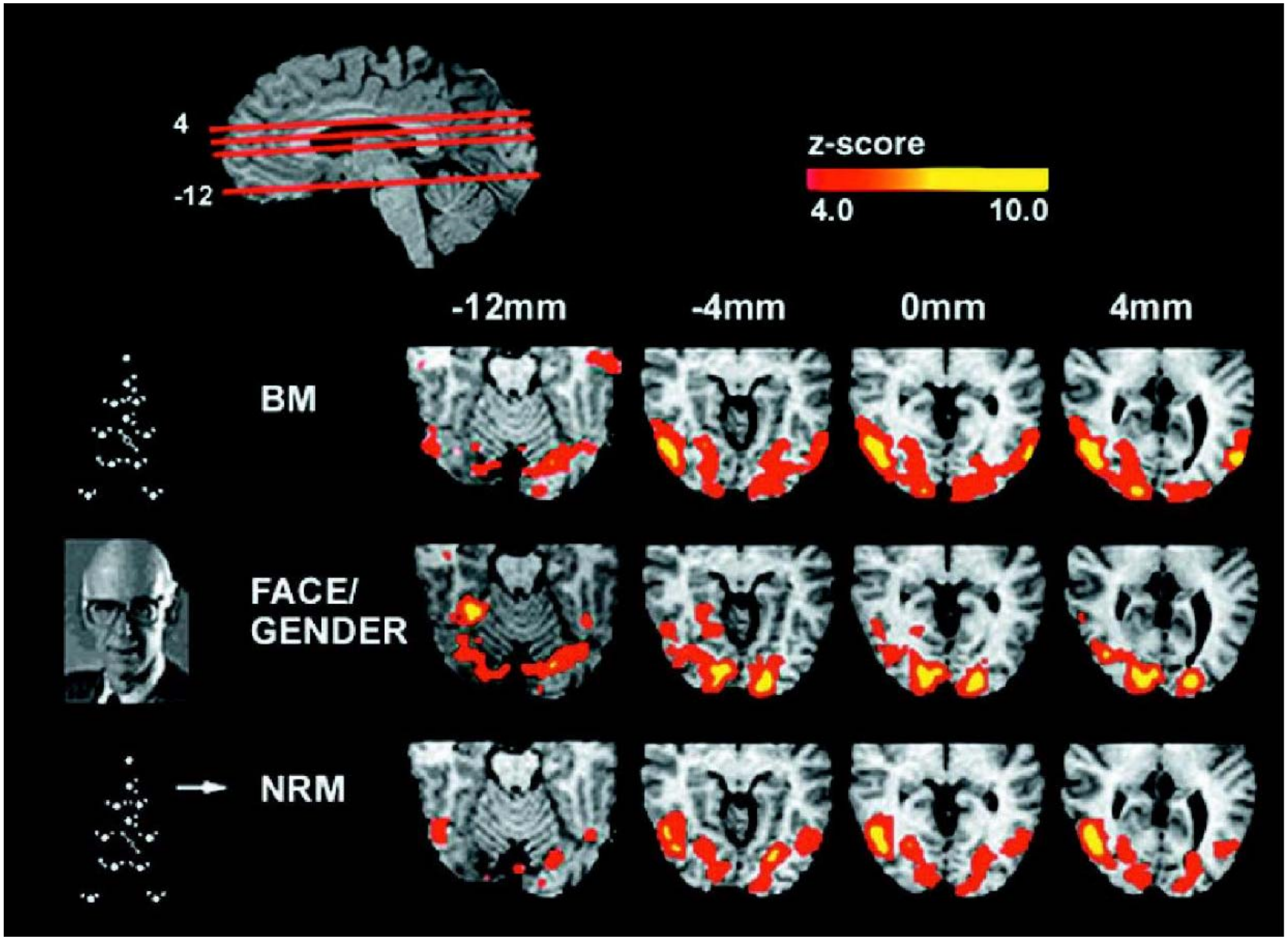
C. L. Leveroni et al. **“Neural Systems Underlying the Recognition of Familiar and Newly Learned Faces”,** The Journal of Neuroscience, January 15, 2000, *20*(2):878–886

**Table 1. Famous faces (FF) vs newly learned (NL) faces**

| Loc. # | Brain region | BA | vol. (ml) | *x* | *y* | *z* |
|---|---|---|---|---|---|---|
| | FF > NL | | | | | |
| | Frontal Lobe | | | | | |
| 1 | L Superior Frontal | 8 | 2.6 | −15 | 33 | 44 |
| 2 | R Medial Frontal | 9 | 2.4 | 10 | 47 | 25 |
| 3 | R Superior Frontal | 8 | 0.5 | 12 | 40 | 45 |
| 4 | L Medial Frontal | 10 | 0.4 | −6 | 49 | −4 |
| 5 | R Precentral | 6 | 0.4 | 49 | −1 | 13 |
| 6 | L Superior Frontal | 8 | 0.4 | −36 | 15 | 50 |
| 7 | R Inferior Frontal | 47 | 0.3 | 32 | 32 | −7 |
| 8 | R Anterior Cingulate | 32 | 0.3 | 11 | 21 | −7 |
| 9 | R Medial Frontal | 11 | 0.3 | 9 | 35 | −13 |
| 10 | L Medial Frontal | 11 | 0.3 | −6 | 39 | −14 |
| | Temporal Lobe | | | | | |
| 11 | L Middle Temporal | 21 | 2.7 | −51 | −11 | −13 |
| 12 | R Middle Temporal | 21 | 1.9 | 52 | −6 | −18 |
| 13 | L Middle Temporal | 21 | 0.6 | −49 | −42 | 7 |
| 14 | L Middle Temporal | 39 | 0.5 | −46 | −68 | 22 |
| 15 | R Superior Temporal | 22 | 0.5 | 54 | −52 | 15 |
| 16 | R Fusiform | 20/37 | 0.4 | 32 | −46 | −16 |
| 17 | R Middle Temporal | 37 | 0.3 | 43 | −64 | 9 |
| 18 | R Insula | – | 0.3 | 37 | 3 | 11 |
| 19 | R Parahippocampal | 35 | 0.2 | 30 | −14 | −23 |
| 20 | R Parahippocampal | 36 | 0.2 | 24 | −43 | −7 |
| 21 | L Hippocampus | 28 | 0.2 | −19 | −12 | −20 |
| | Parietal/Occipital Lobe | | | | | |
| 22 | L Posterior Cingulate | 23/30 | 1.7 | −4 | −57 | 15 |
| 23 | R Inferior Parietal | 40 | 0.5 | 44 | −30 | 22 |
| 24 | R Posterior Cingulate | 31 | 0.3 | 2 | −57 | 29 |
| 25 | L Extrastriate | 18 | 0.3 | −20 | −89 | 20 |
| | Subcortical | | | | | |
| 26 | R Pons | – | 0.4 | 11 | −43 | −34 |
| 27 | L Pons | – | 0.2 | −10 | −43 | −33 |
| 28 | R Putamen | – | 0.3 | 22 | −7 | −6 |
| | NL > FF | | | | | |
| | Parietal Lobe | | | | | |
| 29 | L Inferior Parietal | 40 | 1.0 | −37 | −64 | 40 |
| 30 | R Superior Parietal | 7 | 0.5 | 23 | −66 | 30 |
| 31 | R Inferior Parietal | 40 | 0.3 | 35 | −67 | 42 |

Region is defined as center of mass. The first column refers to location numbers demarcated in Figures 2 and 3 (italicized numbers indicate locations not shown in figures). Coordinates represent distance in millimeters from anterior commissure: $x$ right (+)/left (−); $y$ anterior (+)/posterior(−); $z$ superior (+)/inferior(−).

# Face and motion perception

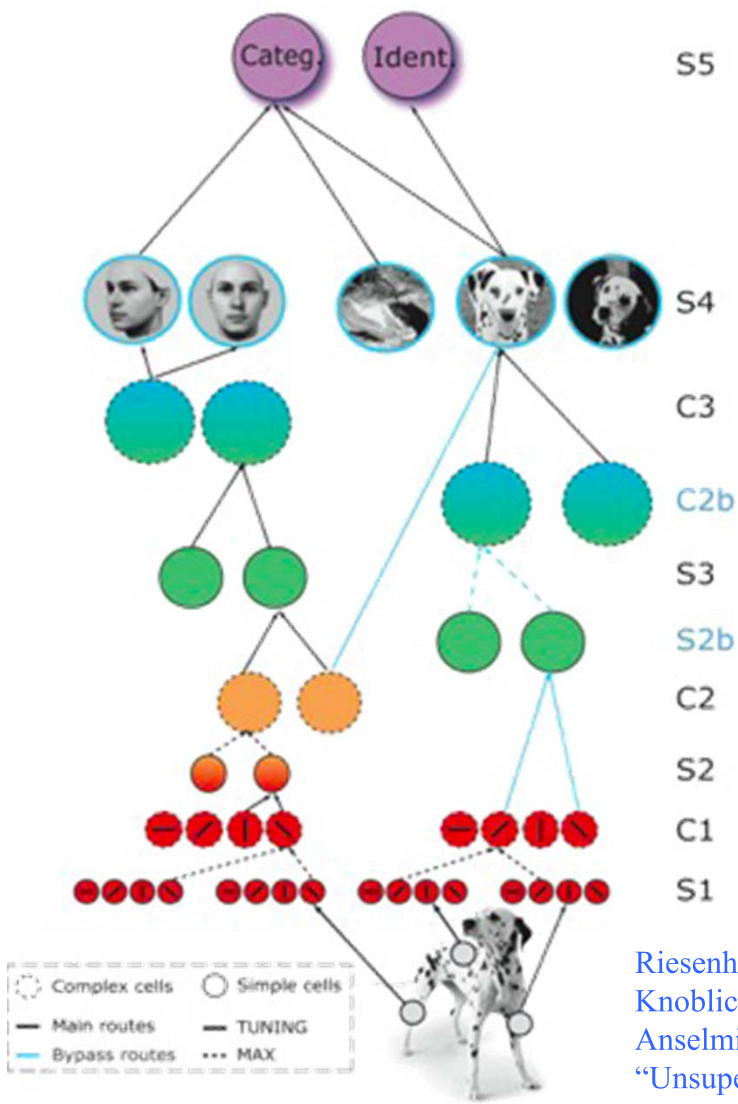


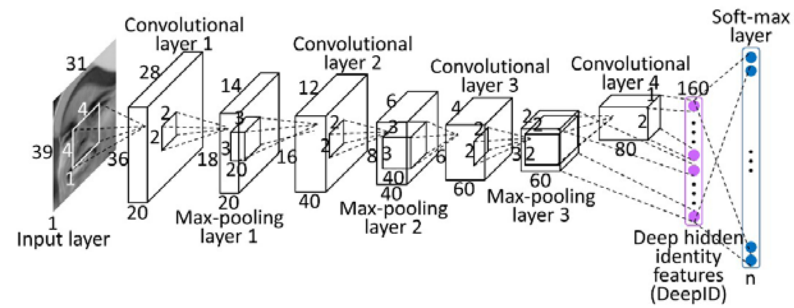

**B**iological **M**otion

**N**on **R**igid **M**otion

**Vaina, L.M., Solomon, J., Chowdhury, S., Sinha, P., Belliveau, J.W.,** "Functional Neuroanatomy of Biological Motion Perception in Humans". *Proc. of the National Academy of Sciences of the United States of America*, Vol. 98, No. 20 (Sep. 25, 2001) , pp. 11656-11661

# Brain models



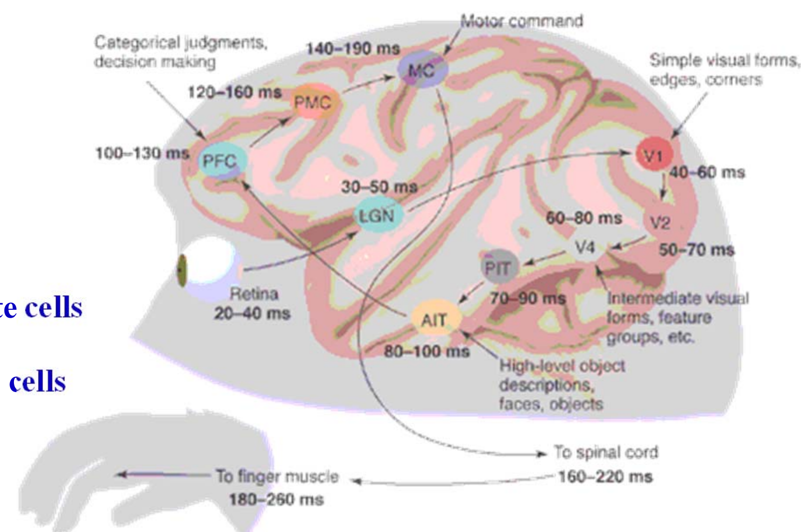




C2 **Complex composite cells**

S2 **Composite feature cells**

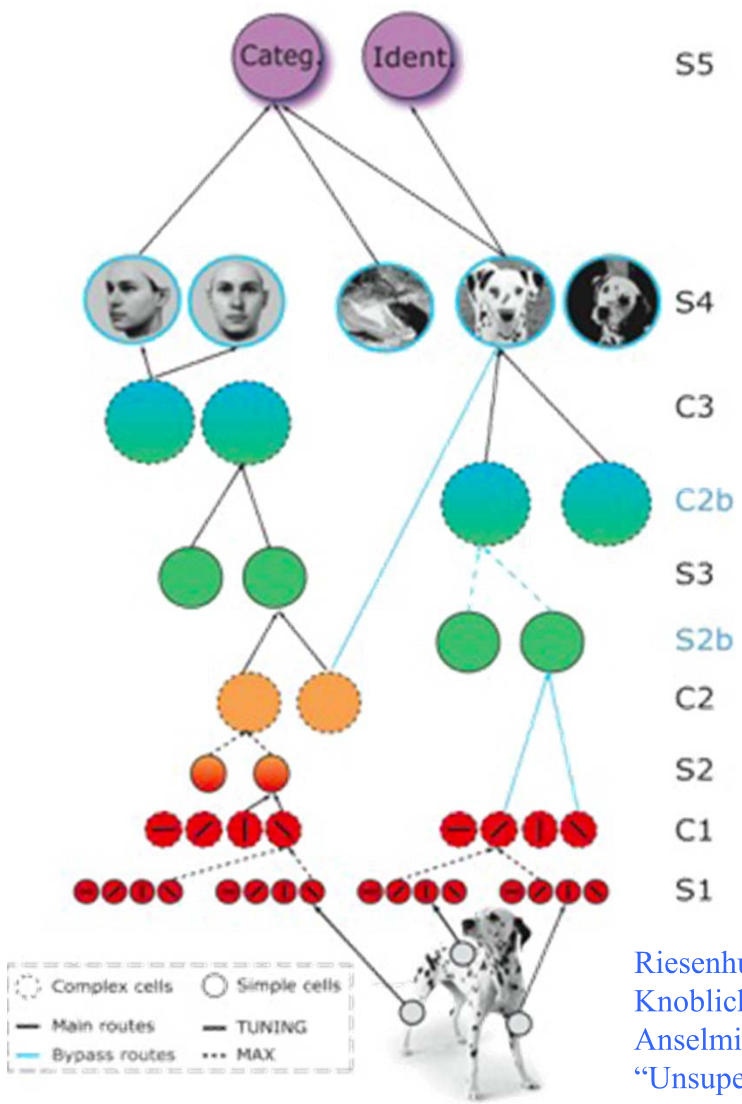
C1 **Complex cells**

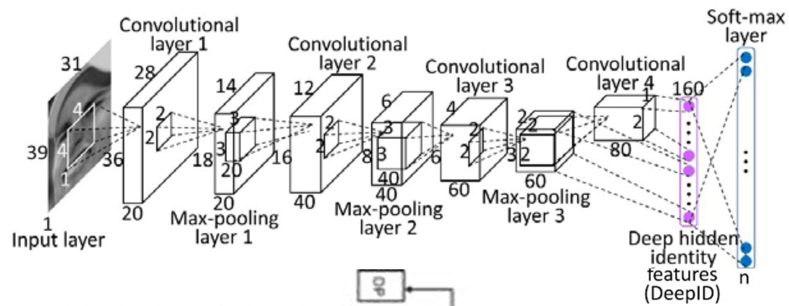
S1 **Simple cells**

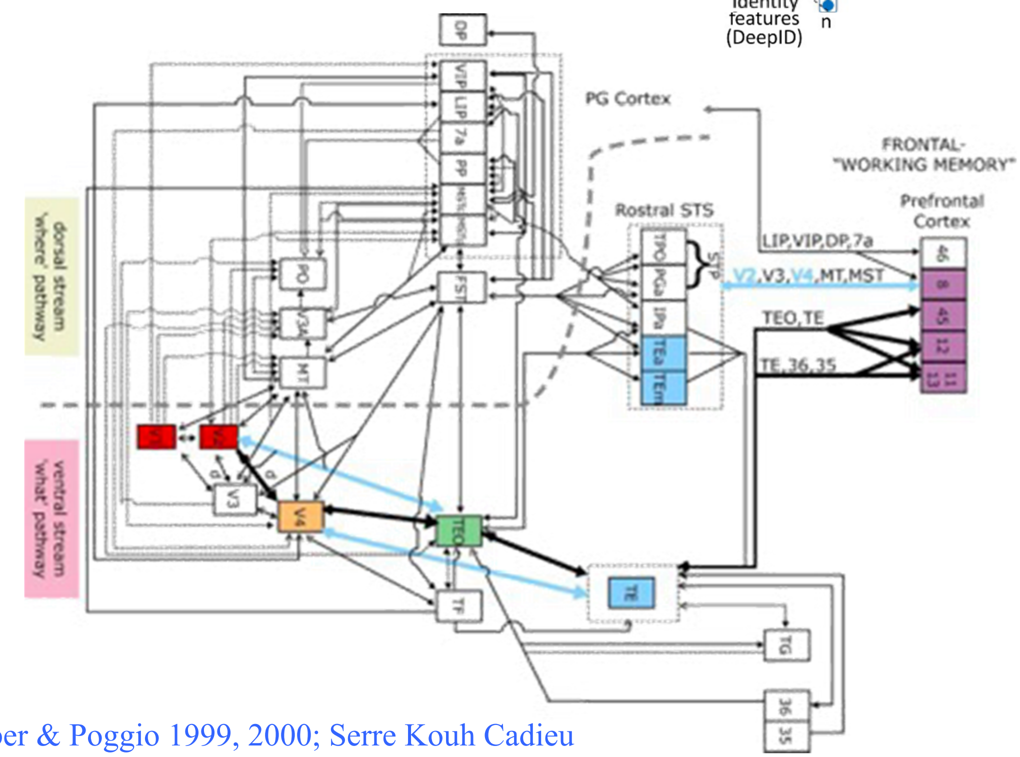



Riesenhuber & Poggio 1999, 2000; Serre Kouh Cadieu
Knoblich Kreiman & Poggio 2005; Serre Oliva Poggio 2007
Anselmi, F., Leibo, J. Z., Rosasco, L., Mutch, J., Tacchetti, A., and Poggio, T., "Unsupervised learning of invariant representations", Theoretical Computer Science, 2015.

# Brain models



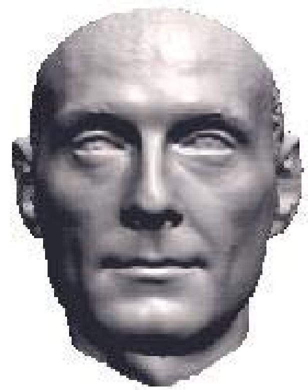

Riesenhuber & Poggio 1999, 2000; Serre Kouh Cadieu
Knoblich Kreiman & Poggio 2005; Serre Oliva Poggio 2007
Anselmi, F., Leibo, J. Z., Rosasco, L., Mutch, J., Tacchetti, A., and Poggio, T., "Unsupervised learning of invariant representations", Theoretical Computer Science, 2015.

# Towards 3D: Morphable models

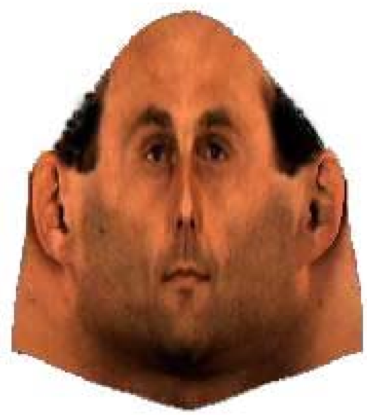

Laser Scanner

Max Planck Institute
Biologische Kybernetik

# Towards 3D: Morphable models



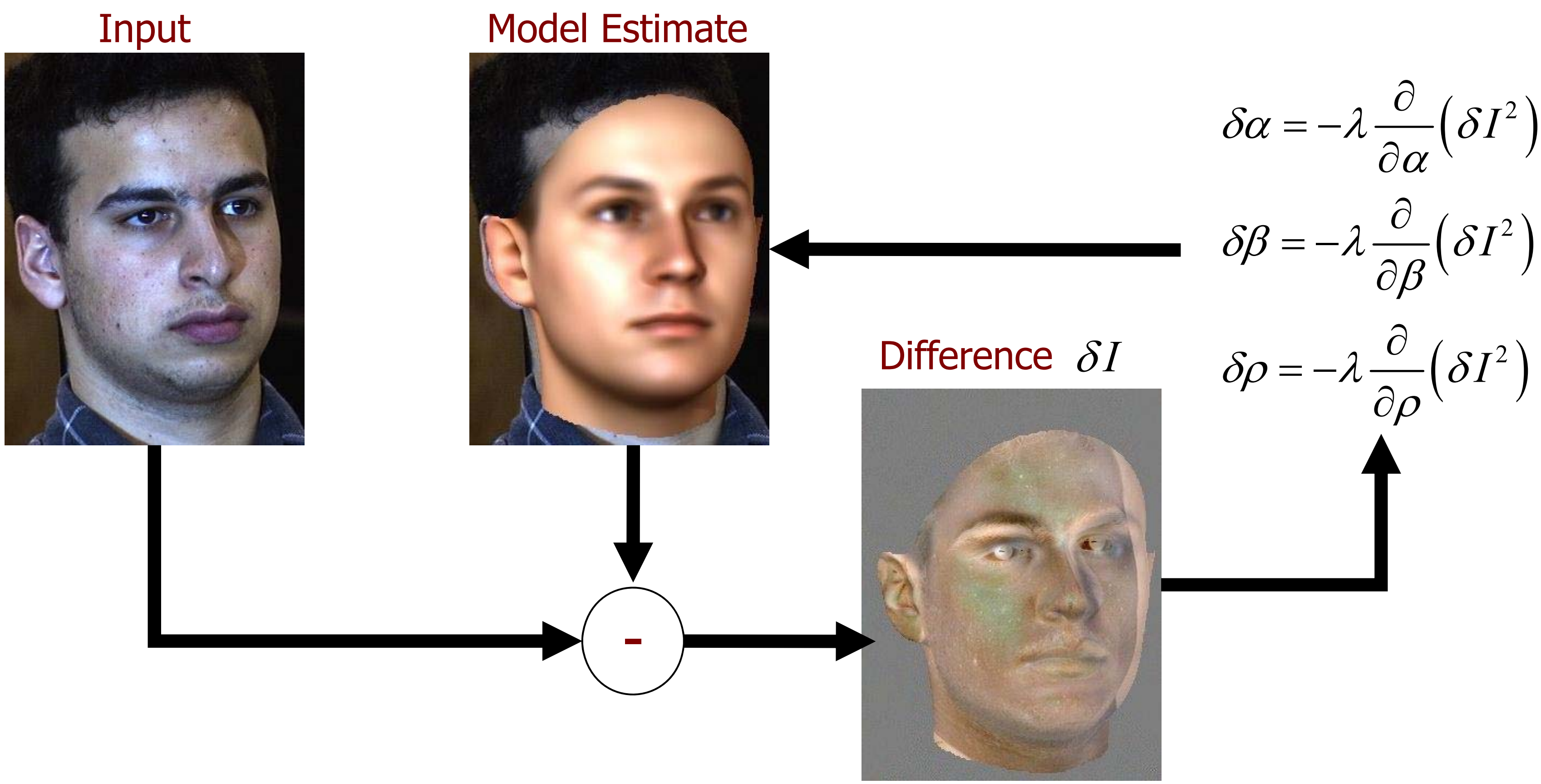
3D shape



surface reflectance

# Towards 3D: Morphable models

- Jones, Poggio 98: Gradient Descent
- Blanz, Vetter 99: Stochastic Gradient Descent
- Pighin, Szeliski, Salesin 99: Levenberg-Marquardt
- Romdhani, Blanz, Vetter 02: Non-linear fitting

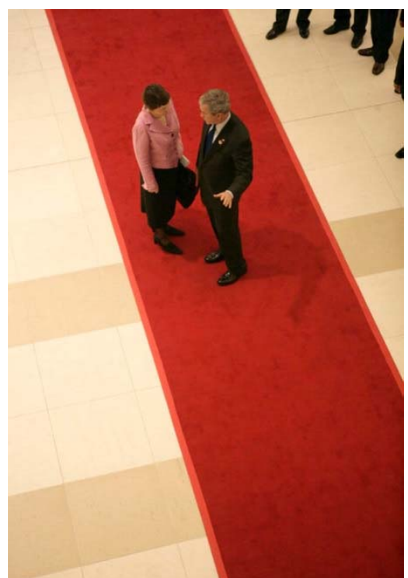


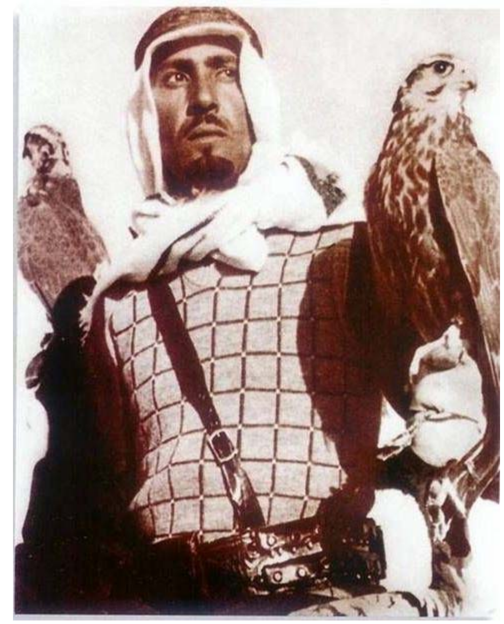
# Biometrics evaluations…

Vision Lab

| Modality | Test Label | Test Parameter | False Reject Rate | False Accept Rate |
|---|---|---|---|---|
| **Fingerprint** | FpVTE 2003 | US Government operational data | 0.1% | 1% |
| **Fingerprint** | FVC 2006 | Heterogeneous population (young, elderly) | 2.2% | 2.2% |
| **Face** | FRGC 2006 | Controlled Illumination, high-resolution images | 0.8-1.6% | 0.1% |
| **Voice** | NIST 2004 | Text independent, multi-lingual | 5-10% | 2-5% |
| **Iris** | ITIRT 2005 | Indoor environment | 0.99% | 0.94% |
| **Iris** | ICE 2006 | Controlled Illumination, broad quality range | 1.1-1.4% | 0.1% |

# The iArpa JANUS project





# The iArpa JANUS project

Dramatically improve face recognition performance in massive video collections through novel approaches capable of leveraging the rich spatial and temporal information available within the multiple views captured in unconstrained video.


Intelligence analysts often rely on facial images to assist in establishing the identity of an individual, but too often, just examining the sheer volume of possibly relevant images and videos can be daunting.

**Phase 2 – 18 months(3/16–9/17)**
datasets challenging for face detection, occlusion, aging

2000+subjects and hundreds of hours of video

**Accuracy**: **85% TAR @ 0.1% FAR**
**Query time**: sublinear

**Phase 3 – 36 months(10/17–9/20)**

10000+subjects and thusands of hours of video

**Accuracy**: **85% TAR @ 0.01% FAR**
**Query time**: logarithmic

# The USC JANUS team

P. Natarajan, PI

G. Medioni, Co-PI

R. Nevatia, Fusion

P. Debevec, Illumination

W. AbdAlmageed
Indexing, LSML

J. Choi
Face Recognition

R. Wu
FD, Systems

H. Li
Expression

L.P. Morency
LM Detection

T. Hassner,
2D matching

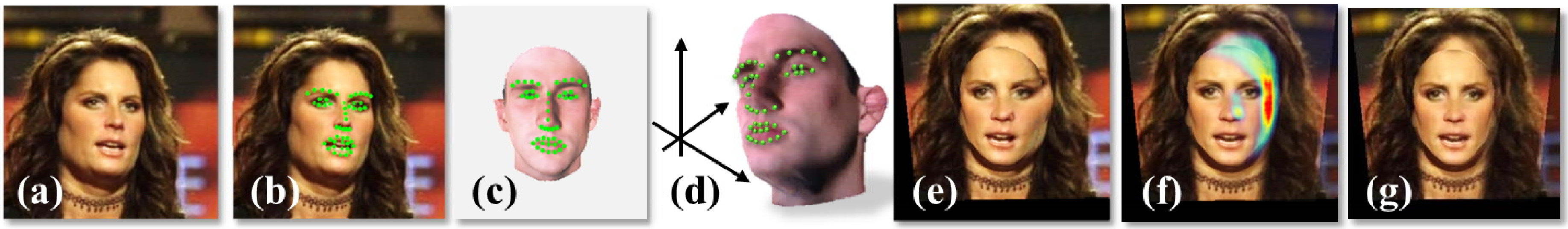
A. Del Bimbo, Firenze
Tracking

U. Park, Sogang U.
Aging, Distinctive

M. Tistarelli, UNISS
Age and Expression

M. Kilmer, Tufts U.
Tensor Approaches

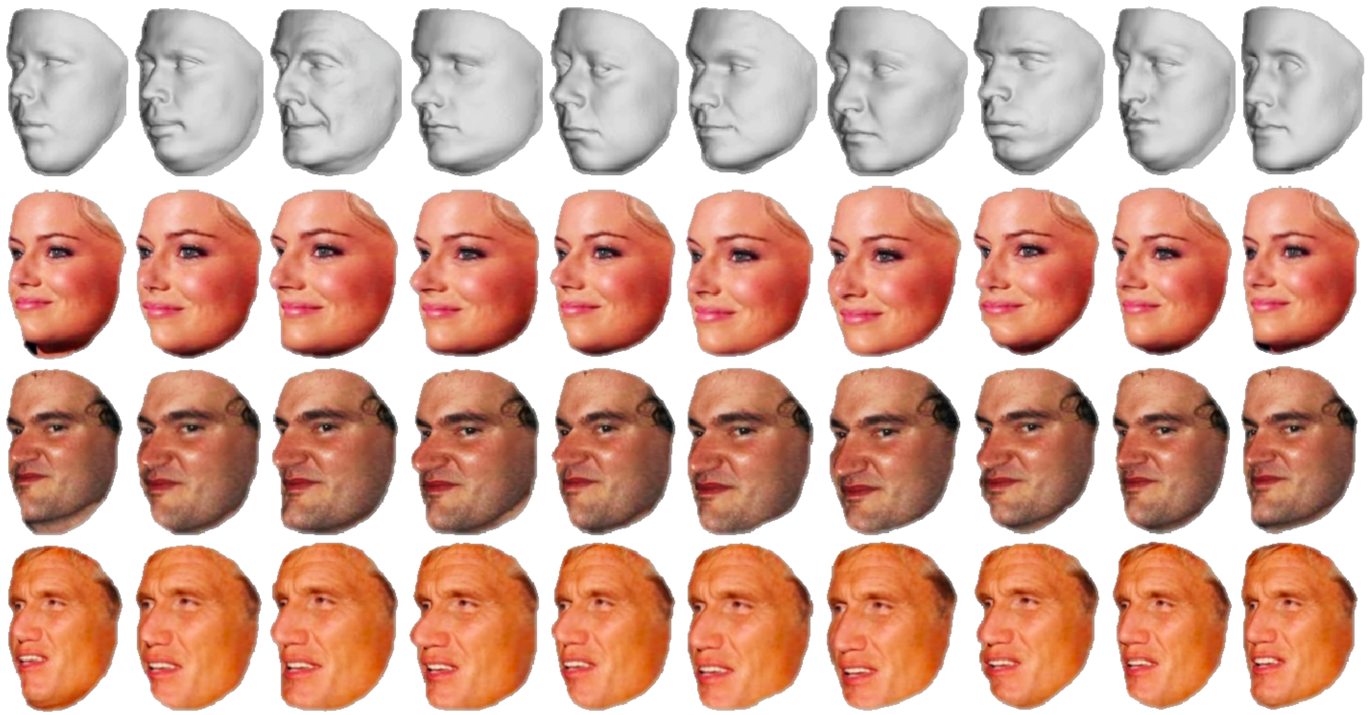
# 2D Frontalization




- (a) Query photo; (b) facial features detection; (c) the same detector used to localize the same facial features in a reference face photo, produced by rendering a textured 3D computer graphics model (d);
- (e) from the 2D coordinates on the query and their corresponding 3D coordinates on the model we estimate a projection matrix which is then used to back-project query intensities to the reference coordinate system;
- (f) estimated visibility due to non-frontal poses, overlaid on the frontalized result. Warmer colors reflect less visible pixels. Facial appearance in these regions is produced by borrowing colors from corresponding symmetric parts of the face; (g) final frontalized result.

Tal Hassner, Shai Harel, Eran Paz, Roee Enbar; "Effective Face Frontalization in Unconstrained Images" The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4295-4304

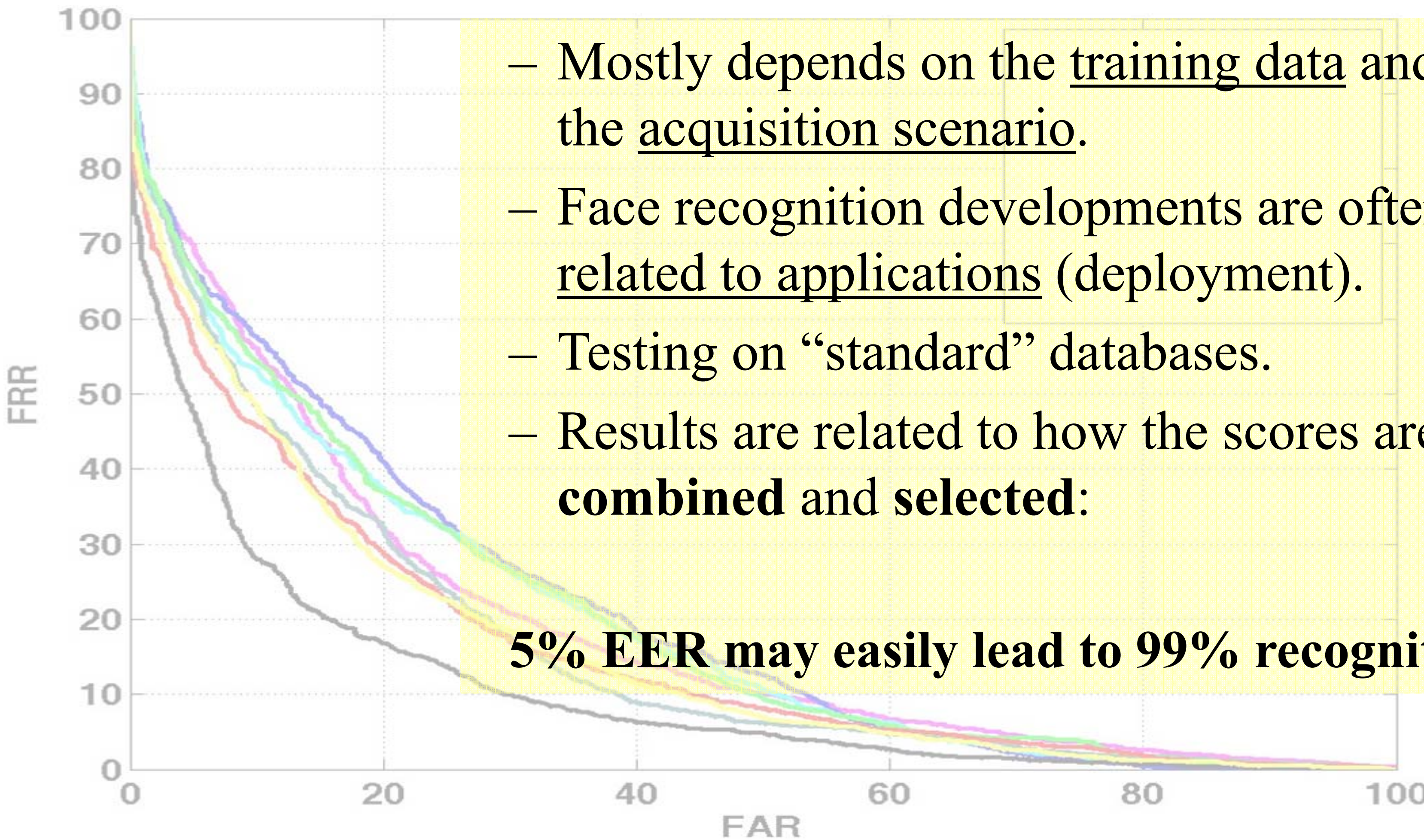
# Efficient CNN learning



**Augmenting faces by using different generic 3D models for rendering.**
**Top**: Ten generic 3D face shapes used for rendering. **Bottom**: Faces rendered with the generic model.
Different shapes induce subtle appearance variations yet do not change the perceived identity of the face in the image.
For training a CNN a single face image is rendered using different generic 3D models, at different poses and different expressions.

Iacopo Masi, Anh Tuan Tran, Jatuporn Toy Leksut, Tal Hassner, Gerard Medioni; "Do We Really Need to Collect Millions of Faces for Effective Face Recognition?" The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. arXiv preprint arXiv:1603.07057, 24 Mar 2016.

# Face recognition performances




- Mostly depends on the <u>training data</u> and the <u>acquisition scenario</u>.
- Face recognition developments are often <u>related to applications</u> (deployment).
- Testing on "standard" databases.
- Results are related to how the scores are **combined** and **selected**:

**5% EER may easily lead to 99% recognition**

# State-of-the-art in face recognition

## Good marks:

- Optimal classifiers (SVM/NN/Bayesian/SRC/DL)
- Advanced face-space representation (ICA/LFA/Kx)
- Best feature extraction methods (Gabor/LBP/MBLBP-SIFT/SURF/etc.) and *visual features*
- *Learning*

## Need for improvements:

- **Video vs mugshots**
- **Subject-based analysis *(familiarity)***
- **Registration**
- **Illumination**
- **Feature *selection***

# Illumination compensation

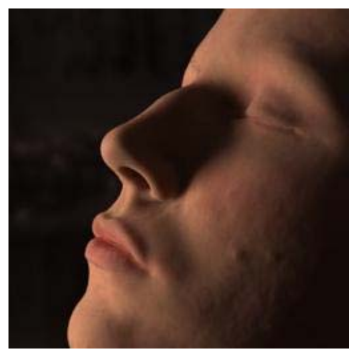
- **Main techniques**:
  - Histogram-based adaptive techniques, applied on image patches
  - Re-lighting techniques
  - Synthesis of illumination-invariant representations (for example the *Hue* component in color space)

# Modelling the face skin

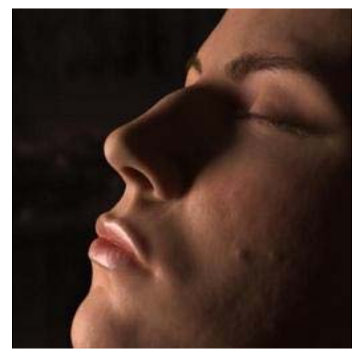
Skin chromaticity map

Diffuse light rendering

Reflectance map of the oily skin layer



Sub-surface reflectance



Final face rendering

***Henrik Wann Jensen,*** *"Digital face cloning", SIGGRAPH'2003 Technical Sketch, San Diego, July 2003. (http://graphics.ucsd.edu/~henrik/papers/face_cloning/)*

# Image re-lighting

$$I(x,y) = R(x,y) \cdot L(x,y) \qquad R(x,y) = \frac{I(x,y)}{L(x,y)} \quad \overset{?}{\Rightarrow} \quad L(x,y)$$

$$F(L) = \iint_{\varpi} \rho(x,y)(L(x,y) - I(x,y))^2 dxdy + \lambda \iint_{\varpi} (L_x^2 + L_y^2) dxdy$$

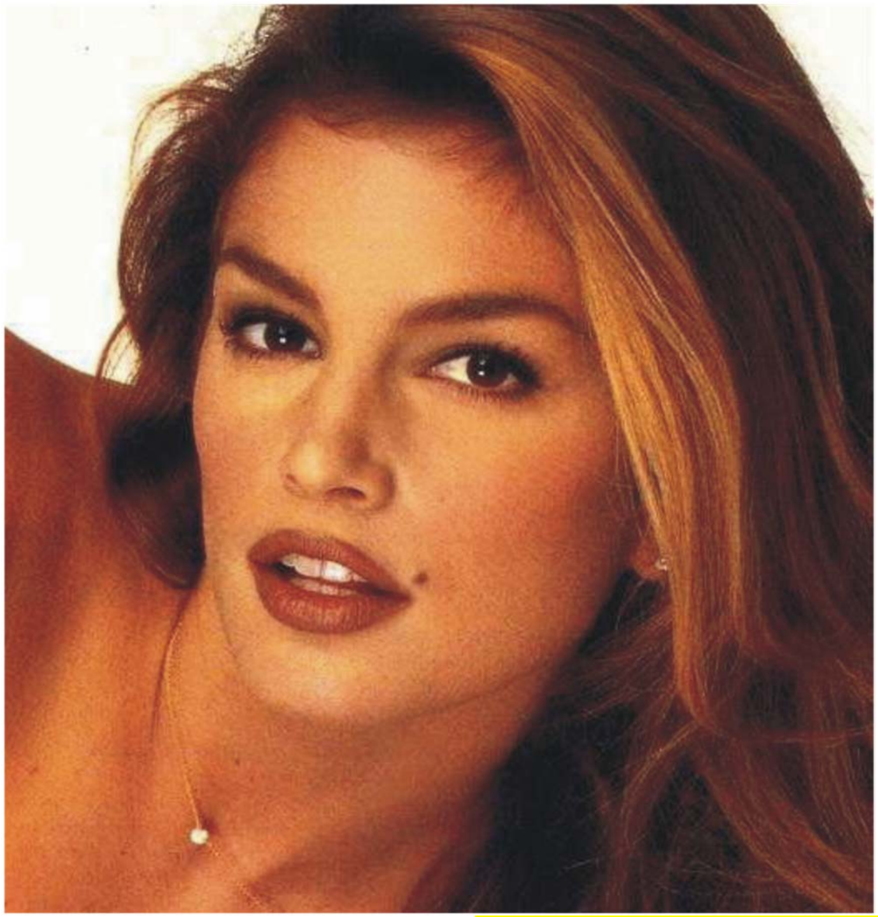
Anisotropic diffusion
(Lagrange solution of (1))

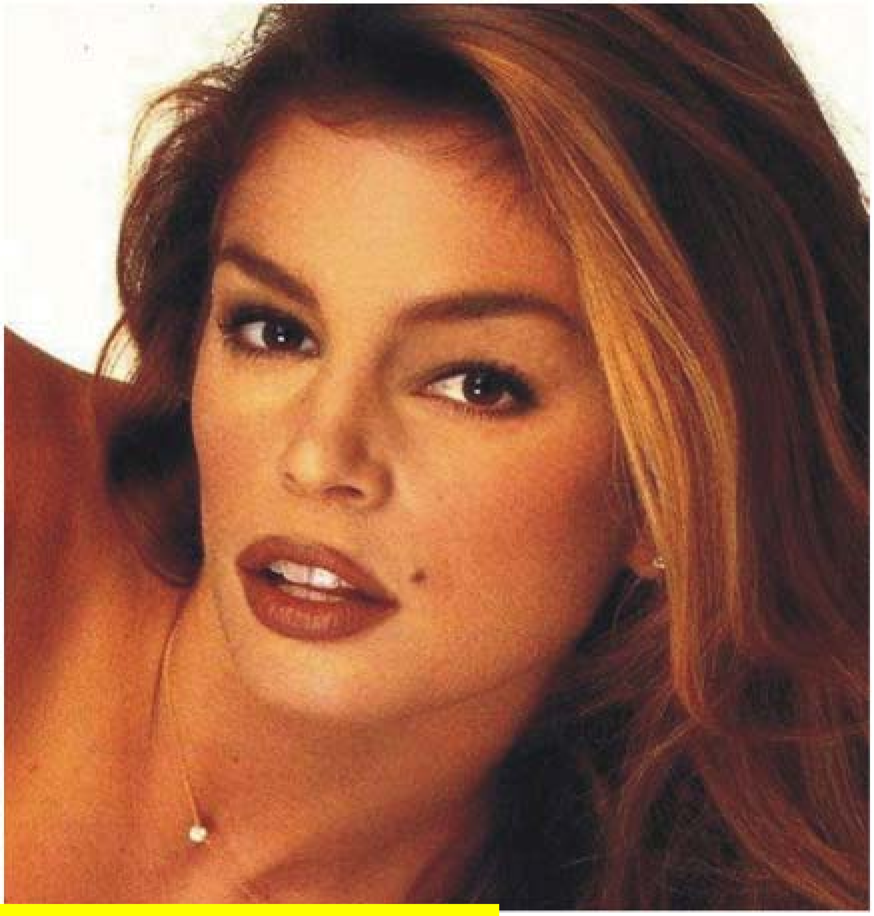
Isotropic diffusion
(Gaussian filtering)

**R. Gross and V. Brajovic**, "An Image Preprocessing Algorithm for Illumination Invariant Face Recognition", International Conference on Audio- and Video-Based Biometric Person Authentication, 2003.

**D. Jobson, Z. Rahmann and G. Woodell**, "A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observations of Scenes", IEEE Transanctions on Image Processing, volume 6, Issue 7, 1997.

# Face alignment

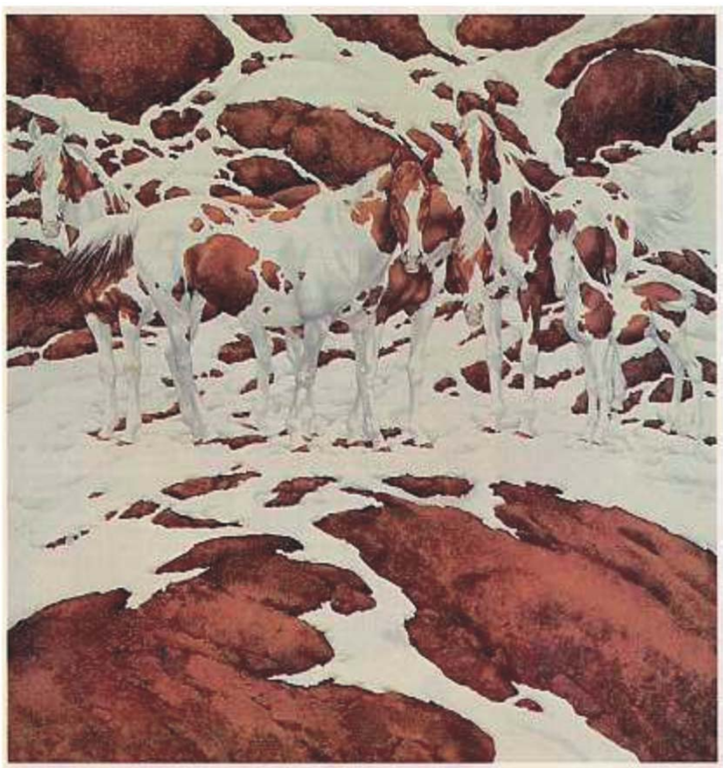




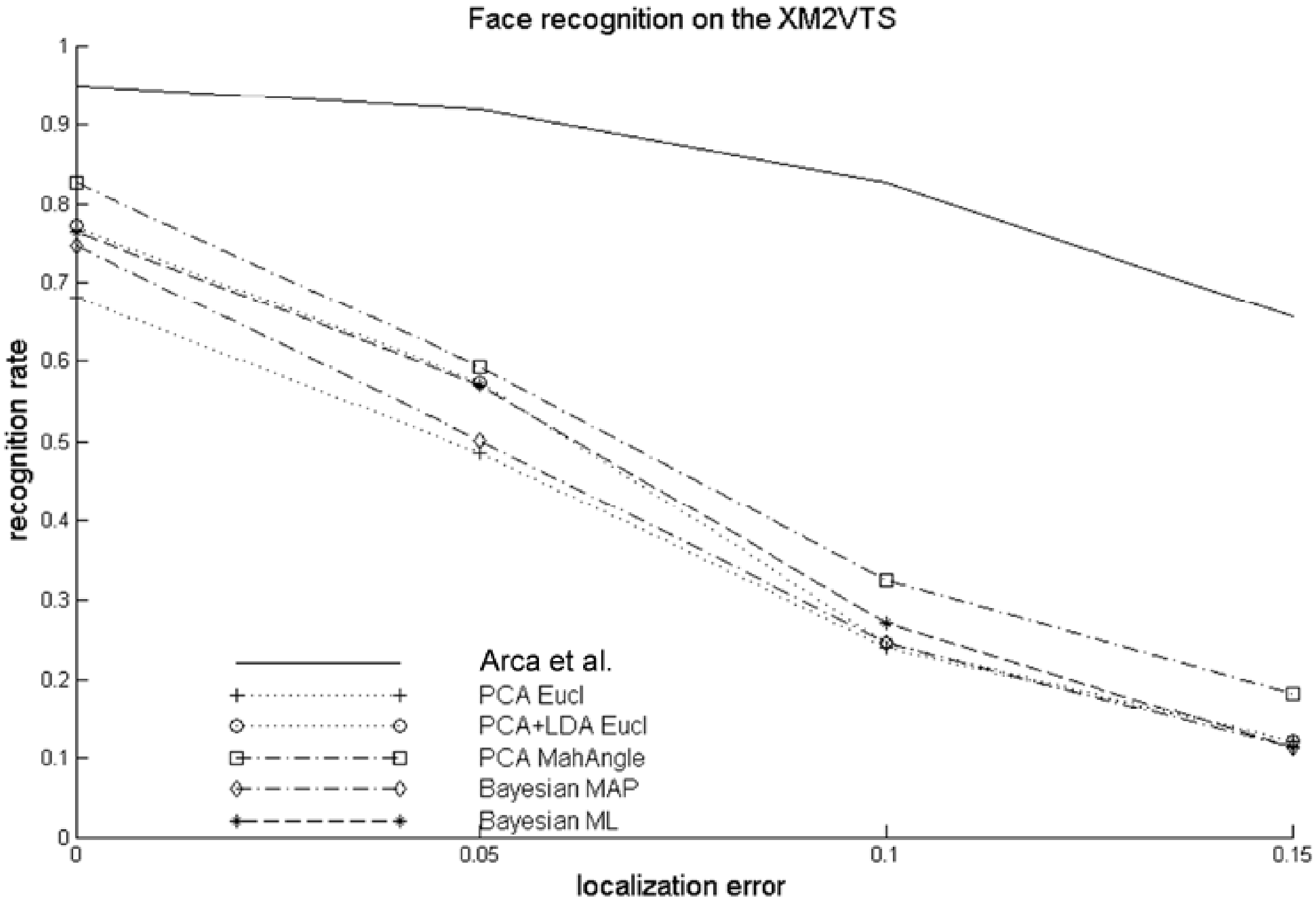
**… But the data is still there**

# The “registration” problem

In cognitive psychology it is called “perceptual organization”



# The "registration" problem



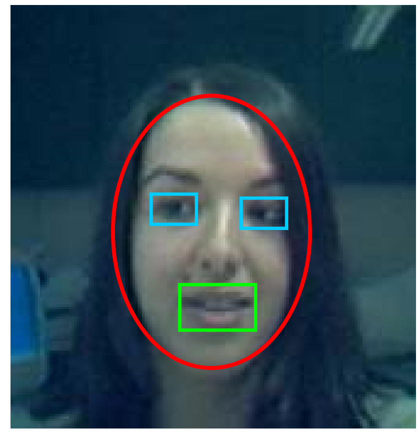

S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on automatically determined facial fiducial points. *Pattern Recognition*, 39(3):432–443, 2006.

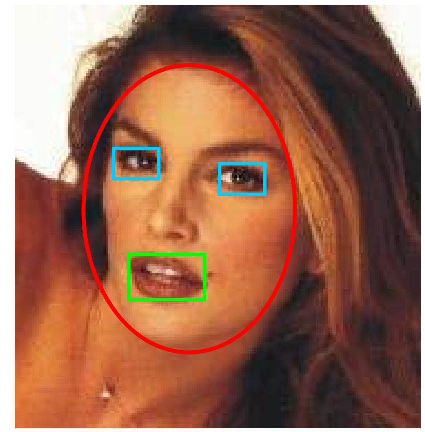
# Recent advances in face recognition

- **Face recognition from video sequences**
- **Subject-based template definition**
- **3D shape and texture**
- **Aging, gender, kinship, expression, intention…**
- Spoofing/Camouflage
- Face registration and Facial symmetry
- Compensation of illumination
  - Multispectral imaging
  - Evaluation of illuminant components
  - Face appearance-invariant models

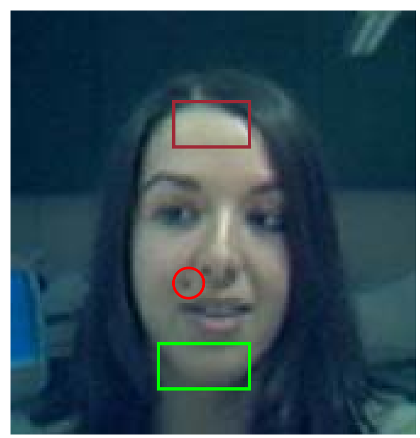
# Subject-specific representation

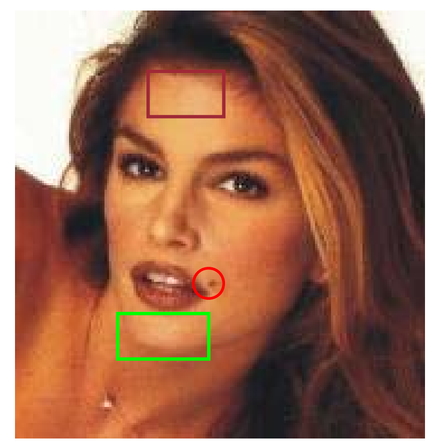
For **localization** and **tracking** we are interested on what every face has **<u>in common</u>** (to tell a face from "non-faces")



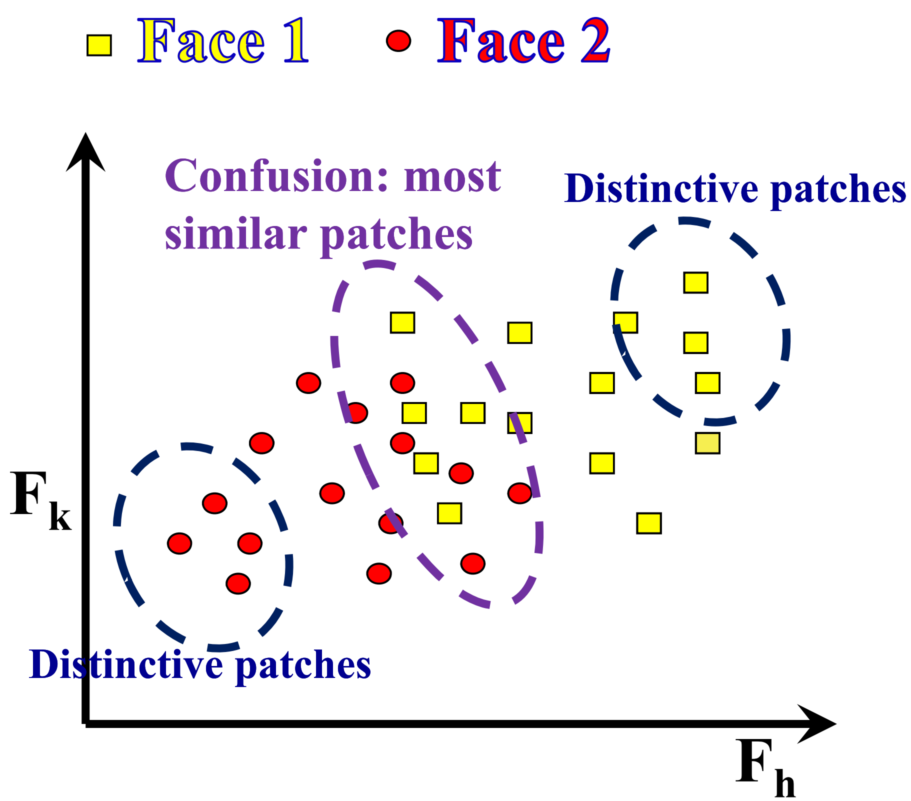


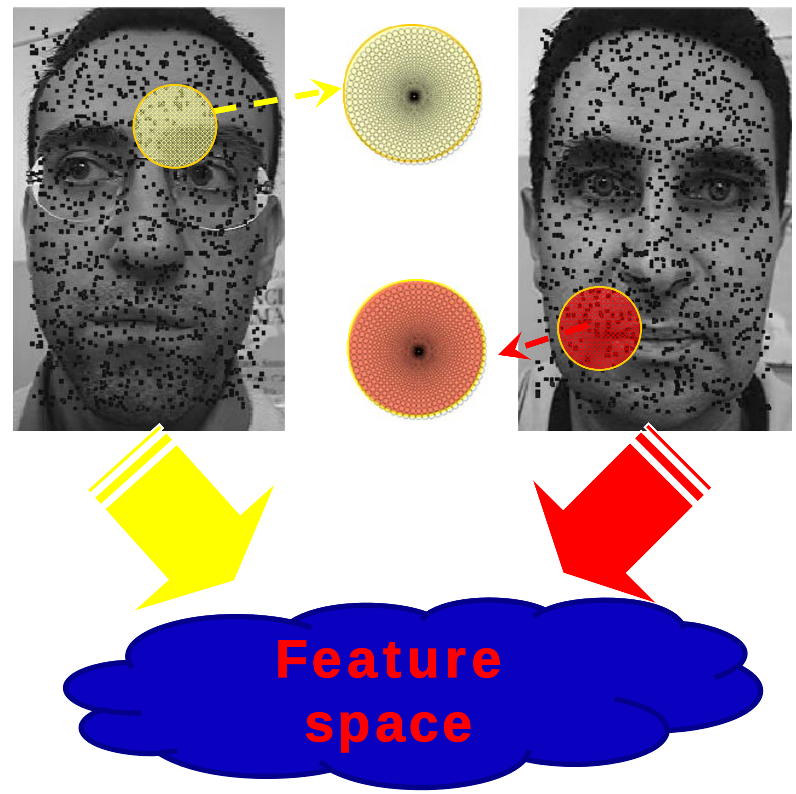
For **identification** we are not interested on what faces have in common but rather **<u>what differentiate</u>** one face from another.





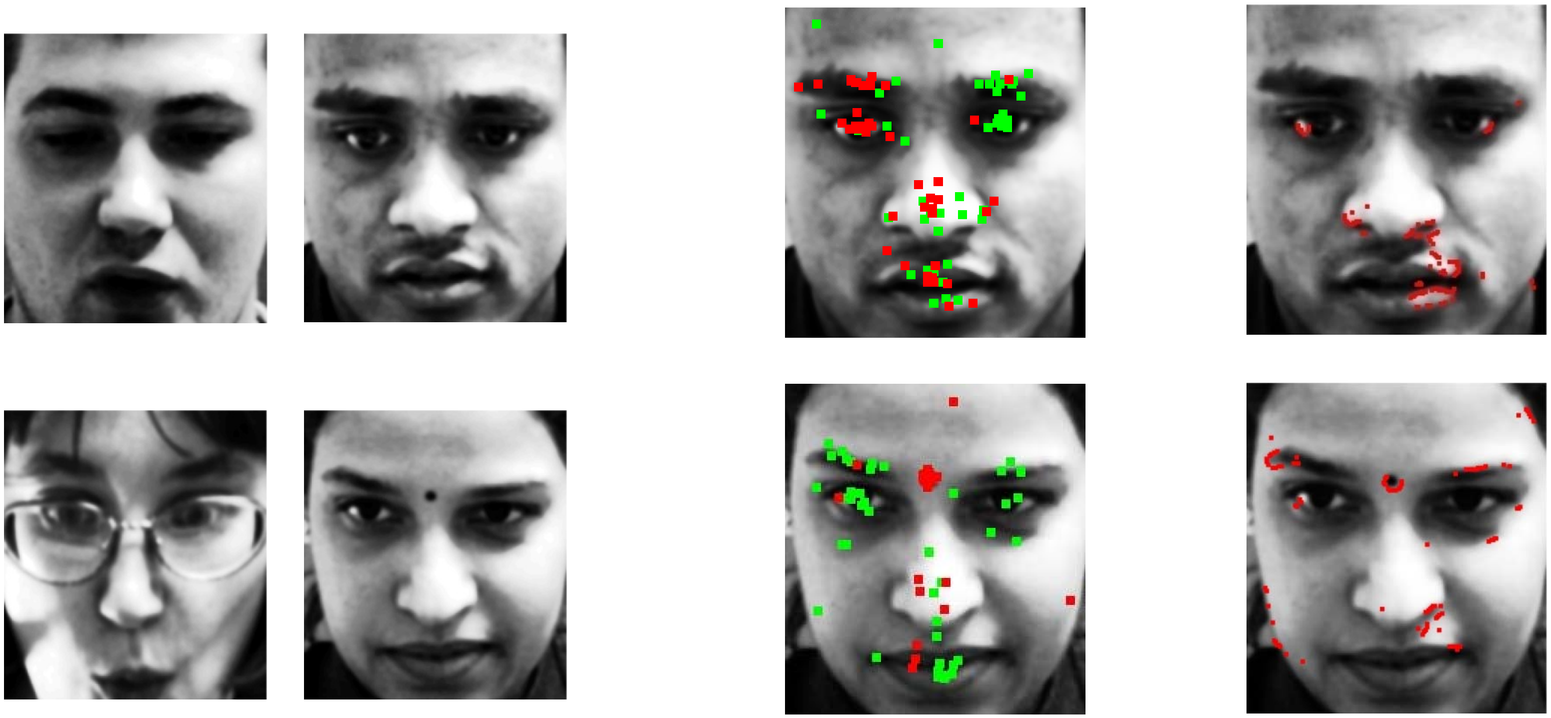
# Subject-specific representation




**Bicego M., Brelstaff G., Brodo L., Grosso E., Lagorio A. and Tistarelli M. (2007) "Distinctiveness of faces: a computational approach", ACM Transactions on Applied Perception, Vol. 5, n. 2, 2008.**

# Subject-specific representation



(A) perceptual and (B) computational results of saliency of local facial features, demonstrate the relevance of *non-standard* facial landmarks

# Selective attention

- Starting point: HMM based classification of faces
- “Walking on the face” for obtaining HMM  sequences

Standard raster scan-path

Saliency-based scan-path

Attention drives face scanning

A. A. Salah, M. Bicego, L. Akarun,  E. Grosso, M. Tistarelli: "Hidden Markov model-based face recognition using selective attention", *Human Vision and Electronic Imaging XII*, Proc. of SPIE, vol. 6492, (2007)

# Attention-based classification

- Experiments on BANCA protocol MC
- Gabor wavelets for saliency map construction
- Employed features: gray levels, DCT coefficients, Haar wavelets

| Window Size | Average Acc. (std) | | Max Acc. | |
|---|---|---|---|---|
| | Biological | Raster | Biological | Raster |
| 7 | 87.62%(2.28%) | 91.92%(1.63%) | 91.15% | 93.08% |
| 9 | 89.31%(1.20%) | 93.92%(0.92%) | 90.38% | 95.00% |
| 11 | 93.69%(1.58%) | 94.46%(1.29%) | 95.77% | 95.77% |
| 13 | 95.23%(0.89%) | 96.08%(0.74%) | 96.15% | 97.31% |
| 15 | 96.85%(1.00%) | 95.85%(1.29%) | 98.08% | 97.31% |
| 17 | 93.15%(1.13%) | 96.69%(0.89%) | 95.00% | 98.08% |

**Table 2.** Comparison between raster and biological scanning

A. A. Salah, M. Bicego, L. Akarun, E. Grosso, M. Tistarelli: "Hidden Markov model-based face recognition using selective attention", *Human Vision and Electronic Imaging XII*, Proc. of SPIE, vol. 6492, (2007)

# FACE FROM VIDEO TECHNOLOGIES


[illegible]

[illegible]

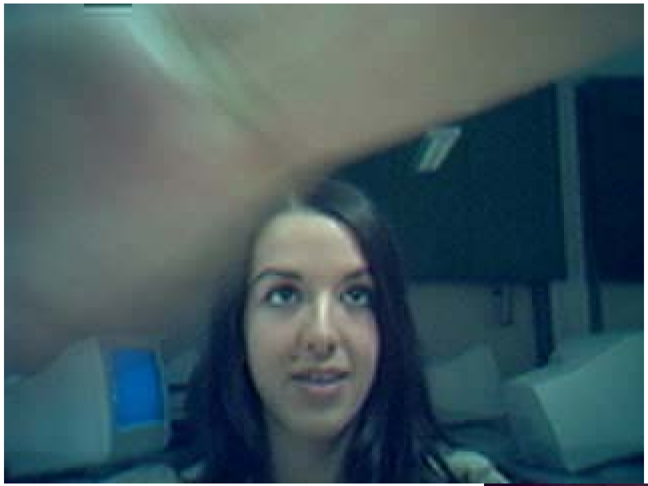
[illegible]

M. Bicego, E.Grosso, M. Tistarelli. “Person authentication from video of faces: a behavioral and physiological approach using Pseudo Hierarchical Hidden Markov Models”, Int.l Conference on Biometric Authentication 2006, Hong Kong, January 2006/IMAVIS 2009

Yi-Chen Chen; Patel, V.M. ; Chellappa, R. ; Phillips, P.J. “Adaptive representations for video-based face recognition across pose”, 2014 IEEE Winter Conference on Applications of Computer Vision (WACV), 984-991, 24-26 March 2014.

# Face recognition from video

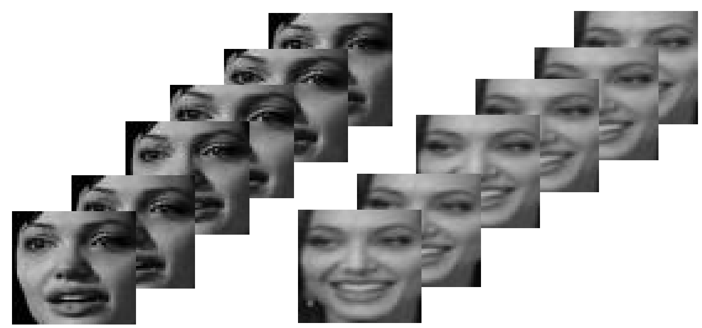
Dynamics in a video stream conveys far more information than a collection of single snapshots



# Video-Based Face Recognition

- General face recognition problem: Identifying faces in the query (probe) given a stored database of ID-labeled faces (gallery).
- Why videos?
  - Video *naturally* arises in many applications
  - Videos contain more information: Spatio-temporal patterns, evidence accrual, 3D information etc.

**Still to Video**



**Video to Video**

# Video-Based Face Recognition

Not just *more data* to be processed:

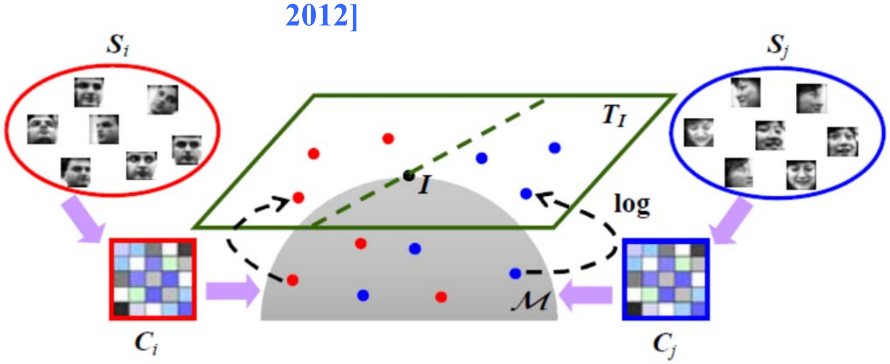
- Select the "best" sensory data (pose, expression, illumination, noise…)
- Multiple-data fusion (decision/score/feature level)
- 3D reconstruction/virtual views
- Resolution enhancement
- Expression and emotion analysis
- Behavioral analysis
- *Dynamic* video templates…?

# Representations

- **Linear subspaces**: Discriminative canonical correlation [Kim et al. PAMI 2007]
- **Affine subspaces**: Affine hull, convex hull [Cevikalp & Triggs CVPR 2010]
- **Manifolds**: [Lee et al. CVIU 2005; Wang et al. CVPR 2008, CVPR 2009]
- **Probability distributions**: [Zhou & Chellappa PAMI 2006]
- **Covariance matrices**: [Wang et al. CVPR 2012]
- **Dictionaries**: [Chen et al. ECCV 2012]
- **Temporal models**: HMM [Liu & Chen CVPR 2003], PH-HMM [Bicego et al. ICBA 2006], ARMA [Aggarwal et al. ICPR 2004]
- **3D models**: [Park & Jain ICB 2007]

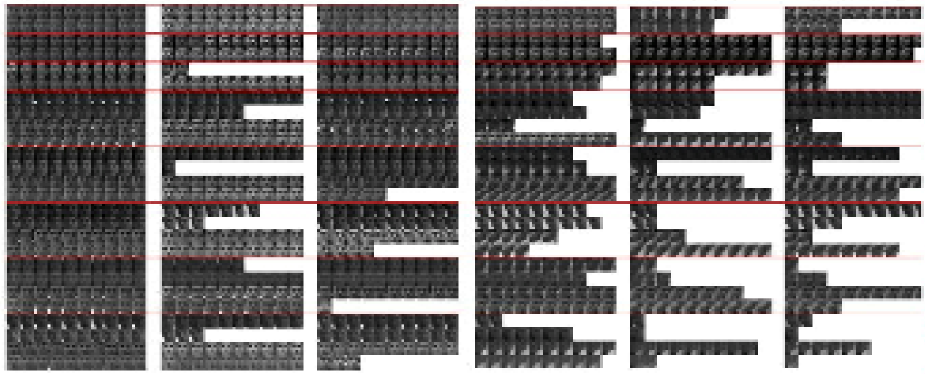
[Chen et al. ECCV 2012]

[Wang et al. CVPR 2012]



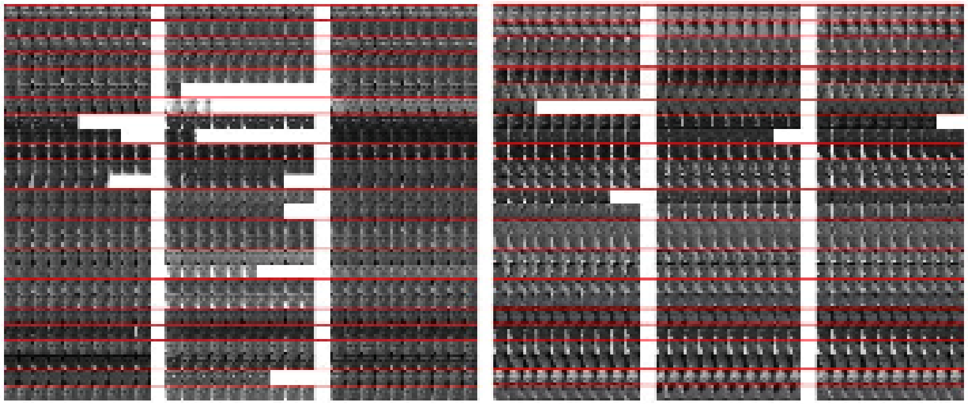
# Video preprocessing/partitioning

- Given a video sequence, extract all frames from it, from which the human face regions are then detected and cropped.
- Based on video summarized algorithm [1], partition cropped face images into *K* partitions.
  - Different partitions exhibit different pose/lighting conditions.



(a) MBGC Notre Dame frontal face partitions (b) MBGC Notre Dame profile face partitions

Figure 3. MBGC Notre Dame partition results



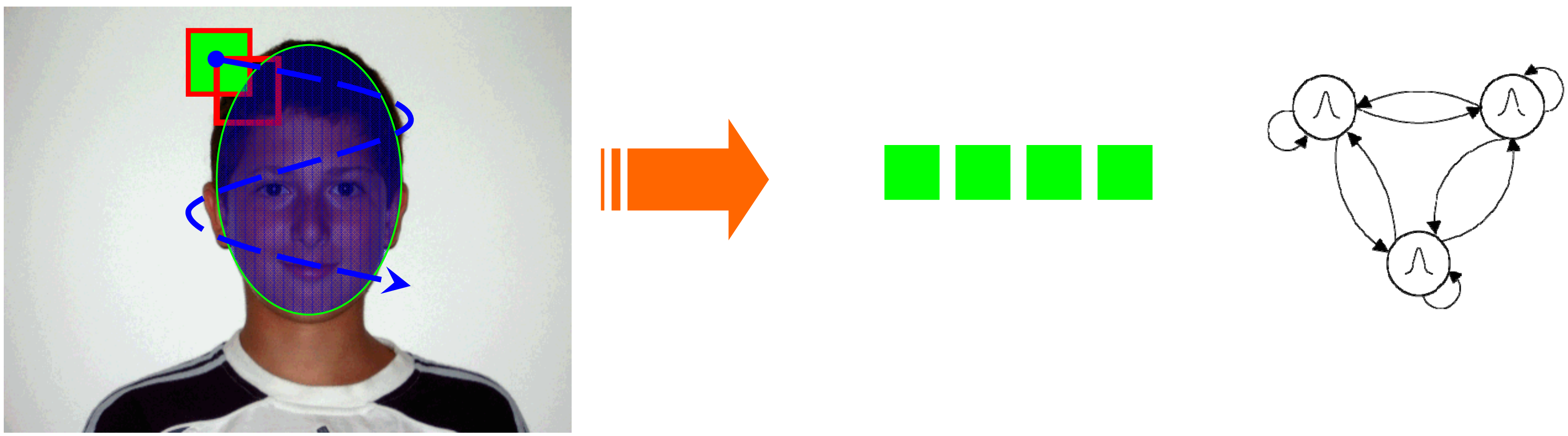
(a) FOCS UT-Dallas walking partitions (b) FOCS UT-Dallas activity partitions

Figure 9. FOCS UT-Dallas partitions

[1] N. Shroff, P. Turaga, and R. Chellappa, "Video précis: Highlighting diverse aspects of videos," *IEEE Transactions on Multimedia*, 2010.

# Hidden Markov Models

Statistical analysis of sequences of patterns



- This idea can be extended to multi-dimensional patterns and sequences … *in several ways*

# Hidden Markov Models

1. Each image is modeled as a single HMM and the sequence of images as a sequence of HMMs

   A. Hadid and M. Pietikainen. "An experimental investigation about the integration of facial dynamics in video-based face recognition". *Electronic Letters on Computer Vision and Image Analysis*, 5(1):1-13, 2005.

2. The entire video is modeled as a single HMM

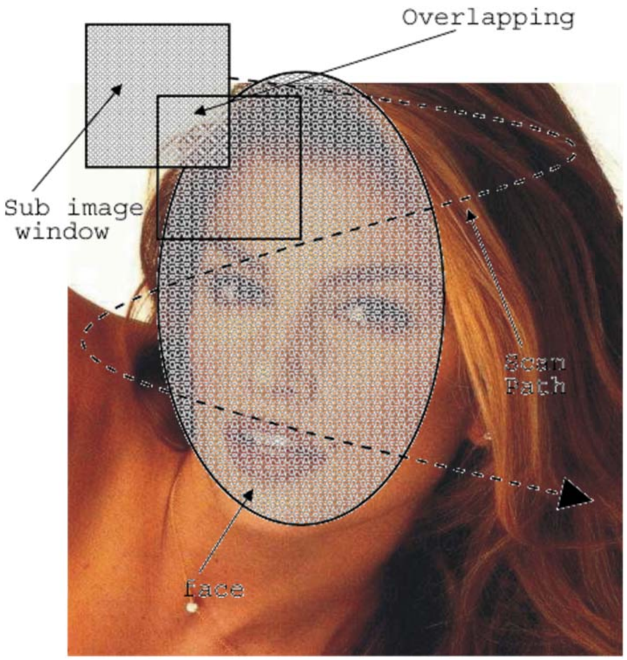
   X. Liu and T. Chen. "Video-based face recognition using adaptive hidden Markov models". In *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, 2003.

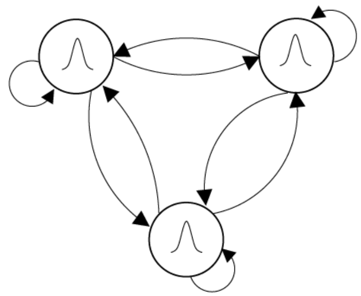
3. The images and the sequence itself are modeled as a complex, hierarchical HMM-based structure

   M. Bicego, E.Grosso, M. Tistarelli. "Person authentication from video of faces: a behavioral and physiological approach using Pseudo Hierarchical Hidden Markov Models", Int.l Conference on Biometric Authentication 2006, Hong Kong, China, January 2006.

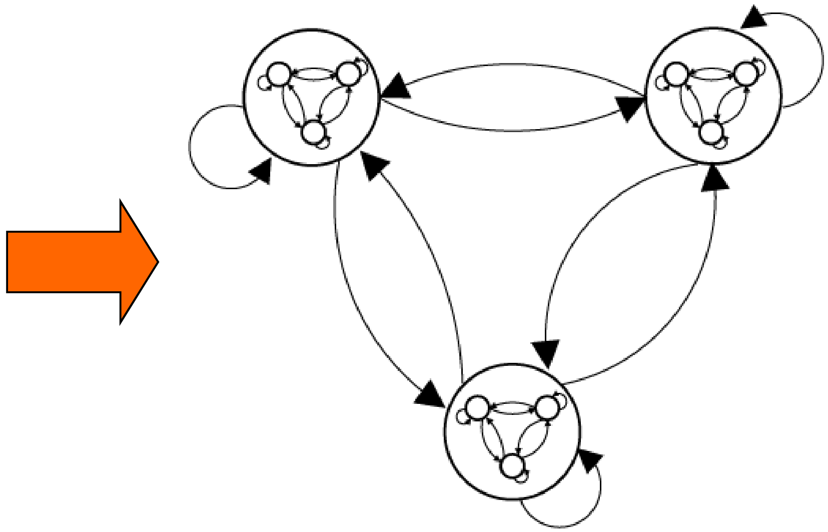
# Pseudo Hyerarchical HMM

## Physiological and behavioral features

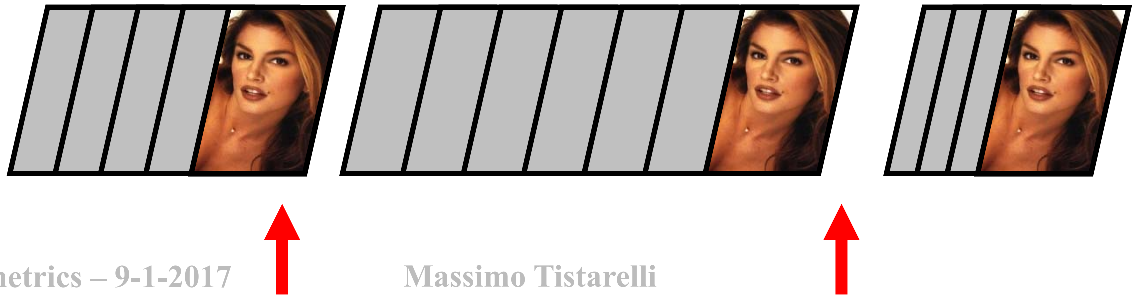


HMM



PH-HMM



# …ANOTHER WAY OF USING MORE IMAGES:

# COHORT NORMALIZATION

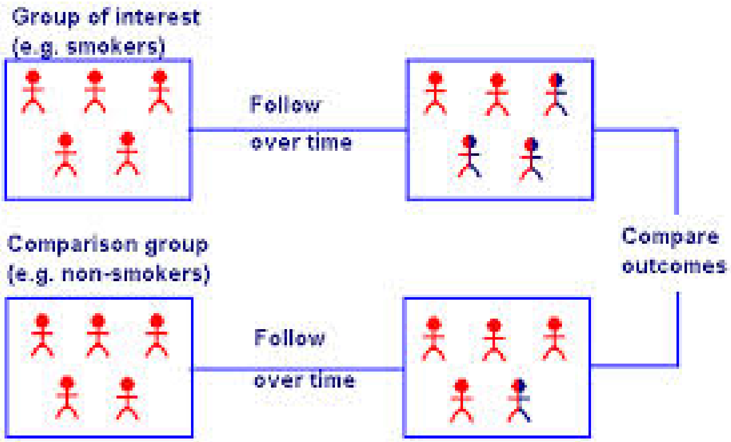

Y. Sun, M. Tistarelli, N. Poh (2013); "Picture-Specific Cohort Score Normalization for Face Pair Matching" Proc IEEE 6th Int.l Conference on Biometrics: Theory, Applications and Systems - BTAS 2013 Washington DC, USA; September 29 - October 2, 2013.

M. Tistarelli, Y. Sun, N. Poh (2014) "On the Use of Discriminative Cohort Score Normalization for Unconstrained Face Recognition", IEEE Trans. on Information Forensic and Security, 9(12):2063-2075, 2014.

# What is a Cohort?

- Ancient Roman military unit, comprising six centuries, equal to one tenth of a legion.

- A group of people banded together or treated as a group.



# What is a Cohort set?

- ***Cohort samples*** are non-matching samples of the same kind of the test samples.

Fingerprint 1
(from subject A)

Fingerprint 2
(from subject A/B)

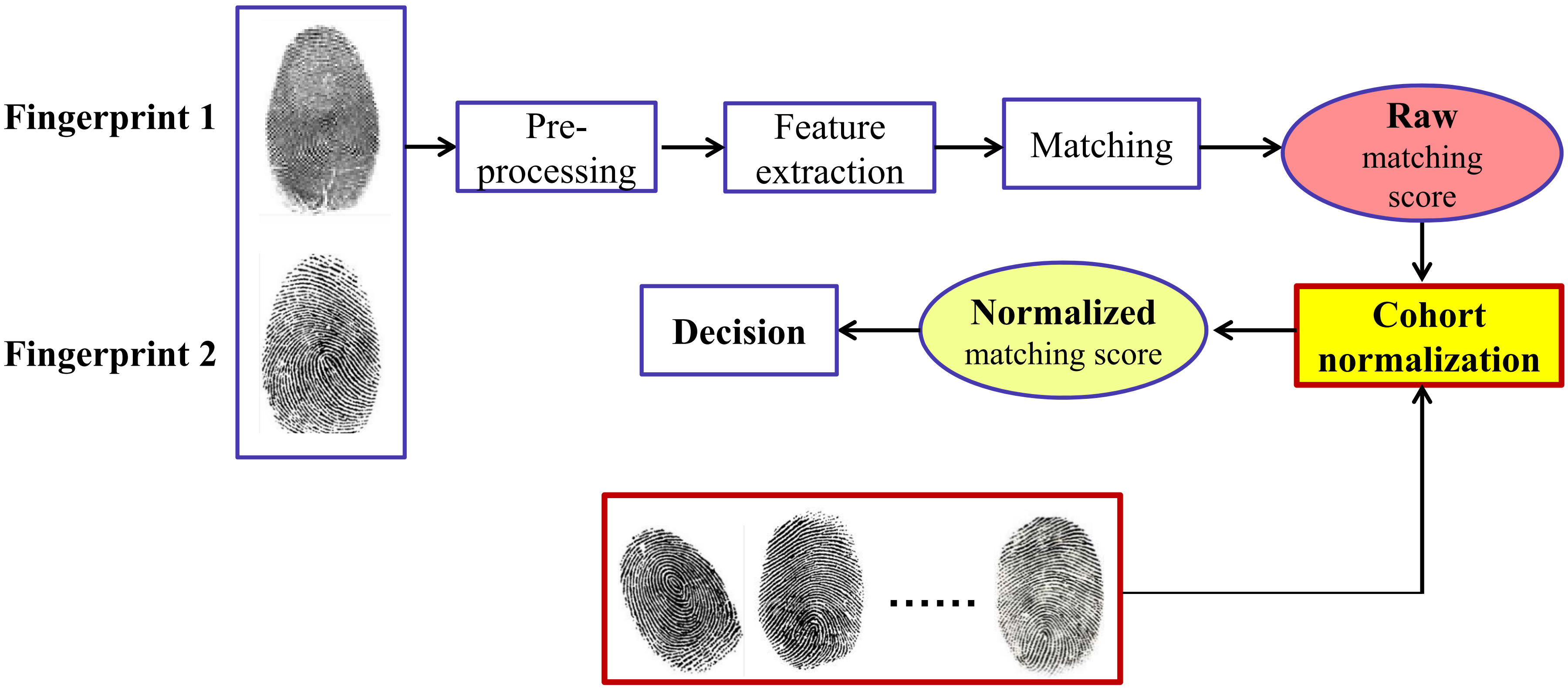
**Two fingerprints being compared**

**Cohort samples:**
fingerprints from ***external*** database

**An *independent* data set**

# What is Cohort normalization?

- Cohort score normalization is a procedure, which aims to post-process the matching score, using information from a set of *cohort samples*.




**Pool of cohort samples**

# How Cohort modeling works?

**Score normalization** can *map* the raw matching score to a domain where the degradation effects caused by the sample variations are reduced.

**Cohort score normalization** can exploit the *discriminative information* by a pool of non-matching samples.
This information can be used to normalize the raw matching scores.

# Existing literature

- Cohort models have been proposed to model **language processing** and lexical retrieval.

Marslen-Wilson, W. (1987). "Functional parallelism in spoken word recognition" *Cognition, 25, 71-102*

- Initially adopted in biometrics for **speaker verification**.

A. E. Rosenberg, J. DeLong, C. H. Lee, B. Juang, and F. K. Soong (1992). “The use of cohort normalized scores for speaker verication” In Int.l Conf. on Spoken Language Processing,1992.

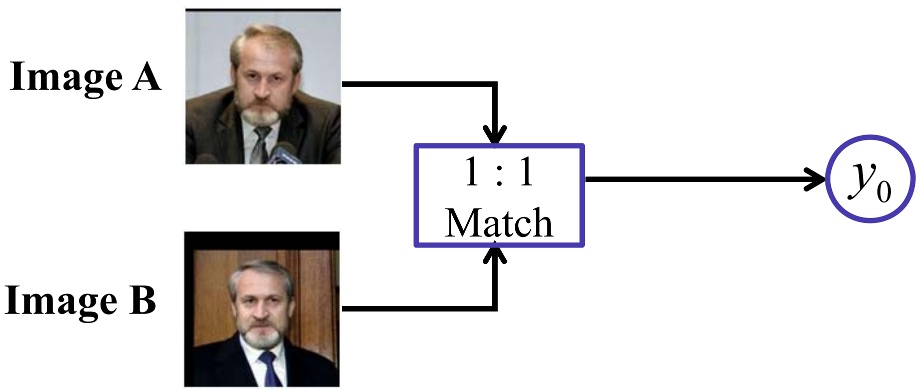
- Afterwards applied for **fingerprint verification** and for multi-biometrics.

G. Aggarwal, N. Ratha, R. M. Bolle, and R. Chellappa (2008). “Multibiometric cohort analysis for biometric fusion”. In IEEE Int.l Conf. on Acoustics, Speech and Signal Processing, 5224–7, 2008.

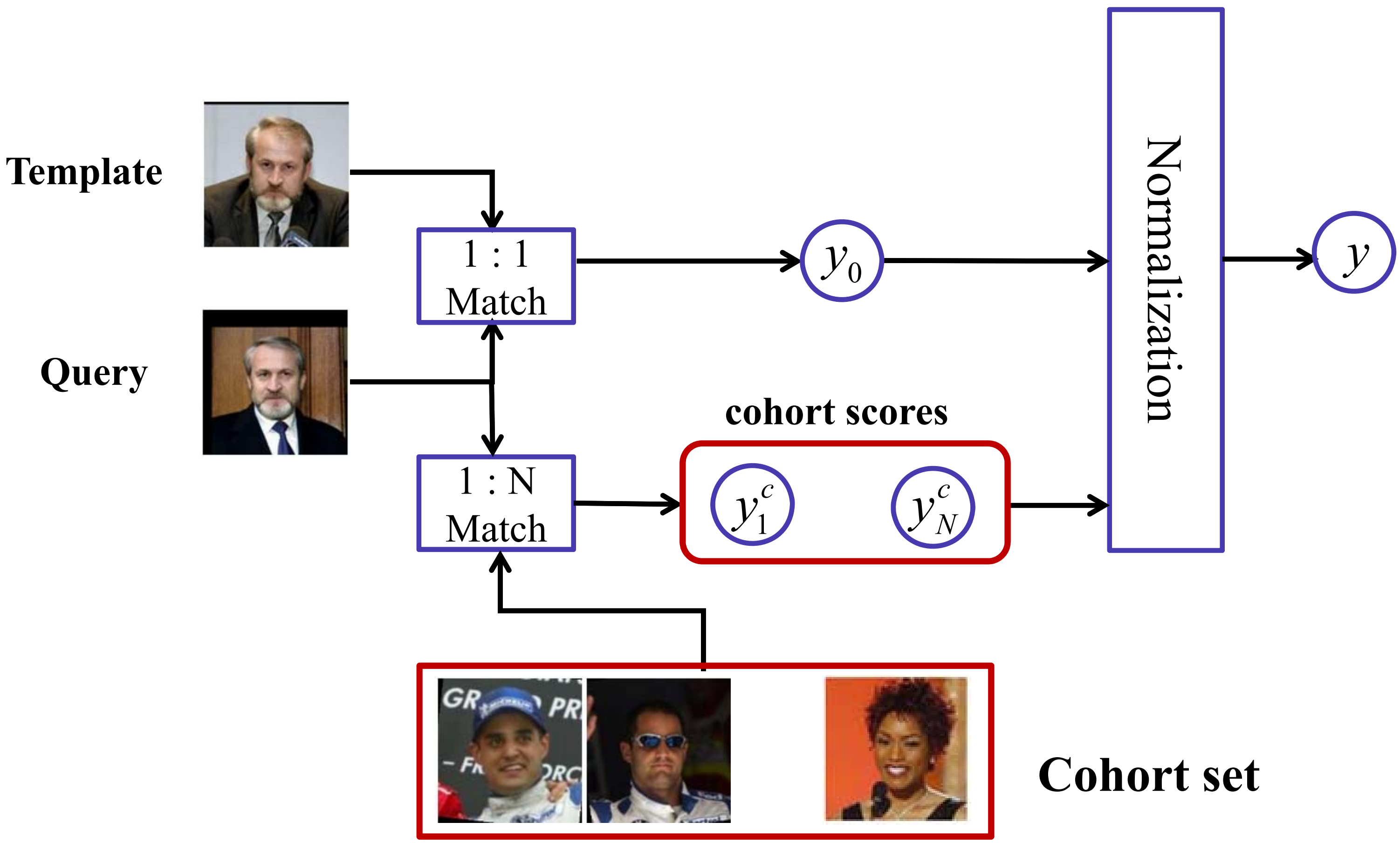
- **Two representative cohort normalization methods:**
  - T-norm (Test-norm …*with Gaussian assumption*)
  - Polynomial regression-based cohort normalization

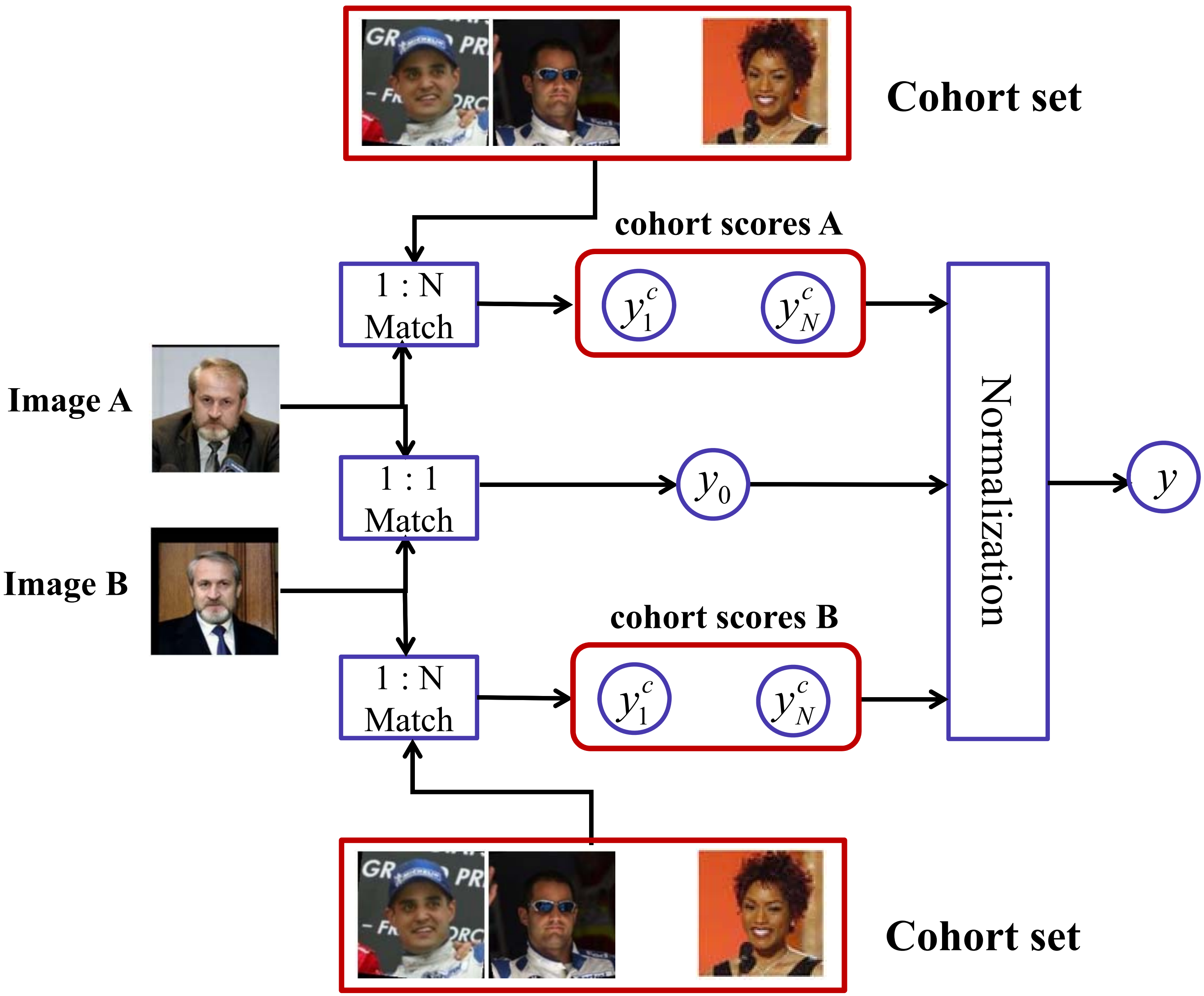
# Face pair matching

- Only two images are given, no other information is provided.
- Large variations may be found in the image pair.

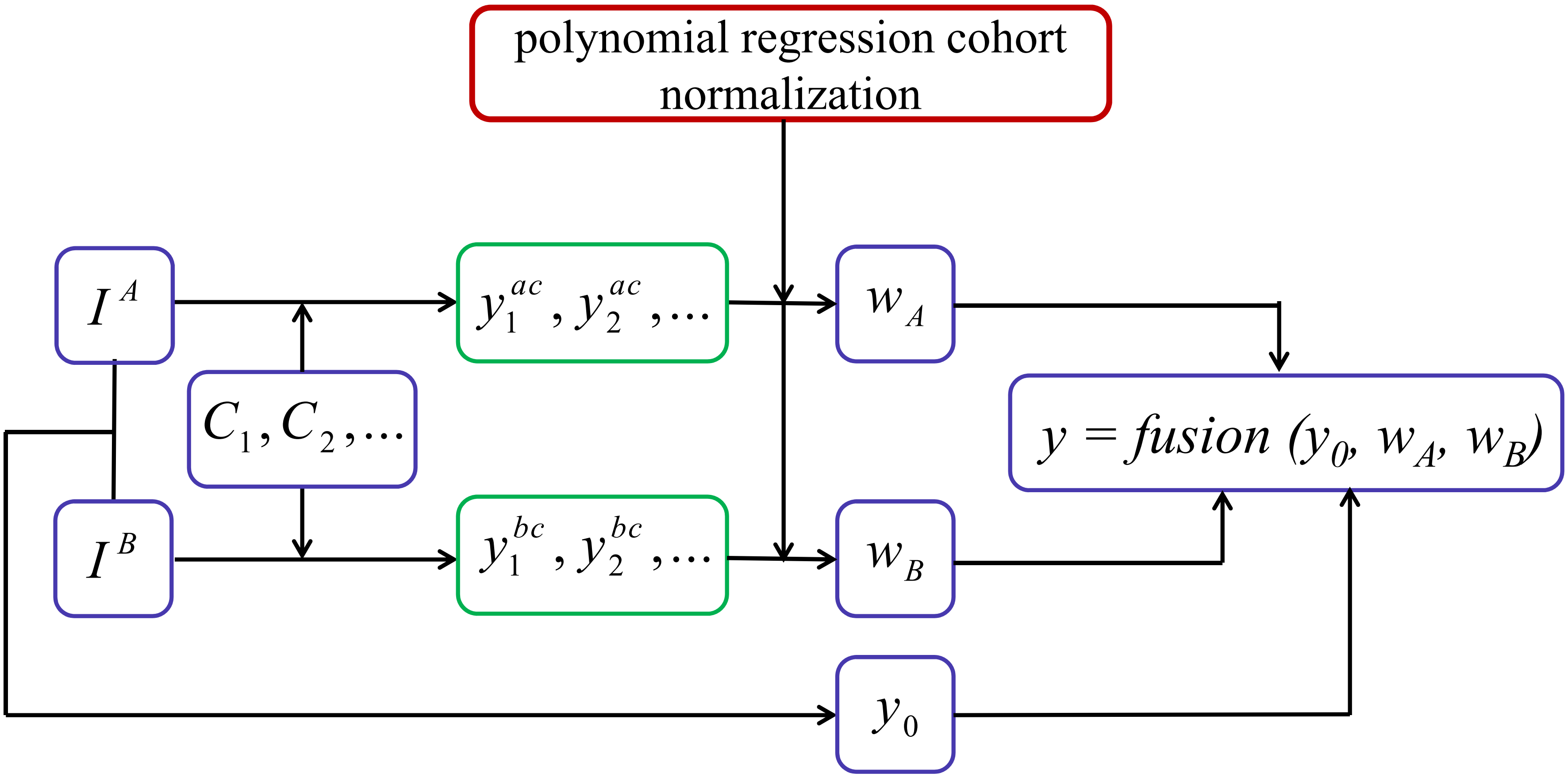


# Face verification




Cohort set
cohort scores A
1 : N
Match
$y_1^c$
$y_N^c$
Image A
1 : 1
Match
$y_0$
Normalization
$y$
Image B
cohort scores B
1 : N
Match
$y_1^c$
$y_N^c$
Cohort set

# Picture-specific Cohort normalization



# Picture-specific Cohort normalization

- **Database**: *Labeled Faces in the Wild*
- **Protocol**: image-restricted setting; 10-fold cross validation
- **Feature extraction**: Intensity, LBP, Gabor, SIFT
- **Matching score**: Euclidean or Hellinger distance

| | Intensity | Gabor | LBP | SIFT |
|---|---|---|---|---|
| Euclidean (no cohort) | 0.6502 | 0.6985 | 0.6500 | 0.7140 |
| Euclidean (with cohort) | **0.6830** | **0.7560** | **0.7443** | **0.7703** |
| Hellinger (no cohort) | 0.6497 | 0.7100 | 0.7132 | 0.7183 |
| Hellinger (with cohort) | **0.6913** | **0.7680** | **0.7707** | **0.7738** |

**Comparative verification accuracy with and without cohort normalization**

Vision Lab

# FACE RECOGNITION ACROSS PLASTIC SURGERY

M. Nappi, S. Ricciardi, M. Tistarelli, (2013); “Deceiving Faces: When Plastic Surgery Challenges Face Recognition” Image and Vision Computing, Vol. 54, pp. 71-82, 2016.

Y. Sun, M. Tistarelli, D. Maltoni (2013); “Structural Similarity based Image Quality Map for Face Recognition across Plastic Surgery” Proc IEEE 6th Int.l Conference on Biometrics: Theory, Applications and Systems - BTAS 2013 Washington DC, USA; September 29 - October 2, 2013.

UNIVERSITATIS TURRITANAE SACERENSIS

# Common cosmetic procedures

Botulinum toxin

Chemical peel

Dermal fillers

Dermoabrasion

# Common cosmetic procedures

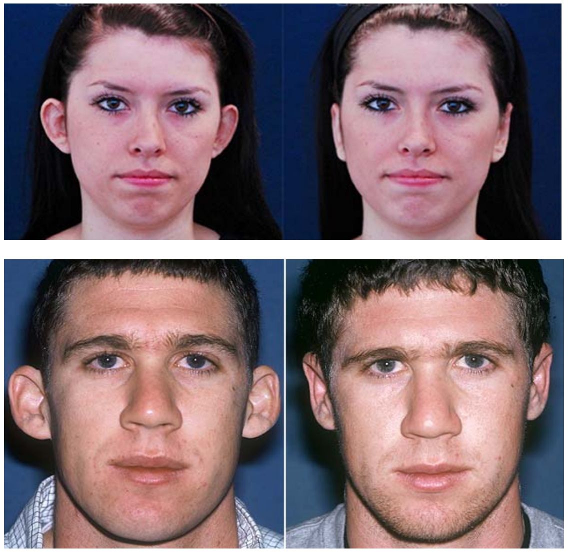
Rhinoplasty

Rhytidectomy (face lift)

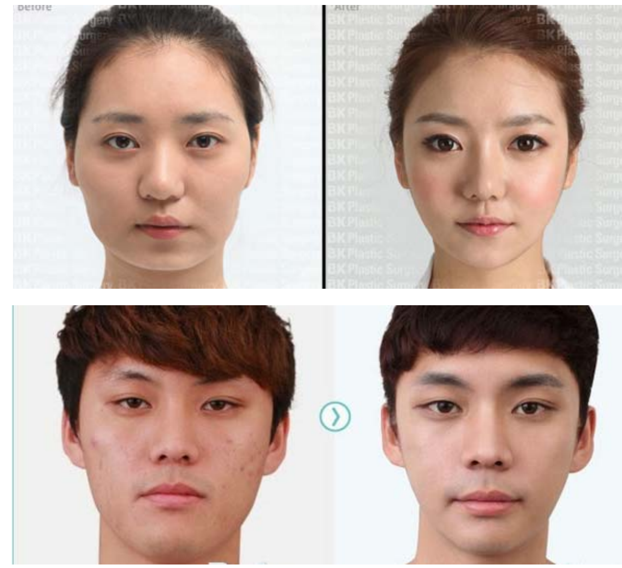
Blepharoplasty

Brow/Forehead lift

# Common cosmetic procedures



Otoplasty



Cheek bones reshaping

Mentoplasty

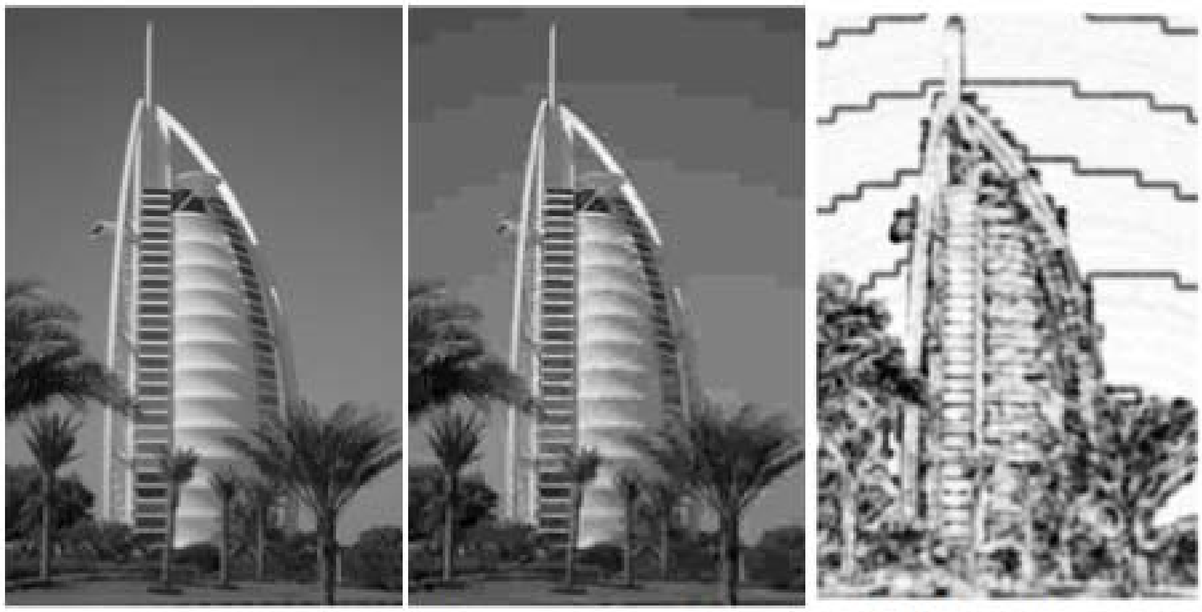
# Common cosmetic procedures

| Surgical procedure | Facial region | Spatial frequencies | Extension of face surface | Potential impact | Relative diffusion |
|---|---|---|---|---|---|
| **Botulinum toxine** | Forehead | High | Medium | Low to Medium | 52% |
| **Dermal fillers** | Periocular / smile lines | High | Limited | Medium | 19% |
| **Chemical peel** | Whole face | High | Wide | Low | 9% |
| **Dermoabrasion (*Resurfacing*)** | Whole face | High/Medium | Wide | Low to Medium | 0,6% |
| **Microdermoabrasion** | Whole face | High | Wide | (very) Low | 8% |
| **Nose reshaping (*Rhinoplasty*)** | Nose | Low | Limited | Medium | 1,8% |
| **Eyelid surgery (*Blepharoplasty*)** | Periocular re-gion | Medium | Limited | Low to Medium | 1,8% |
| **Facelift (*Rhytidectomy*)** | Whole face | Low to High | Wide | High | 1.1% |
| **Brow lift (*Forehead lift*)** | Forehead | Medium/High | Limited | Medium | 0.4% |
| **Chin surgery (*Mentoplasty*)** | Lower face re-gion | Low | Medium | High | 0.15% |
| **Cheekbones reshaping** | Zygomatic re-gion | Low | Medium | High | 0.1% |
| **Ear surgery (*Otoplasty*)** | Ears | Low | Limted | Low | 0.2% |

# Face recognition algorithms

| # | Reference | Dataset | Key Features | | | | | Algorithm |
|---|---|---|---|---|---|---|---|---|
| | | | GLOBAL | LOCAL | TEX | 3D | RR% | |
| 1 | Aggarwal et al. [26] | Plastic Surgery Face Database | N | Y | N | N | *77.9* | Part-wise and Sparse representation |
| 2 | Bhatt et al. [40] | Plastic Surgery Face Database | Y | Y | Y | N | *78.6* | Uniform Circular Local Binary Pattern (UCLBP) + Speeded Up Robust Features (SURF) + genetic alghorithm |
| 3 | De Marsico et al. [20,21] | Plastic Surgery Face Database | Y | Y | N | N | *70.0* | PIFS + region-based correlation index |
| 4 | El–said, Abol Atta [39] | Plastic Surgery Face Database | N | Y | N | N | *76.1* | geometrical descriptors of ROIs + minimum distance classifiers |
| 5 | Ibrahim et al. [14] | Plastic Surgery Face Database | Y | N | Y | N | *83.2* | PCA, KPCA, KFA, Gabor |
| 6 | Karuppusamy and Ponmuthuramalingam [44] | NA | N | Y | Y | N | - | Extended Uniform Circular Local Binary Pattern (EUCLBP) + SIFT + Particle Swarm Optimization (PSO) |
| 7 | Lakshmiprabha et al. [34] | Plastic Surgery Face Database | N | Y | Y | N | *74.4* | Gabor/LBP + PCA + Euclidean Distance |
| 8 | Liu et al. [29] | Plastic Surgery Face Database | Y | Y | Y | N | *86.1* | Gabor Patch classifiers via Rank-Order list Fusion (GPROF) |
| 9 | Mun and Deorankar [33] | Web available Before/After Surgery photos | Y | Y | Y | N | - | Multimodal biometric featuresPCA (face)+LBP (periocular region) |
| 10 | Singh, Vatsa and Noore [4] | NA | Y | Y | Y | N | *40* | PCA, FDA, GF, LFA, LBP, GNN |
| 11 | Singh et al. [11] | Plastic Surgery Face Database | Y | Y | Y | N | *40* | PCA, FDA, LFA, CLBP, SURF, GNN |
| 12 | Sun et al. [36] | Plastic Surgery Face Database | Y | Y | Y | N | *77.5* | Structural Similarity (SSIM) index +weighted patch fusion |
| 13 | Verghis et Bhuvaneshwari [16] | Plastic Surgery Face Database | Y | Y | N | Y | *87.3* | Evolutionary granular algorithm + SIFT and EUCLBP |

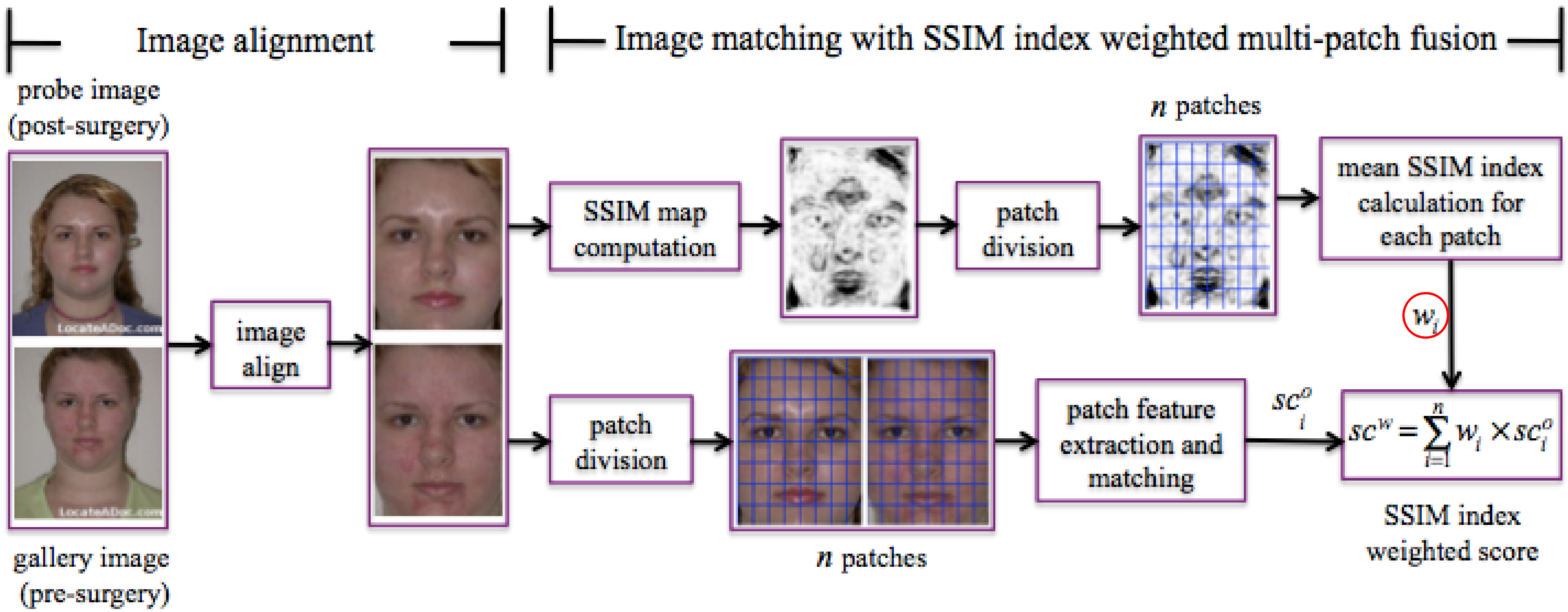
# Structural Similarity Map



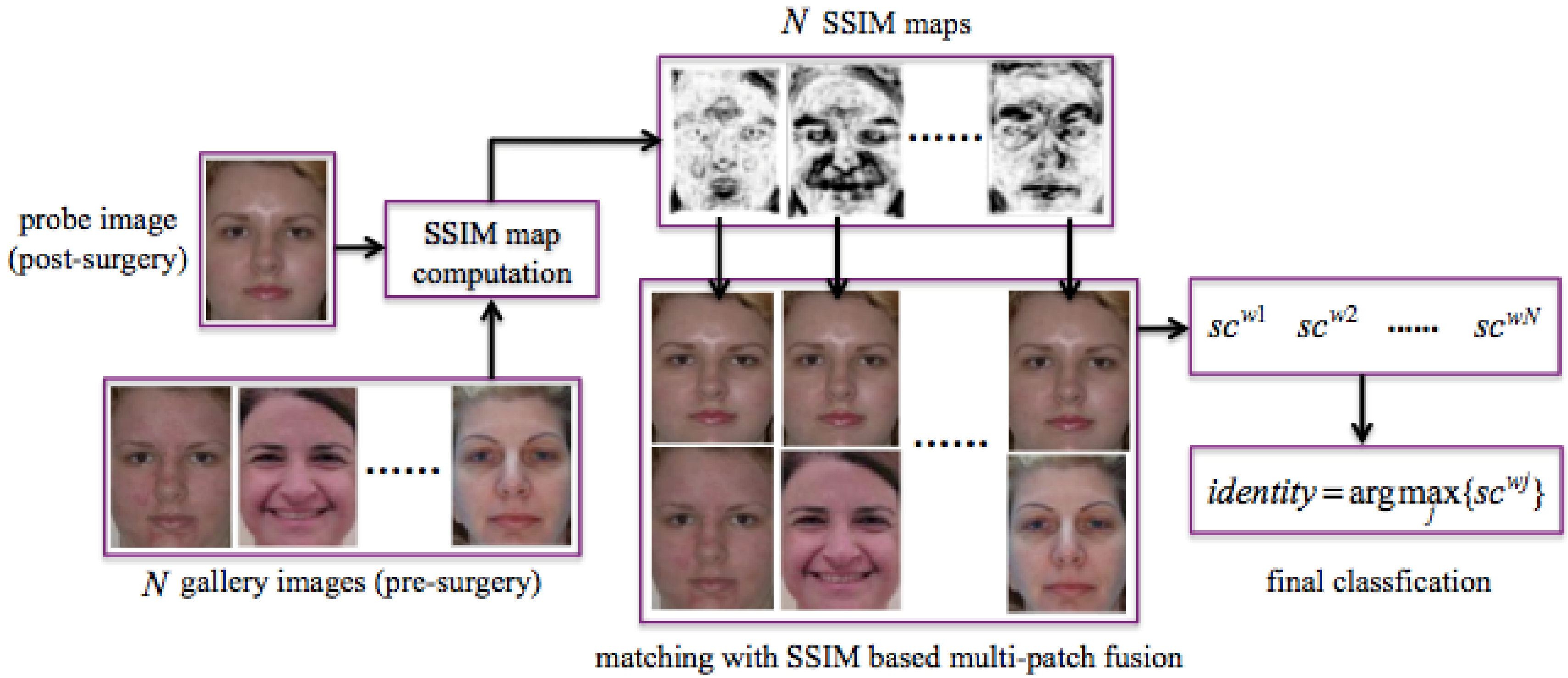
Reference image JPEG compressed image SSIM quality map

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

# Face recognition using SSIM

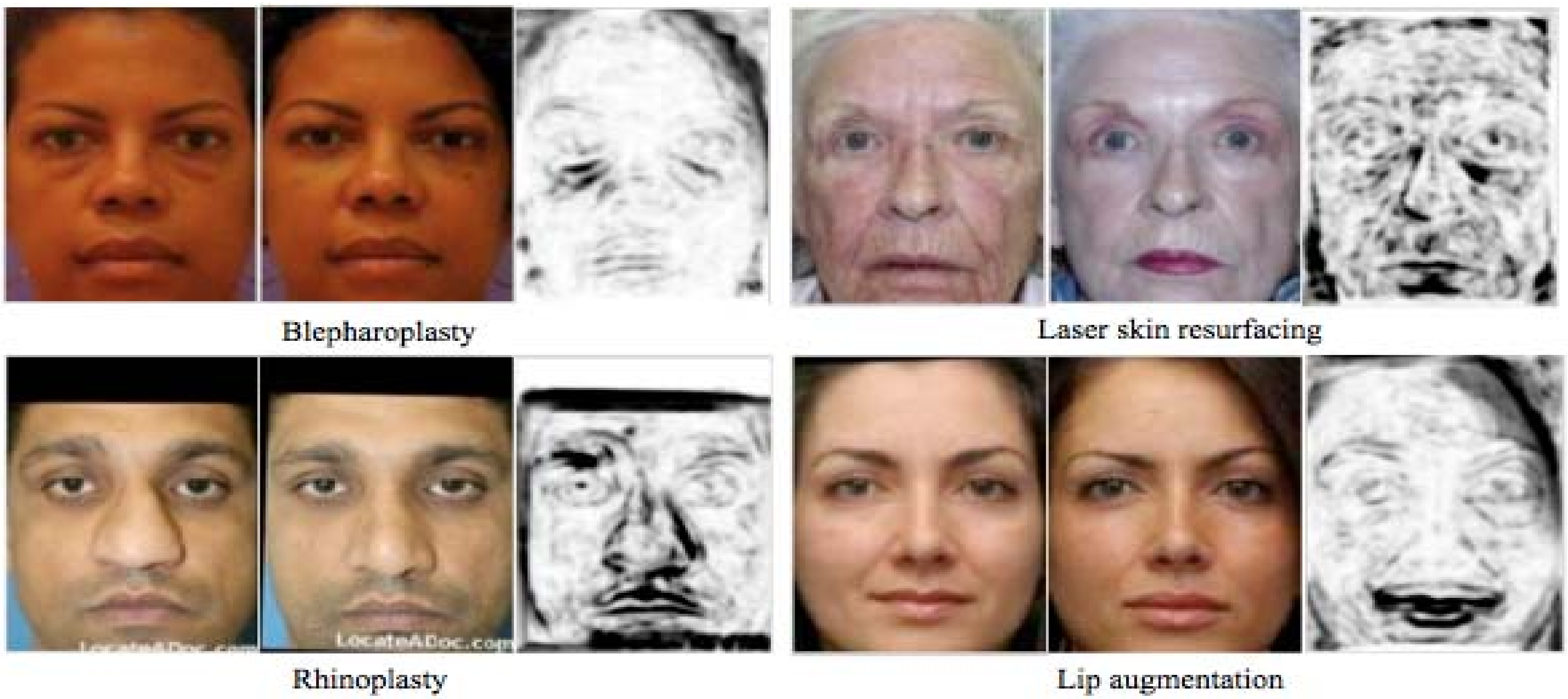


# Face recognition using SSIM



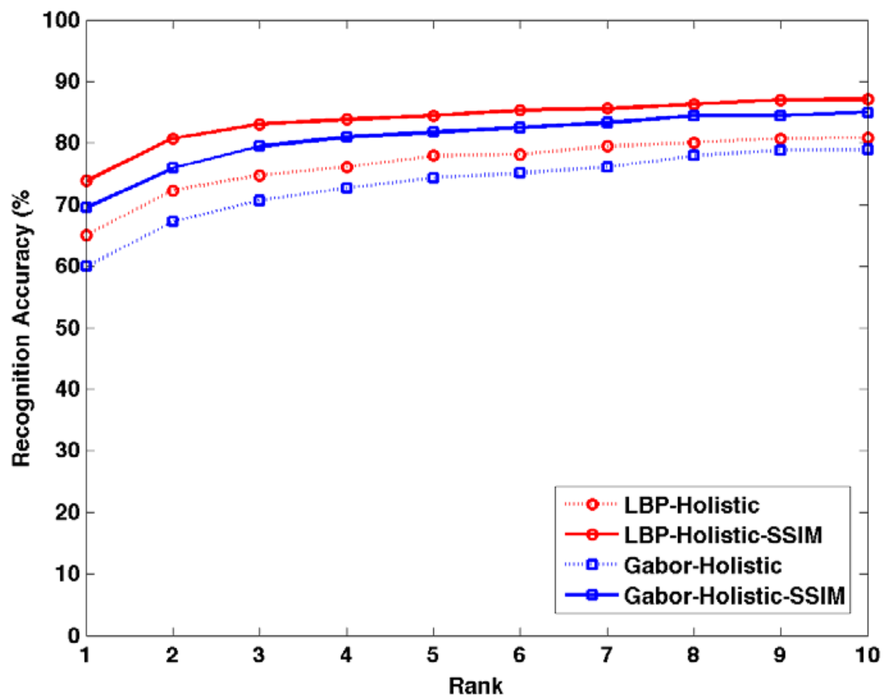
Vision Lab

# SSIM examples



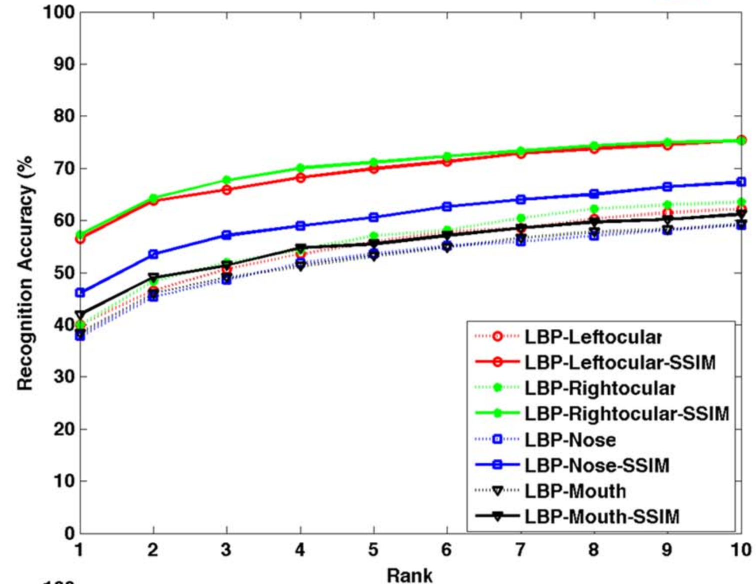
# Experimental results

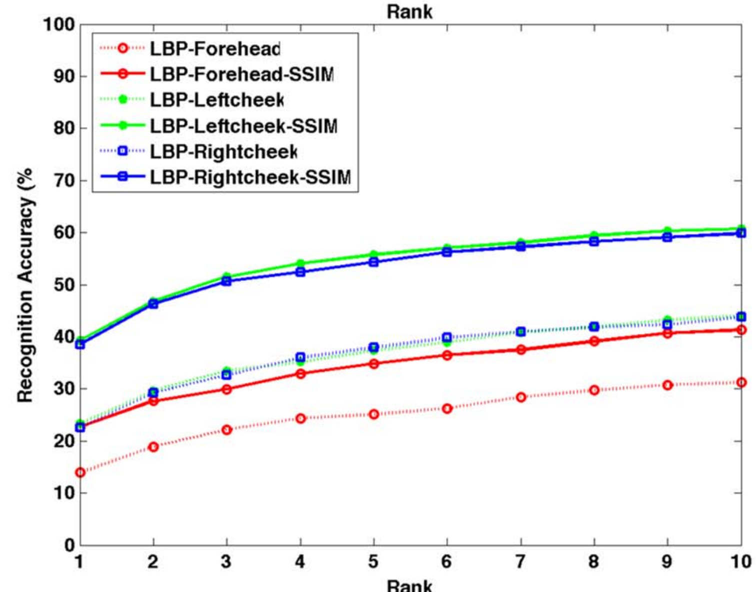
**Plastic surgery database containing 576 images of 784 subjects taken from the web.**



# Component-based matching

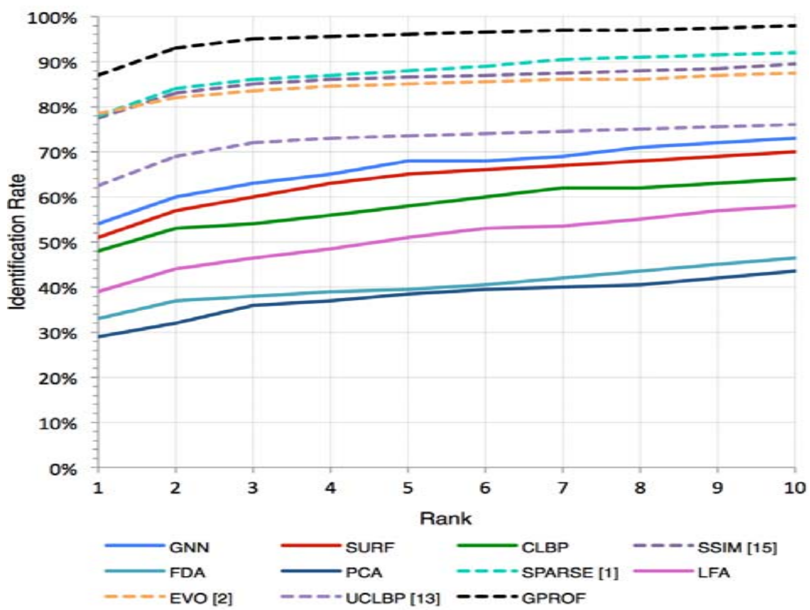
| | |
|---|---|
| Forehead | 48×64 |
| Oculars | 48×64 |
| Nose | 48×48 |
| Cheeks | 48×48 |
| Mouth | 48×64 |

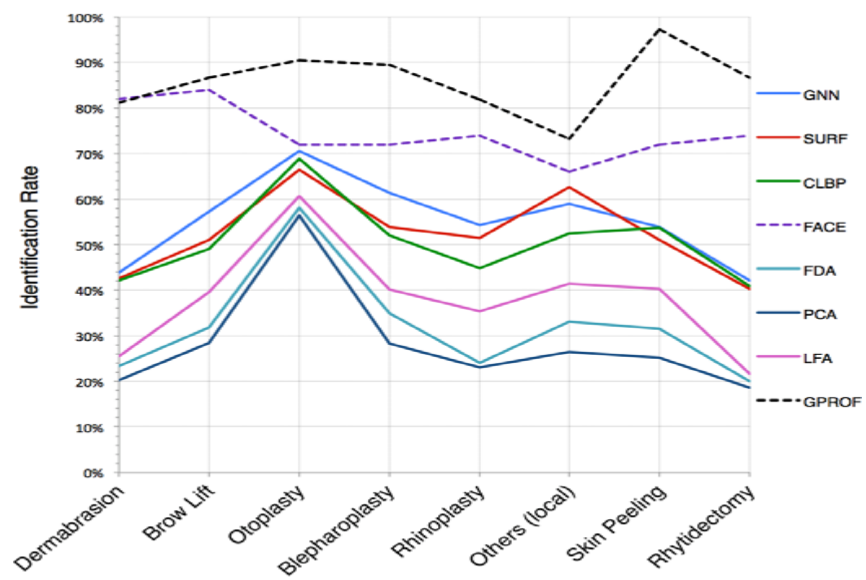


# Performance

Vision Lab

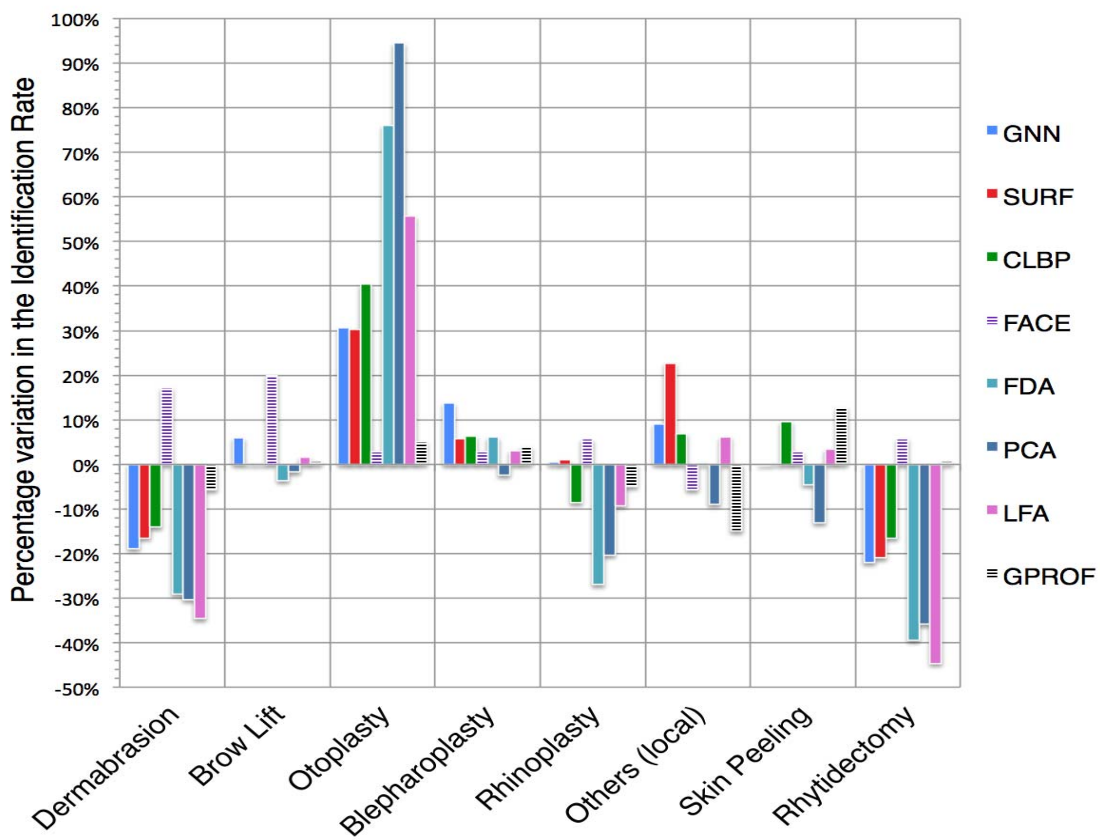



Comparison of the Cumulative Match Characteristic curves computed from eleven different face recognition algorithms applied to the same plastic surgery database. Dashed lines refer to region-based approaches, while solid lines refer to holistic approaches. The CMC curves are those reported in their original research papers.

Identification error, as reported by eight different recognition algorithms, categorized by eight different surgical procedures. The six leftmost procedures are local, while the two rightmost procedures are global.

# Performance



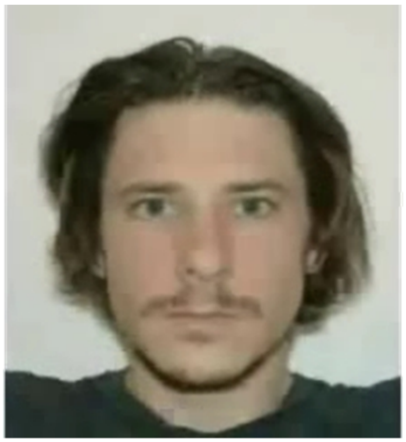
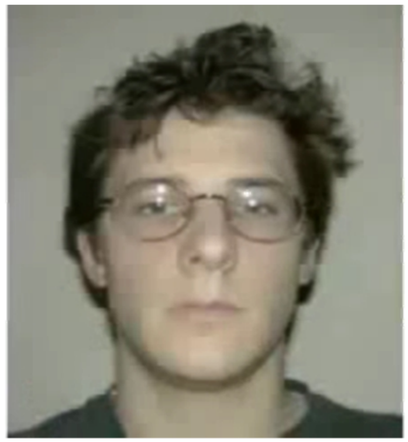
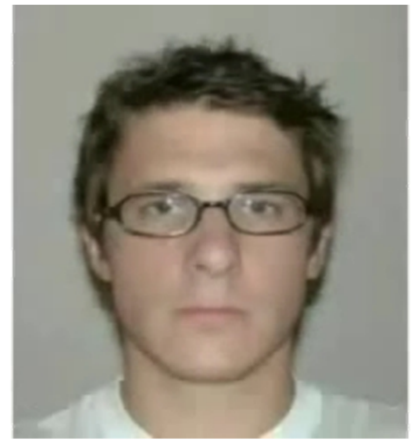

Comparison of overall vs. procedure-wise performance of eight different algorithms. The identification rate is normalized to 1.

# DEALING WITH AGE PROGRESSION

M. Ortega, L. Brodo, M. Bicego, M. Tistarelli, "Measuring changes in face appearance through aging,", Proc. of IEEE Computer Vision and Pattern Recognition Workshop, pp. 107-113, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (2009)

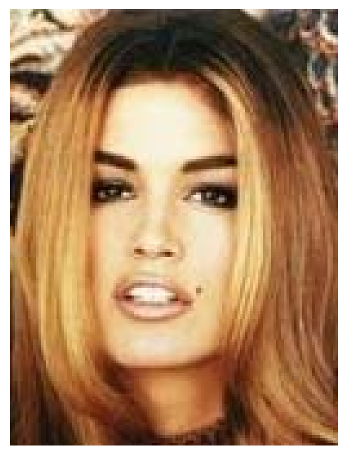
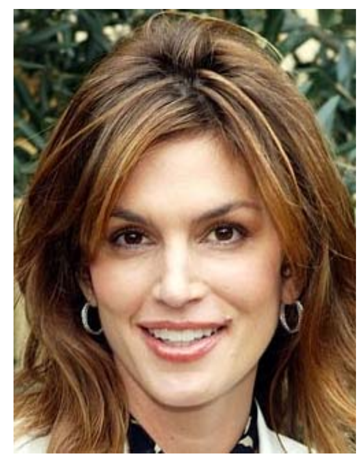
# Aging effects

**Time duration: 2 years**

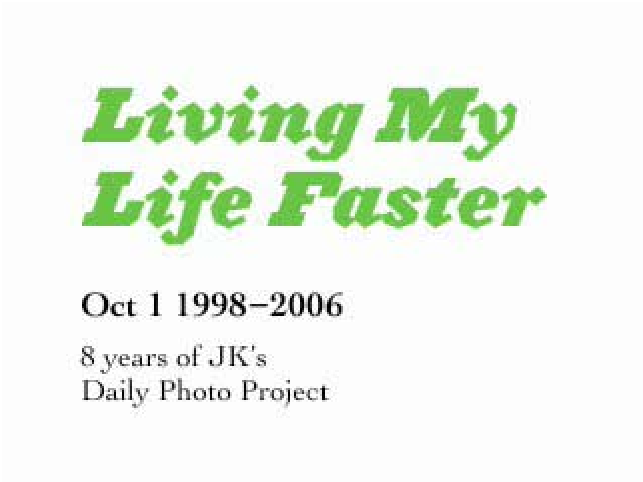




**Time duration: several years**

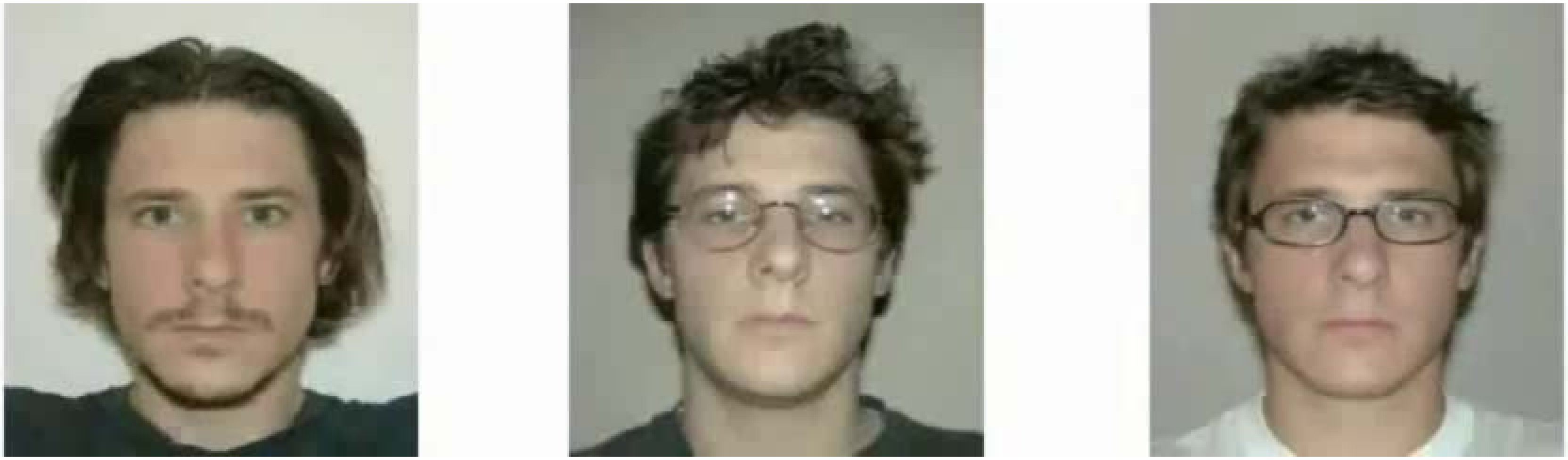


# Aging … over time



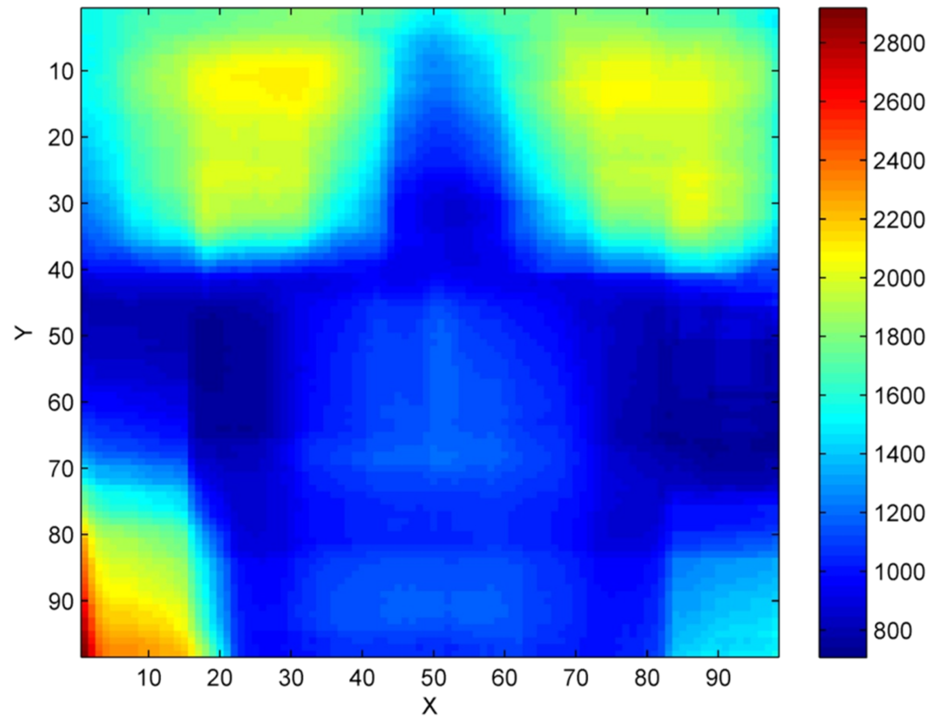

# Aging … over time



$$d_{I_1,I_2}(x,y) = \frac{1}{2}\left(\frac{1}{|P_{I_1}|}\sum_{p\in P_{I_1}}\omega(p) + \frac{1}{|P_{I_2}|}\sum_{q\in P_{I_2}}\omega(q)\right)$$

# Photometric effects

## Time evolution of facial features over 4 years

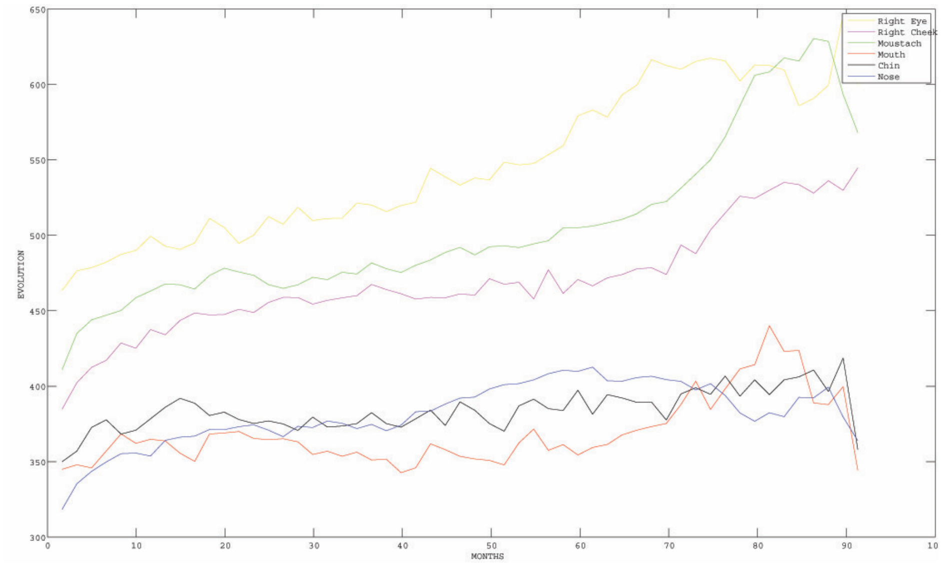



$$E(x,y,t) = \frac{1}{T}\sum_{i=1}^{T} d_{I_i,I_{i+t}}(x,y)$$

M. Ortega, L. Brodo, M. Bicego, M. Tistarelli, "Measuring changes in face appearance through aging,", Proc. of IEEE Computer Vision and Pattern Recognition Workshop, pp. 107-113, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (2009)

# Photometric effects

## Time evolution of facial features over 8 years

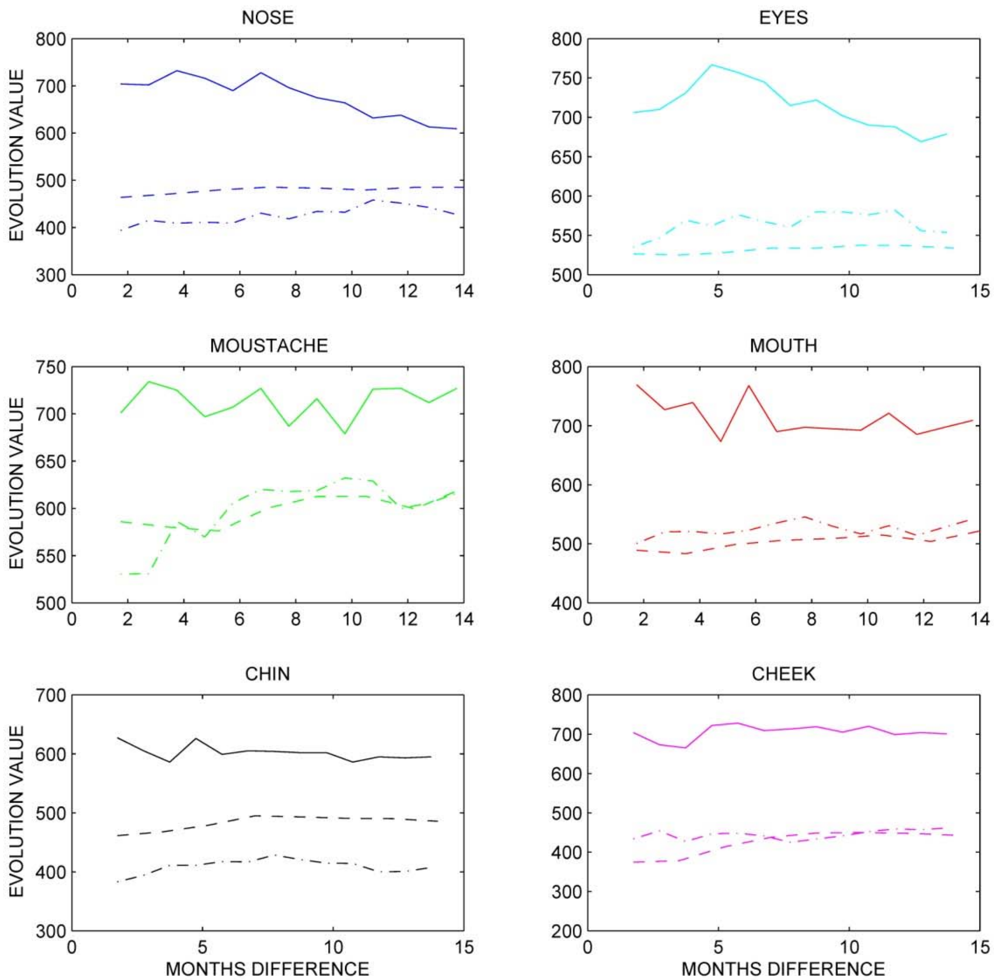



$$E(L,t) = \frac{1}{w_L h_L} \sum_{(x,y)\in L} E(x,y,t)$$

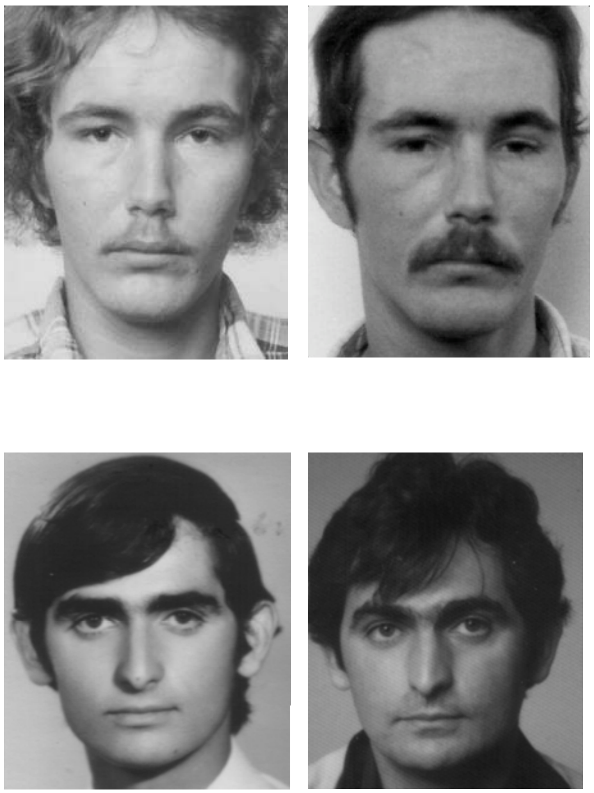
M. Ortega, L. Brodo, M. Bicego, M. Tistarelli, "Measuring changes in face appearance through aging,", Proc. of IEEE Computer Vision and Pattern Recognition Workshop, pp. 107-113, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (2009)

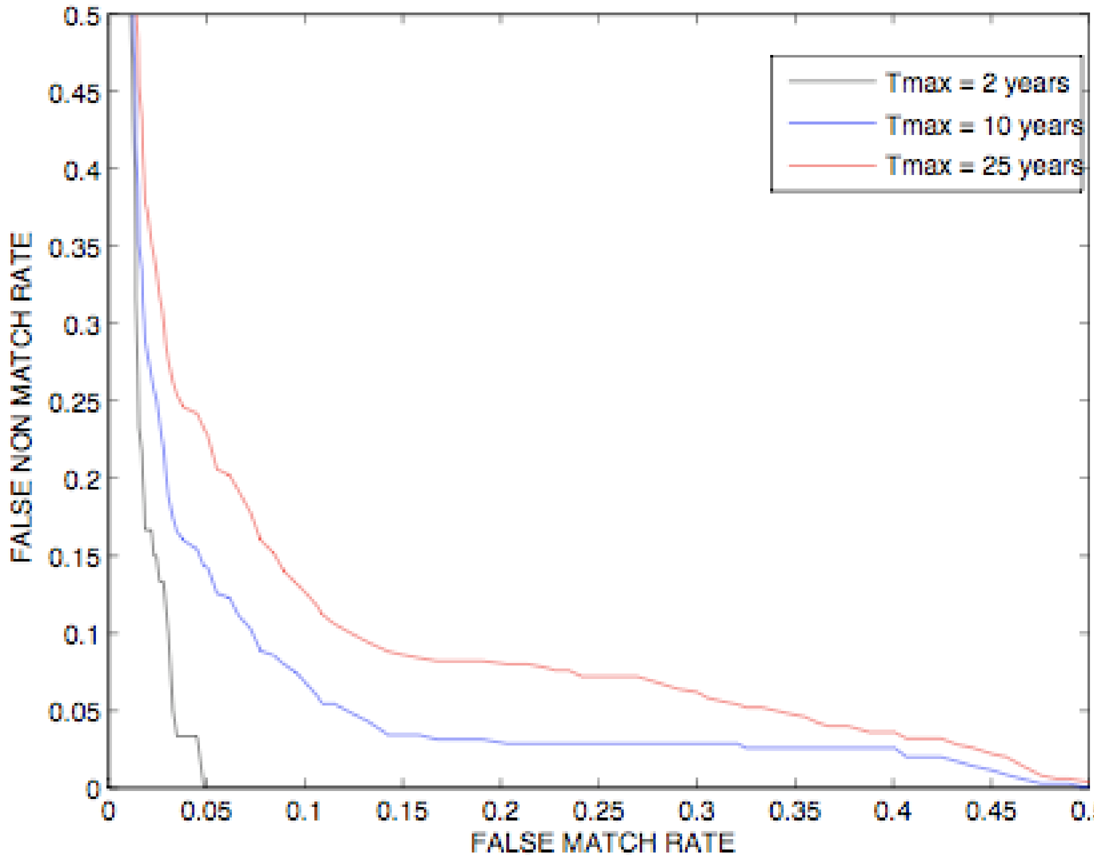
# Photometric effects

## Comparative time evolution of features for different subjects



# Face matching across age






M. Ortega, L. Brodo, M. Bicego, M. Tistarelli, "Measuring changes in face appearance through aging,", Proc. of IEEE Computer Vision and Pattern Recognition Workshop, pp. 107-113, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (2009)

# 3D FACE RECOGNITION

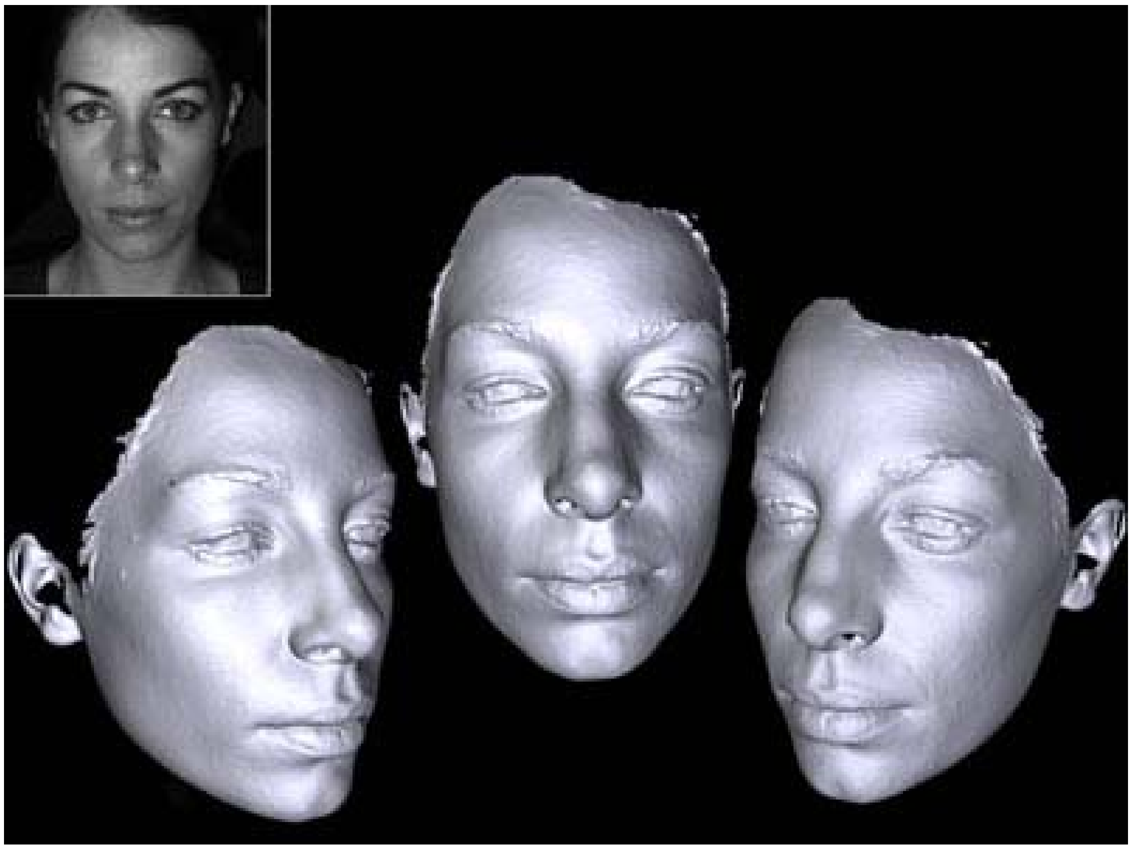

Cadoni M., Grosso E., Lagorio A., “Large scale face identification by combined iconic features and 3D joint invariant signatures”, Image and Vision Computing, Vo. 52, pp. 42-55, 2016.

Cadoni M., Grosso E., Lagorio A., Tistarelli M., “Blending 2D and 3D Face Recognition”, T. Bourlai Ed. *Face Recognition Across the Imaging Spectrum*, pp. 305-331, Springer, 2016.

Cadoni M., Grosso E., Lagorio A., Tistarelli M., “From 3D Faces to Biometric Identities”, Proc. of BioId European Workshop, pp. 156-167, LNCS 6583, Springer, 2011.

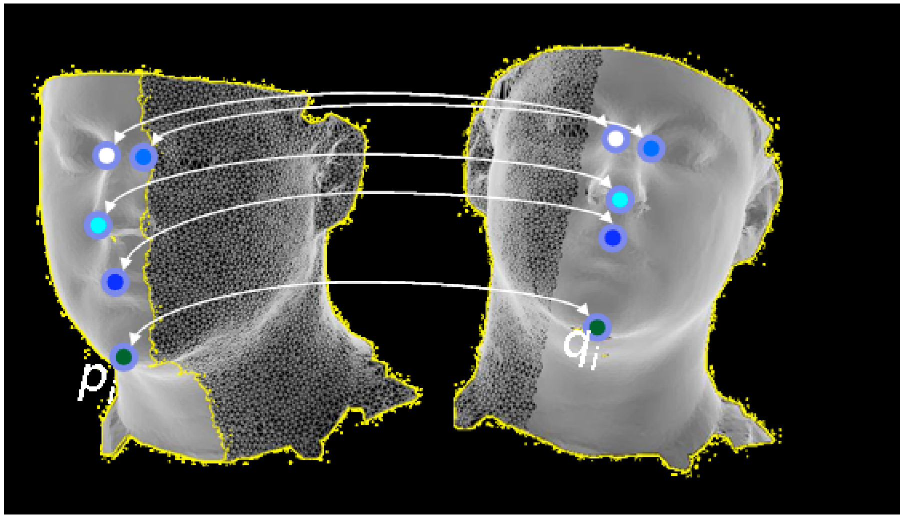
Cadoni M., Bicego E., Grosso E., “3D Face Recognition Using Joint Differential Invariants”, Proc. of Third International Conference on Biometrics, ICB 2009, pp. 279-288, LNCS 5558, Springer, 2009.

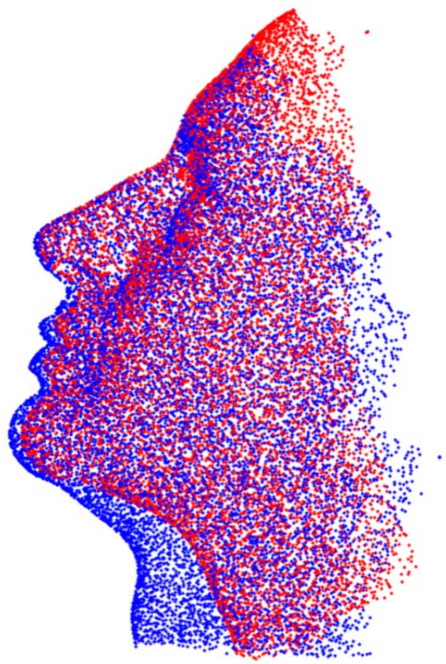
# 3D face recognition: from ill-posed to well-posed



# 3D face matching

- Recognition of faces from 3D data can be achieved by pairing a set of points from two individuals and measuring the goodness of fit.
- This process requires to identify anchor points on the faces



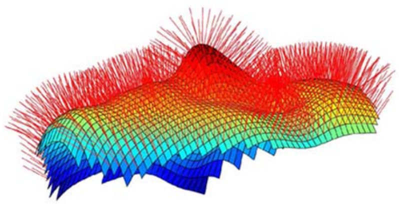
**Point clouds:**
- ICP gives the distance
- Hausdorff
- Too many points

PC

TEX

**Texture:**
- Fusion with shape

SN



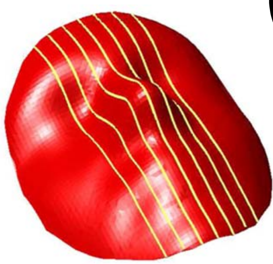
**Surface normals:**
- Too many …

PCA

RI

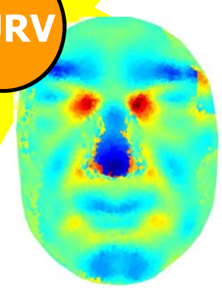
**Range images:**
- Benefit from 2D literature
- Easy to fuse with texture
- Applicable to 2.5D only

PRO



**Face profile:**
-Sparse
-Easy to compare
-Not fully descriptive

CURV



**Curvature:**
- Useful for landmarks detection
- Sensitive to noise and the quality of data

# 3D shape matching

Vision Lab

| Author, year, reference | Persons in dataset | Images in dataset | Image size | 3D face data | Core matching algorithm | Reported performance |
|---|---|---|---|---|---|---|
| Cartoux, 1989 [12] | 5 | 18 | Not available | Profile, surface | Minimum distance | 100% |
| Lee, 1990 [26] | 6 | 6 | $256 \times 150$ | EGI | Correlation | None |
| Gordon, 1992 [21] | 26 train 8 test | 26 train 24 test | Not available | Feature vector | Closest vector | 100% |
| Nagamine, 1992 [39] | 16 | 160 | $256 \times 240$ | Multiple profiles | Closest vector | 100% |
| Achermann, 1997 [3] | 24 | 240 | $75 \times 150$ | Range image | PCA, HMM | 100% |
| Tanaka, 1998 [52] | 37 | 37 | $256 \times 256$ | EGI | Correlation | 100% |
| Achermann, 2000 [2] | 24 | 240 | $75 \times 150$ | Point set | Hausdorff distance | 100% |
| Chua, 2000 [17] | 6 | 24 | Not available | Point set | Point signature | 100% |
| Hesher, 2003 [22] | 37 | 222 | $242 \times 347$ | Range image | PCA | 97% |
| Lee, 2003 [27] | 35 | 70 | $320 \times 320$ | Feature vector | Closest vector | 94% at rank 5 |
| Medioni, 2003 [34] | 100 | 700 | Not available | Point set | ICP | 98% |
| Moreno, 2003 [38] | 60 | 420 | 2.2K points | Feature vector | Closest vector | 78% |
| Pan, 2003 [42] | 30 | 360 | 3K points | Point set, range image | Hausdorff and PCA | 3–5% EER, 5–7% EER |
| Lee, 2004 [28] | 42 | 84 | $240 \times 320$ | Range, curvature | Weighted Hausdorff | 98% |
| Lu, 2004 [30] | 18 | 113 | $240 \times 320$ | point set | ICP | 96% |
| Russ, 2004 [49] | 200 FRGC v1 | 468 | $480 \times 640$ | Range image | Hausdorff distance | 98% verification |
| Xu, 2004 [57] | 120 (30) | 720 | Not available | Point set + feature vector | Minimum distance | 96% on 30, 72% on 120 |
| Bronstein, 2005 [11] | 30 | 220 | Not available | Point set | "canonical forms" | 100% |
| Chang, 2005 [16] | 466 FRGC v2 | 4007 | $480 \times 640$ | Point set | multi-ICP | 92% |
| Gökberk, 2005 [20] | 106 | 579 | Not available | Multiple | Multiple | 99% |
| Lee, 2005 [29] | 100 | 200 | Various | Feature vector | SVM | 96% |
| Lu, 2005 [31] | 100 | 196 probes | $240 \times 320$ | Surface mesh | ICP, TPS | 89% |
| Pan, 2005 [41] | 276 FRGC v1 | 943 | $480 \times 640$ | Range image | PCA | 95%, 3% EER |
| Passalis, 2005 [44] | 466 FRGC v2 | 4007 | $480 \times 640$ | Surface mesh | Deformable model | 90% |
| Russ, 2005 [50] | 200 FRGC v1 | 398 | $480 \times 640$ | Range image | Hausdorff distance | 98.5% |

**Bowyer et al. CVIU 101: 1-15 (2006)**

# Fusing 3D shape and texture

| Author, year, reference | Persons in dataset | Images in dataset | Image size | 3D face data | Core matching algorithm | Reported performance |
|---|---|---|---|---|---|---|
| Lao, 2000 [25] | 10 | 360 | 480 × 640 | Surface mesh | Minimum distance | 91% |
| Beumier, 2001 [4] | 27 gallery 29 probes | 81 gallery, 87 probes | Not available | Multiple profiles | Minimum distance | 1.4% EER |
| Wang, 2002 [56] | 50 | 300 | 128 × 512 | Feature vector | SVM, DDAG | >90% |
| Bronstein, 2003 [10] | 157 | Not available | 2250 points | Range, point set | PCA | Not reported |
| Chang, 2003 [14] | 200 (275 train) | 951 | 480 × 640 | Range image | PCA | 99% 3D + 2D, 93% 3D only |
| Tsalakanidou, 2003 [55] | 40 | 80 | 100 × 80 | Range image | PCA | 99% 3D + 2D, 93% 3D only |
| Godil, 2004 [19] | 200 | 400 | 128 × 128 | Range image | PCA | 82% rank 1 |
| Papatheodorou, 2004 [43] | 62 | 806 | 10,000 points | Point set | ICP | 100–66% |
| Tsalakanidou, 2004 [54] | 50 | 3000 | 571 × 752 | Range image | EHHM per mode | 4% EER |
| Hüsken, 2005 [23] | 466 | 4,007 FRGC v.2 | 480 × 640 | hier. graph | graph match | 93% verification at 0.01 FAR |
| Lu, 2005 [32] | 100 | 598 | 320 × 240 | Point set | ICP, LDA | 91% |
| Maurer, 2005 [33] | 466 | 4007 FRGC v.2 | 480 × 640 | Surface mesh | ICP, Neven | 87% verification at 0.01 FAR |

**Bowyer et al. CVIU 101: 1-15 (2006)**

# 3D Shape invariants

Multiscale features extracion

3D face registration using geometrical invariants

Compute the face similarity

Cadoni M., Grosso E., Lagorio A., "Large scale face identification by combined iconic features and 3D joint invariant signatures", Image and Vision Computing, Vo. 52, pp. 42-55, 2016.
Cadoni M., Bicego E., Grosso E., "3D Face Recognition Using Joint Differential Invariants", Proc. of Third International Conference on Biometrics, ICB 2009, pp. 279-288, LNCS 5558, Springer, 2009.

# 3D face *registration*

- **For each triplet (*p1, p2, p3*) of features points a set of nine invariants is computed:**

$$(I_1, I_2, I_3, J_1, J_2, J_3, \tilde{J}_1, \tilde{J}_2, \tilde{J}_3)$$

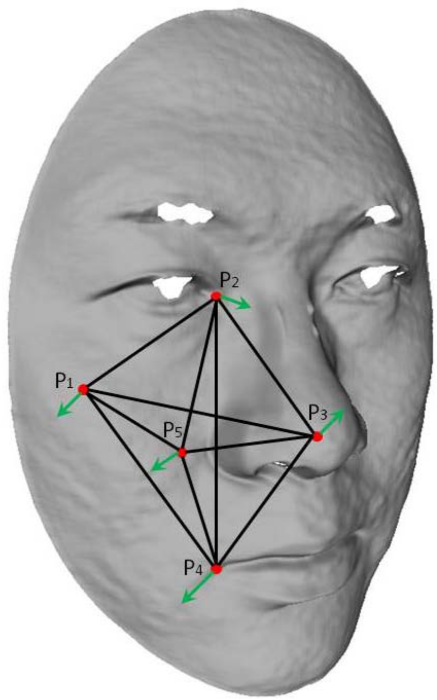
  - 3 of differential order zero:

$$I_1 = \|p_2 - p_1\|$$

  - 6 of differential order one:

$$u = \frac{p_2 - p_1}{\|p_2 - p_1\|} \qquad v_t = \frac{(p_2 - p_1) \wedge (p_3 - p_1)}{\|(p_2 - p_1) \wedge (p_3 - p_1)\|}$$

$$J_k = \frac{(v_t \wedge v) \cdot v_k}{v_t \cdot v_k} \qquad \tilde{J}_k = \frac{u \cdot v_k}{v_t \cdot v_k}$$



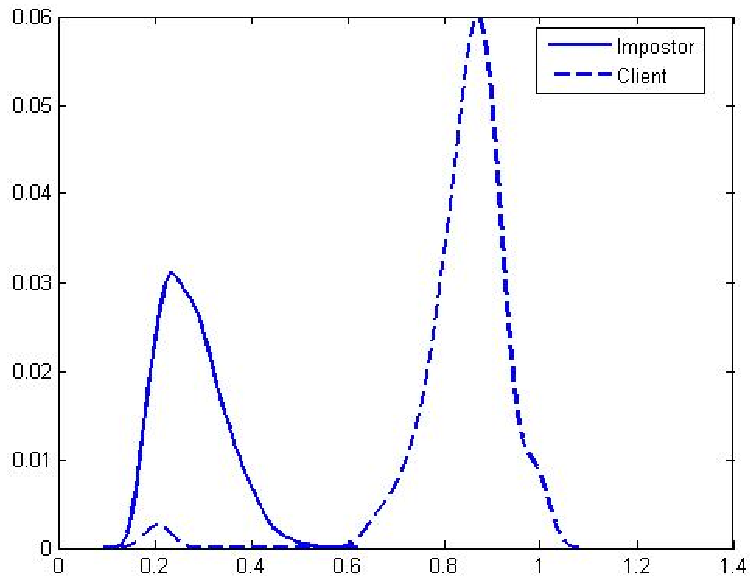
# Results

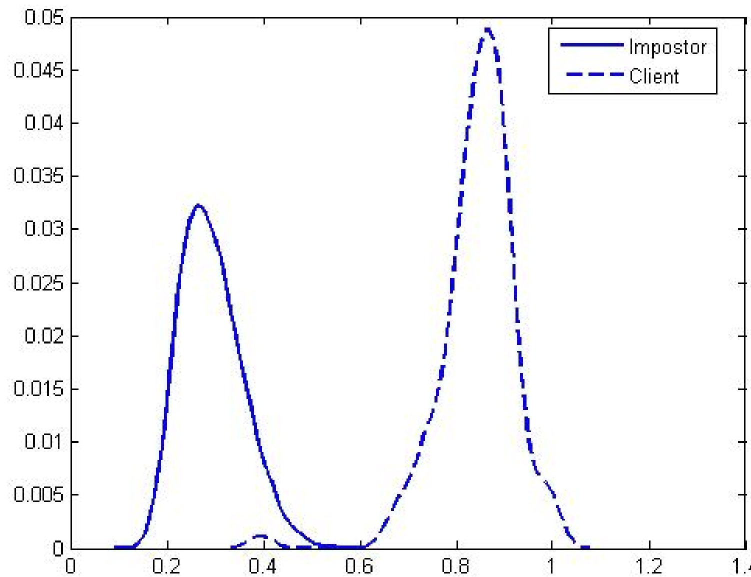
| Exp. | FR | AR | TP | FP | FN | TN | Acc. |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0.981 | 103 | 0 | 2 | 10920 | 0.9998 |
| 2 | 0 | 0.924 | 97 | 0 | 8 | 10920 | 0.9992 |
| 3 | 0 | 0.99 | 104 | 0 | 1 | 10920 | 0.9999 |

**FR: Failed registration**
**AR: Authentication rate**
**TP: True positives**
**FP: False positives**
**FN: False negatives**
**TN: True negatives**
**Acc.: Accuracy = (TP + TN)/ (P + N)**


- M. Cadoni, M. Bicego, E. Grosso, "3D face recognition using joint differential invariants", *Proc. Int. Conf. on Biometrics (ICB2009)*, pp. 279-288, (2009)
- Marinella Cadoni, Enrico Grosso, Andrea Lagorio, Massimo Tistarelli: "From 3D Faces to Biometric Identities". Proc. of BIOID 2011: 156-167, 2011

# Results



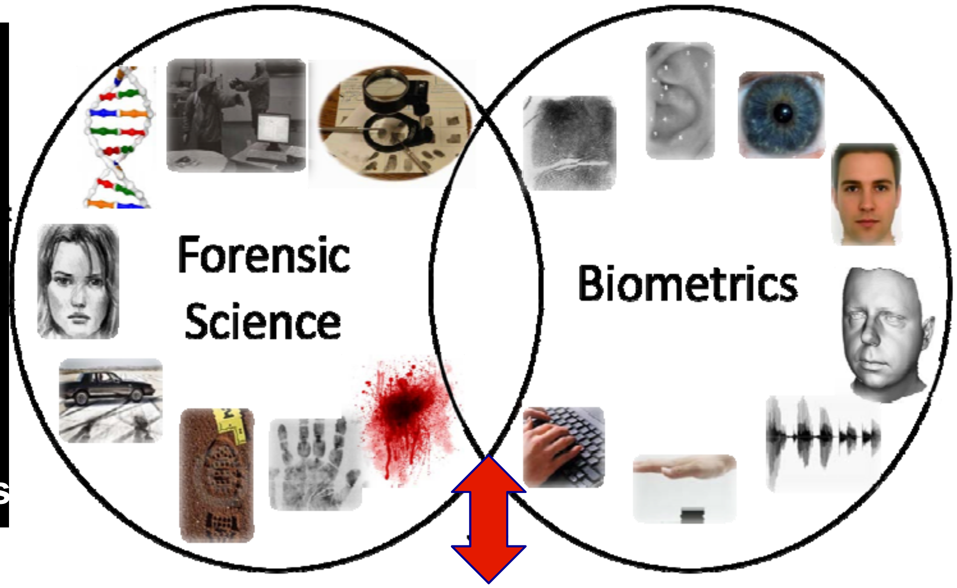

Impostor and client distributions for experiment 1 (left), and 3 (right)


- **M. Cadoni, M. Bicego, E. Grosso, "3D face recognition using joint differential invariants",** ***Proc. Int. Conf. on Biometrics (ICB2009)*****, pp. 279-288, (2009)**
- **Marinella Cadoni, Enrico Grosso, Andrea Lagorio, Massimo Tistarelli: "From 3D Faces to Biometric Identities". Proc. of BIOID 2011: 156-167, 2011**

# APPLICATIONS IN FORENSIC SCIENCE

# Biometrics and Forensic Science

- Latent fingerprint
- Latent palmprint
- Fibers
- Explosive residue
- Paint chips
- DNA
- Tire marks
- Shoe prints
- Bite marks
- Scars Marks Tatoos




- Fingerprint
- Palmprint
- 2D Face
- 3D Face
- Iris
- Speech
- Signature
- Gait
- Ear
- Keystroke

- Improve matching accuracy
- Automated matching
- Minimize human bias and sources of human error
- Validate basis for evidence

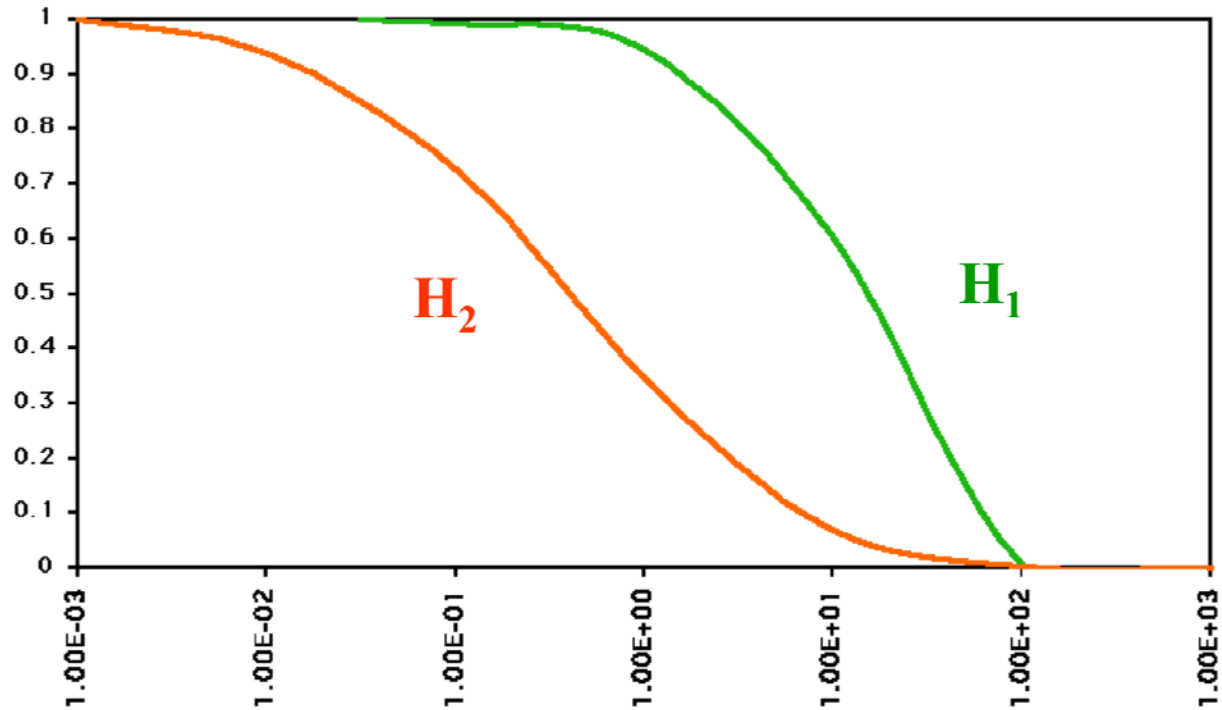
**Forensics:** Use of "trace evidence" from the crime scene to identify objects or persons

**Biometrics:** Identification of living persons by their traits in "real-time"

# Biometrics and Forensic Science

- Tippet plots
  - Representation of the results proposed in the field of interpretation of forensic DNA analysis

**Vertical axis**: Indication of the probability that the result of the experiment exceeds a given value of likelihood ratio




**Horizontal axis**: Graduation with increasing values of likelihood ratios

# Working environments

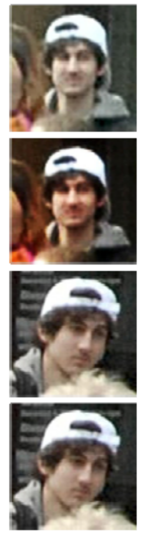
# Working environments


# Videos & Sketches

Vision Lab


**First Composite Sketch Created in 1987** **Later Sketch which became widely circulated**

The search for the Unabomber generated one of the most well known composite sketches in recent history. The sketch is based on a single 1987 witness, who saw an individual drop off a package containing a bomb. Although the identity is obscured by a hooded sweatshirt and dark glasses, there are some very obvious differences between the sketch and Ted Kaczynski, the man who was later arrested for the crimes. The man in the sketch seems to have a very narrow nose, yet Kacyknski has a prominent, almost bulbous nose. Kaczynski also has a broad chin and prominent age lines extending from above the flare of his nose down. The basic shape of the face is different in the two sketches, but neither accurately captures Kaczynski's likeness.



# Face “Recognition”

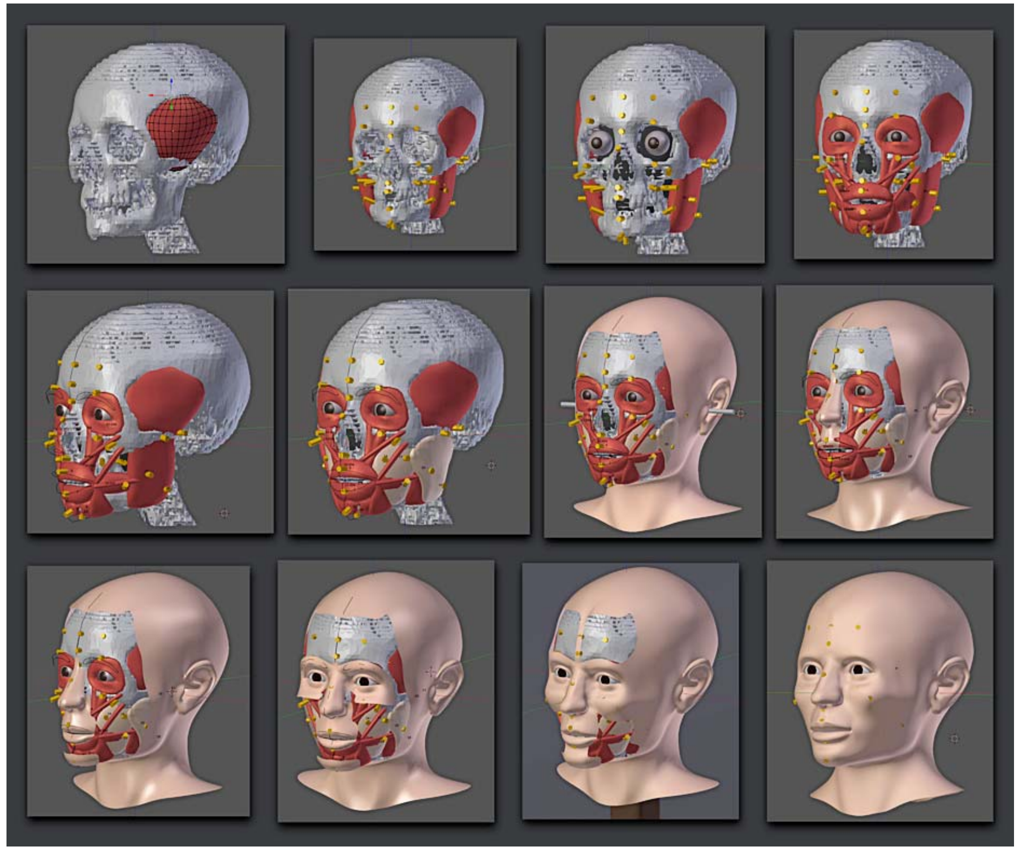
## Forensic face evaluation

| Fea-ture | Details | Similarities and differences | | | | | |
|---|---|---|---|---|---|---|---|
| | | S | NO | D | SI | Explanation for differences | |
| | | | | | | Assumption of same source | Assumption of different source |
| Face | Shape | X | | | | | |
| | Proportions | X | | | | | |
| | Hairline Hairgrowth | | | X | | Age, hairdo | |
| Fore-head | Shape | X | | | | | |
| | Bumps | X | | | | | |
| | Horizontal creases | | X | | | | |
| | Eyebrows | X | | | | | |

52 Didier Meuwly | The use of biometric information in forensic practice | 12.06.2014 | UNIVERSITY OF TWENTE. EEMCS | Netherlands Forensic Institute *Ministry of Security and Justice*

Courtesy of Didier Meuwly NFI

# Face “Reconstruction”



https://www.youtube.com/watch?v=TOgGTgYXTys&list=PL1ptAvVz1FL61_KtmPgqKr-RjTMQiuwJ5&feature=player_detailpage

# Face representation in the HVS

**Table 1. Famous faces (FF) vs newly learned (NL) faces**

| Loc. # | Brain region | BA | vol. (ml) | *x* | *y* | *z* |
|---|---|---|---|---|---|---|
| | FF > NL | | | | | |
| | Frontal Lobe | | | | | |
| 1 | L Superior Frontal | 8 | 2.6 | | | |
| 2 | R Medial Frontal | 9 | 2.4 | | | |
| 3 | R Superior Frontal | 8 | 0.5 | | | |
| 4 | L Medial Frontal | 10 | 0.4 | | | |
| 5 | R Precentral | 6 | 0.4 | | | |
| 6 | L Superior Frontal | 8 | 0.4 | | | |
| 7 | R Inferior Frontal | 47 | 0.3 | | | |
| 8 | R Anterior Cingulate | 32 | 0.3 | | | |
| 9 | R Medial Frontal | 11 | 0.3 | | | |
| 10 | L Medial Frontal | 11 | 0.3 | | | |
| | Temporal Lobe | | | | | |
| 11 | L Middle Temporal | 21 | 2.7 | | | |
| 12 | R Middle Temporal | 21 | 1.9 | | | |
| 13 | L Middle Temporal | 21 | 0.6 | | | |
| 14 | L Middle Temporal | 39 | 0.5 | | | |
| 15 | R Superior Temporal | 22 | 0.5 | | | |
| 16 | R Fusiform | 20/37 | 0.4 | | | |
| 17 | R Middle Temporal | 37 | 0.3 | | | |
| 18 | R Insula | – | 0.3 | | | |
| 19 | R Parahippocampal | 35 | 0.2 | | | |
| 20 | R Parahippocampal | 36 | 0.2 | | | |
| 21 | L Hippocampus | 28 | 0.2 | −19 | −12 | −20 |
| | Parietal/Occipital Lobe | | | | | |
| 22 | L Posterior Cingulate | 23/30 | 1.7 | −4 | −57 | 15 |
| 23 | R Inferior Parietal | 40 | 0.5 | 44 | −30 | 22 |
| 24 | R Posterior Cingulate | 31 | 0.3 | 2 | −57 | 29 |
| 25 | L Extrastriate | 18 | 0.3 | −20 | −89 | 20 |
| | Subcortical | | | | | |
| 26 | R Pons | – | 0.4 | 11 | −43 | −34 |
| 27 | L Pons | – | 0.2 | −10 | −43 | −33 |
| 28 | R Putamen | – | 0.3 | 22 | −7 | −6 |
| | NL > FF | | | | | |
| | Parietal Lobe | | | | | |
| 29 | L Inferior Parietal | 40 | 1.0 | −37 | −64 | 40 |
| 30 | R Superior Parietal | 7 | 0.5 | 23 | −66 | 30 |
| 31 | R Inferior Parietal | 40 | 0.3 | 35 | −67 | 42 |

Region is defined as center of mass. The first column refers to location numbers demarcated in Figures 2 and 3 (italicized numbers indicate locations not shown in figures). Coordinates represent distance in millimeters from anterior commissure: *x* right (+)/left (−); *y* anterior (+)/posterior(−); *z* superior (+)/inferior(−).

**BRAIN Total 1400 ml**
**100 billion neurons**
**71.5 Mneurons/ml**

**Maybe we can sketch the network size devoted to process faces….**

C. L. Leveroni et al. **"Neural Systems Underlying the Recognition of Familiar and Newly Learned Faces"**, The Journal of Neuroscience, January 15, 2000, *20*(2):878–886

# Face representation in the HVS

ices (FF) vs newly learned (NL) faces

| Brain region | BA | vol. (ml) |
|---|---|---|
| FF > NL | | |
| Frontal Lobe | | |
| L Superior Frontal | 8 | 2.6 |
| R Medial Frontal | 9 | 2.4 |
| R Superior Frontal | 8 | 0.5 |
| L Medial Frontal | 10 | 0.4 |
| R Precentral | 6 | 0.4 |
| L Superior Frontal | 8 | 0.4 |
| R Inferior Frontal | 47 | 0.3 |
| R Anterior Cingulate | 32 | 0.3 |
| R Medial Frontal | 11 | 0.3 |
| L Medial Frontal | 11 | 0.3 |
| Temporal Lobe | | |
| L Middle Temporal | 21 | 2.7 |
| R Middle Temporal | 21 | 1.9 |
| L Middle Temporal | 21 | 0.6 |
| L Middle Temporal | 39 | 0.5 |
| R Superior Temporal | 22 | 0.5 |
| R Fusiform | 20/37 | 0.4 |
| R Middle Temporal | 37 | 0.3 |
| R Insula | – | 0.3 |
| R Parahippocampal | 35 | 0.2 |
| R Parahippocampal | 36 | 0.2 |
| L Hippocampus | 28 | 0.2 |
| Parietal/Occipital Lobe | | |
| L Posterior Cingulate | 23/30 | 1.7 |
| R Inferior Parietal | 40 | 0.5 |
| R Posterior Cingulate | 31 | 0.3 |
| L Extrastriate | 18 | 0.3 |
| Subcortical | | |
| R Pons | – | 0.4 |
| L Pons | – | 0.2 |
| R Putamen | – | 0.3 |
| NL > FF | | |
| Parietal Lobe | | |
| L Inferior Parietal | 40 | 1.0 |
| R Superior Parietal | 7 | 0.5 |
| R Inferior Parietal | 40 | 0.3 |

center of mass. The first column refers to location numbers demarcated in Figures 2 and
t distance in millimeters from anterior commissure: $x$ right (+)/left (−); $y$ anterior (+)/

**The BRAIN mass is equal to 1400 ml**
**Composed of some 100 billion neurons**
**71.5 Mneurons/ml**

**Summing up the volumes of all active areas, the total volume is 21,2 ml**

**or … 1.5 Bneurons**

**… with 12K Synapses/neuron!**

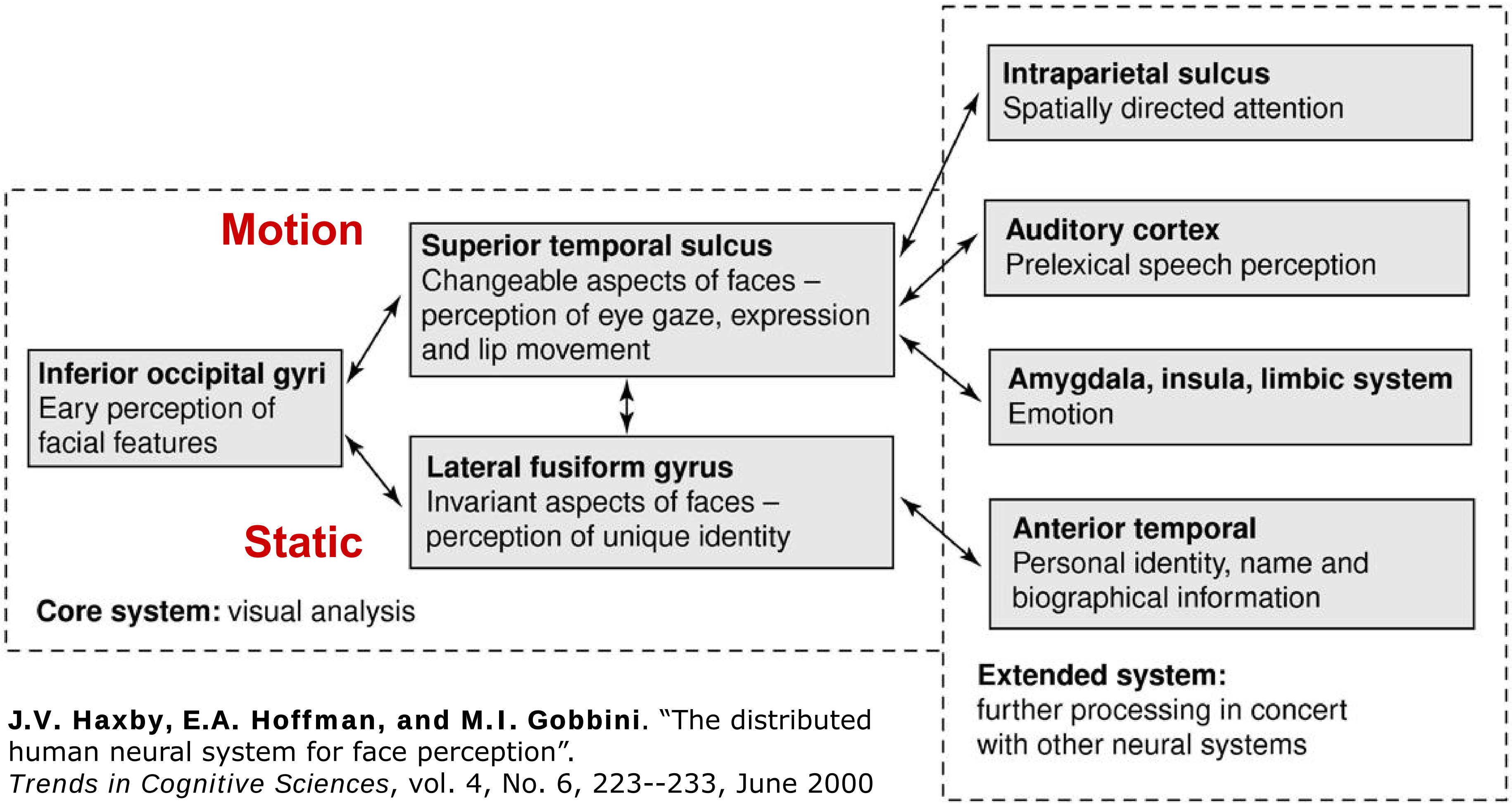
**= 18 trillion synapses!**
**= 2.3 trillion Bytes?**

**If we can learn, say 10,000 faces this corresponds to**

**220 MB/face**

**(or a 7 sec. video stream of 1Kx1K images)**

# Face representation in the HVS




**J.V. Haxby, E.A. Hoffman, and M.I. Gobbini**. "The distributed human neural system for face perception".
*Trends in Cognitive Sciences*, vol. 4, No. 6, 223--233, June 2000

# Face recognition

I. What happened in 20+ years of research in face recognition?

II. What can we learn?

III. What is still to be done?

# Summary and Conclusion

- Face recognition is one of the most appealing biometric modality
  … and one of the most challenging,
  sometimes beyond expectations
- Several successful working systems
  … still several drawbacks and limitations
- Many directions for research to follow
  - Feature extraction and selection
  - Subject-based representation and classification
  - Illumination and registration
  - Complex behaviors
  - Learn from external factors and expectations
  - Exploit contextual information

# THANK YOU
# FOR YOUR ATTENTION

# …AND PATIENCE

# APPENDIX 1

# DATASETS

# Face databases

**The appearance of a face is affected by many factors**

- Identity
- Face pose
- Illumination
- Facial expression

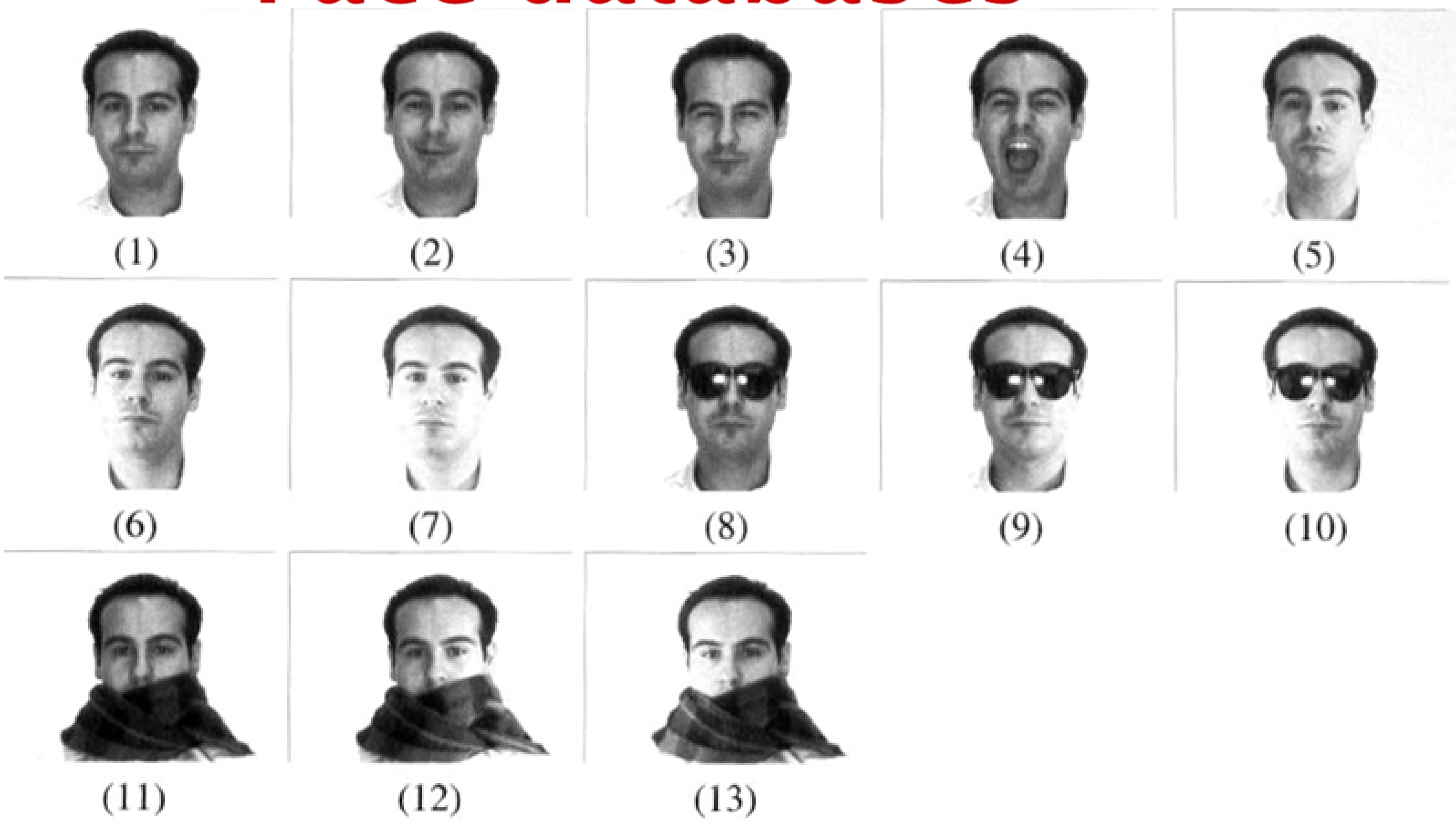
- Age
- Occlusion
- Facial hair

The development of algorithms robust to these variations requires databases of sufficient size that include carefully controlled variations of these factors.

Common databases to comparatively evaluate algorithms.

Collecting a high quality database is a resource-intensive task.

# Face databases




**AR database. The conditions are (1) neutral, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) both lights on, (8) sun glasses, (9) sun glasses/left light (10) sun glasses/right light, (11) scarf, (12) scarf/left light, (13) scarf/right light**

# Face databases

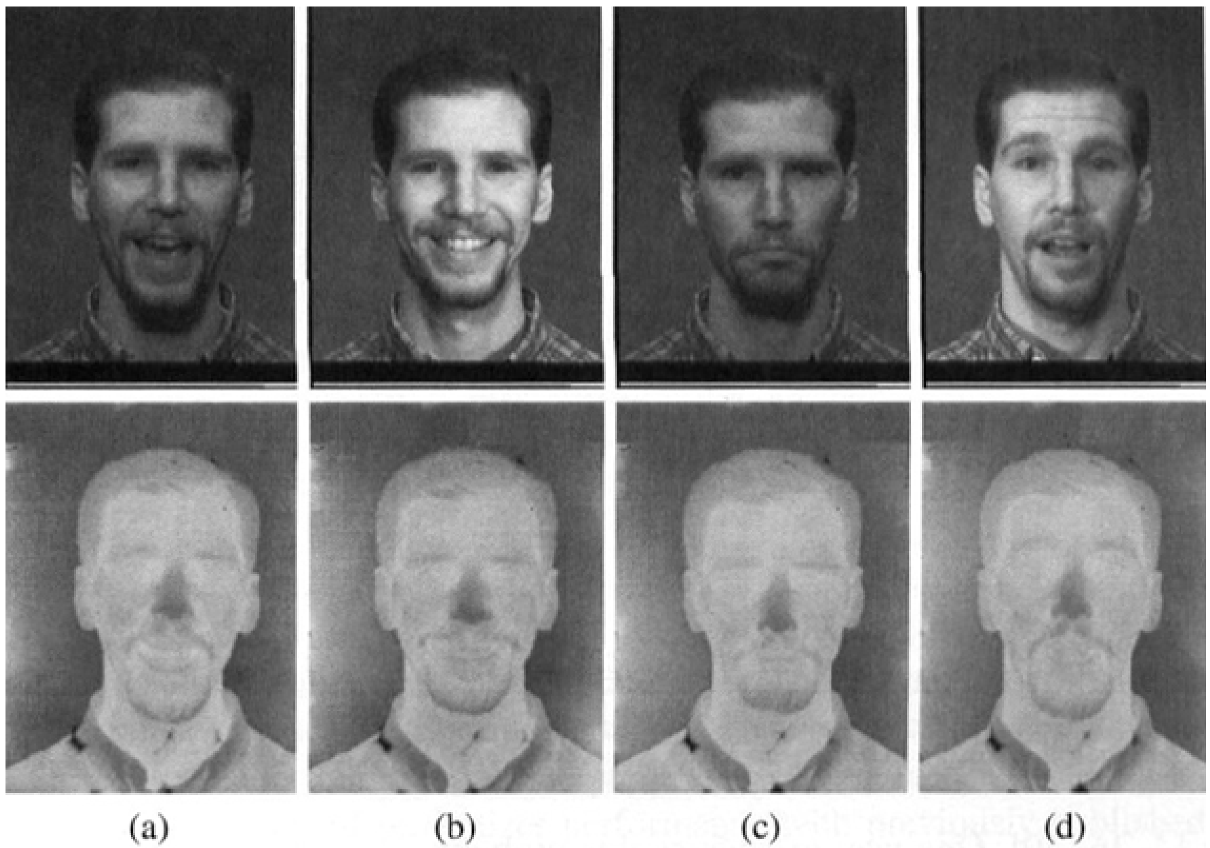
**CAS-PEAL database.** **The images were recorded using separate cameras triggered in close succession. The cameras are about 22.5$^0$ apart. Subjects were asked to look up, to look straight ahead, and to look down. Shown here are seven of the nine poses currently being distributed.**

# Face databases

**Illumination variation in the CAS-PEAL database.** The images were recorded with constant ambient illumination and manually triggered fluorescent lamps.

Also the **CMU PIE database** has been designed to include illumination variations

# Face databases



**Equinox IR database. The upper row contains visible images and the lower row long-wave infrared images. The categories are (a) vowel (frontal illumination), (b) "smile" (right illumination), (c) "frown" (frontal illumination), (d) "surprise" (left illumination).**

# Face databases

fa fb duplicate I fc duplicate II

**Frontal image categories used in the FERET evaluations**. For images in the fb category, a different facial expression was requested. The fc images were recorded with a different camera and under different lighting conditions. The duplicate images were recorded in a later session, with 0 and 1031 days (duplicate I) or 540 to 1031 days (duplicate II) between recordings.

pr $+90^\circ$ hr $+67.5^\circ$ qr $+22.5^\circ$ ql $-22.5^\circ$ hl $-67.5^\circ$ pl $-90^\circ$

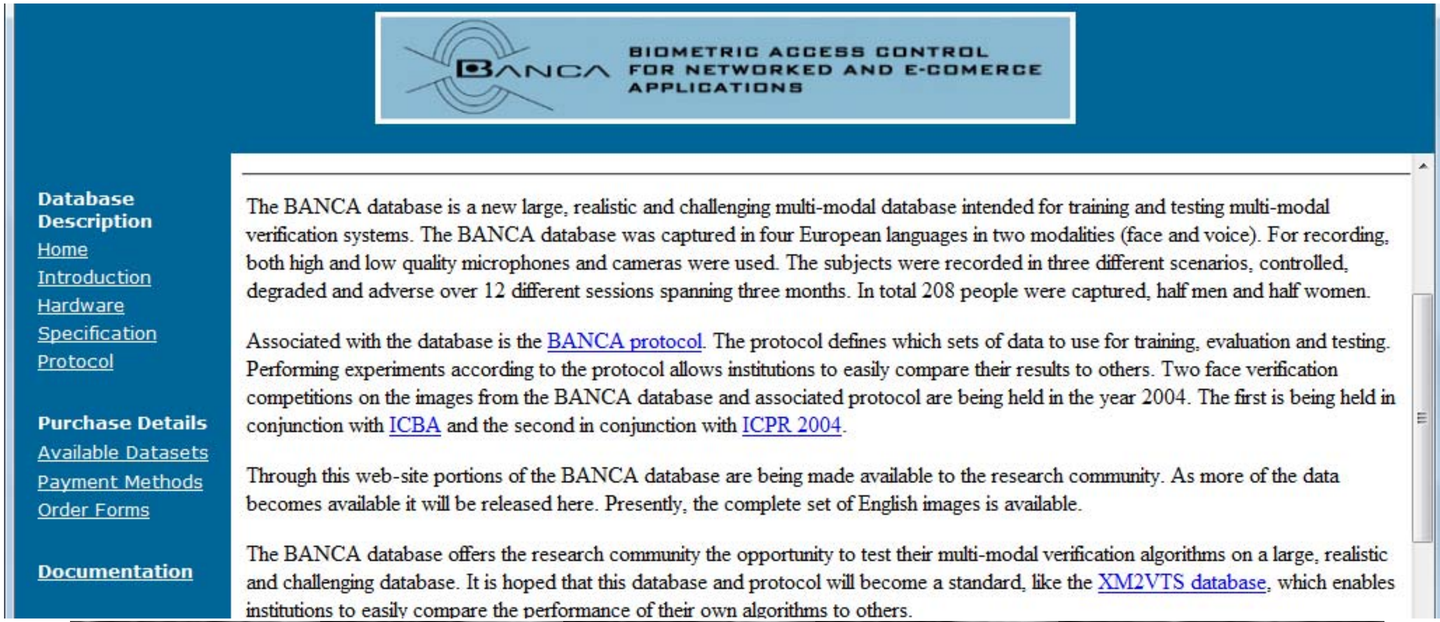
Additional set of pose images from the FERET database: right and left profile (labeled pr and pl), right and left quarter profile (qr, ql), and right and left half profile (hr, hl).

# Face databases

**Notre Dame HumanID database**. Example images of the "unstructured" lighting condition recorded in the hallway outside of the laboratory.

**University of Texas Video Database**. Example images for the different recording conditions of the database. First row: Facial speech. Second row: Laughter. Third row: Disgust.

# Face databases




**The BANCA and XM2VTS video databases distributed by the University of Surrey**

# Face databases

BIOMETRIC ACCESS CONTROL FOR NETWORKED AND E-COMERCE APPLICATIONS

## The BANCA Protocol

An evaluation protocol defines a set of data, how it should be used by a system to perform a set of experiments and how the system performance should be computed.

In verification, two types of protocols exist; closed-set and open-set. In closed-set verification the population of clients is fixed. This means that the system design can be tuned to the clients in the set. Thus both the adopted representation (features) and the verification algorithm applied in the feature space are based on some training data collected for this set of clients. Anyone who is not in the training set is considered an impostor. The XM2VTS protocol is an example of this type of verification problem formulation.

In open-set verification we wish to add new clients to the list without having to redesign the verification system. In particular, we want to use the same feature space and the same design parameters such as thresholds. In such a scenario the feature space and the verification system parameters must be trained using completely independent data from that used for specifying client models. The BANCA protocol is an example of an open-set verification protocol.

# Face detection databases

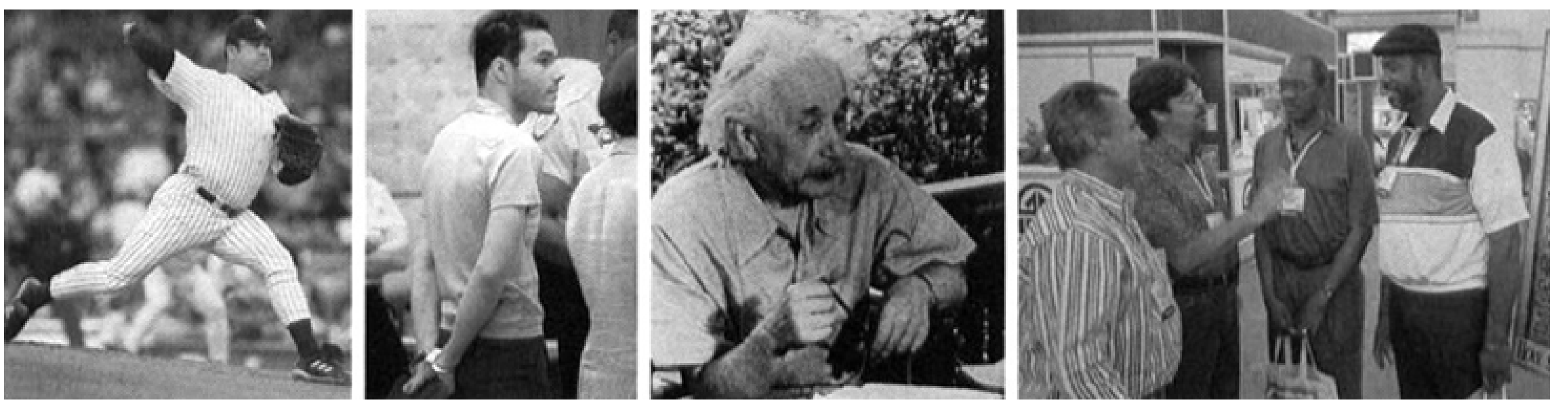
- Face detection algorithms typically have to be trained on face and non-faces images to build up an internal representation of the human face.
- Popular choices are the FERET, MIT, ORL, Harvard, and AR public databases. Nonpublic databases are often also employed.
- These data sets should be representative of real-world data containing faces of various orientations against a complex background.
- In recent years two public data sets emerged as quasi-standard evaluation test sets:
  - The combined MIT/CMU test set for frontal face detection
  - The CMU test set II for frontal and non-frontal face detection

# Face detection databases

**Example images from the Upright Test Set portion of the MIT/CMU test set.**

# Face detection databases

**Example images from the Tilted Test Set portion of the MIT/CMU test set.**

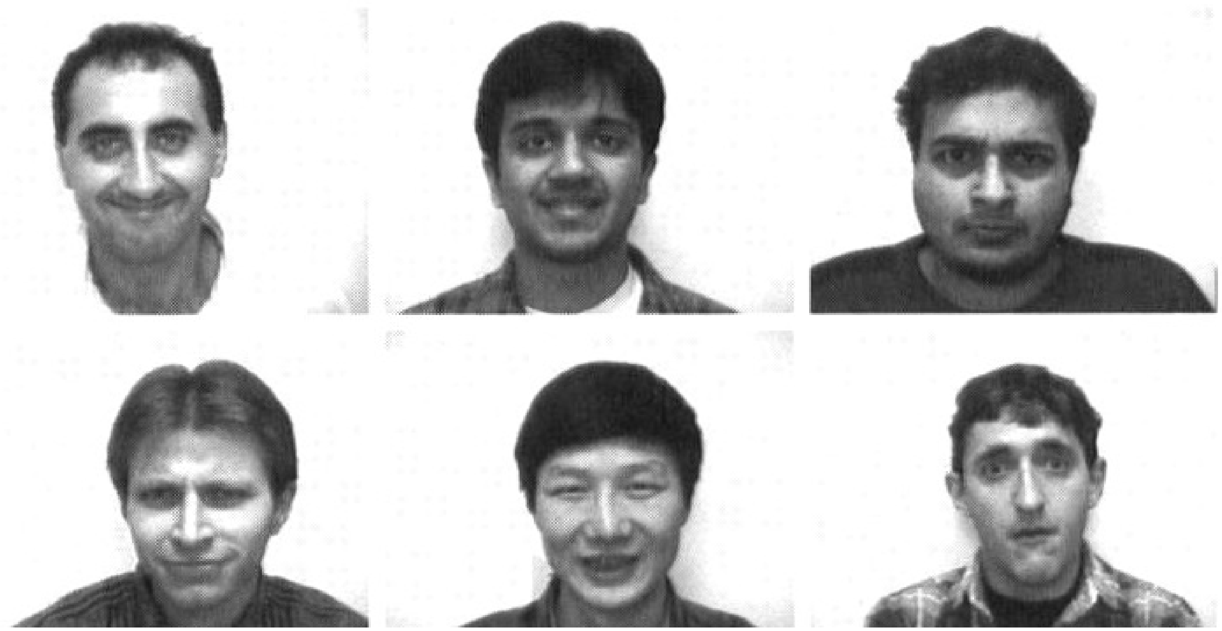


**CMU Test Set II**. Most of the faces in this test set are in profile view.

# Face actions databases

**Cohn-Kanade AU-Coded Facial Expression database.** Examples of emotion-specified expressions from image sequences.

**University of Maryland database.** The images show peak frames taken from an image sequence in which the subjects display a set of facial expressions of their choice.



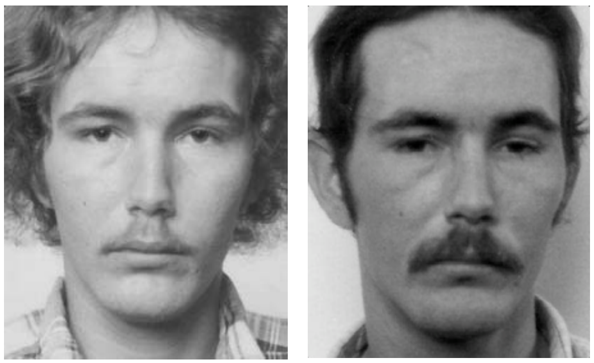
| References | Elicitation method | Size | A/V | Emotion description | Labeling | Accessibility |
|---|---|---|---|---|---|---|
| Cohn-Kanade (CK) '00 [71] | Posed | 210 adults, 3 races; Available: 480 videos | V | Category: 6 basic emotions, and AUs | FACS | Y |
| Sebe et al. (SD) '04 [123] | Natural: Subjects watched emotion-inducing videos | 28 adults | V | Category: Neutral, happy, surprise, disgust | Self-report | N |
| MMI '05[3] [106], [98] | Posed: static images, videos recorded simultaneously in frontal and profile view; Natural: Children interacted with a comedian. Adults watched emotion-inducing videos | Posed: 61 adults Natural: 11 children and 18 adults. Overall: 3 races Available: 1250 videos, 600 static images | V | Category: 6 basic emotions, single AU and multiple AUs activation | FACS, Observers' judgment | Y |
| UT Dallas '06 [95] | Natural: Subjects watched emotion-inducing videos | 229 adults | V | Category: 6 basic emotions, puzzle, laughter, boredom, disbelief | Observers' judgment | Y |
| BU-3DFE (BU)'06 [148] | Posed: 3D range data by using 3DMD digitizer. | 100 adults Mixed races | V | Category: 6 basic emotions. Four levels of intensity | N/A | Y |
| FABO face and body gesture [63] | Posed: two cameras to record facial expressions and body gestures respectively | 23 adults Mixed races Available: 210 videos | V | Category: 6 basic emotions, neutral, uncertainty, anxiety, boredom | N/A | Y |
| Banse-Scherer '96 [8] | Posed | 6 actors & 6 actresses Available: 1344 audio samples | A | Category: hot/cold anger, panic fear, anxiety, despair, sadness, elation, happiness, interest, boredom, shame, pride, disgust, contempt. | Listeners' judgment | Y |
| Danish Emotional Speech Database '96[5] | Posed | 2 actors & 2 actresses; 2 words, 9 sentences, 2 passages; 10 min of audio data. | A | Category: neutral, surprise, happiness, sadness, anger | Listeners' judgment | Y |
| ISL meeting corpus '02 [15] | Natural: meeting corpus | 18 meetings; Available: data of 5 participants per meeting averagely | A | Category: Positive, neutral, negative [3], [90] | Listeners' judgment | Y |
| CSC corpus [65] | Natural: subject was motivated to tell the truth and deceive the interviewers in different tasks | 32 adults, 15.2 h, 3882 speaking turns, 9687 SUs | A | Deceptive, non-deceptive speech | Self-report | N |
| Automatic call center (ACC)'05 [83] | Natural: Human-computer dialogue at a commercial call system | 1187 calls 7200 utterances | A | Category: Negative, non-negative | listeners' judgment | N |
| Bank and Stock Service 04 [34] | Natural: human-human dialogue at call center | 350 dialogues, 10000 speaking turns | A | Category: fear, anger, stress | Listeners' judgment | N |
| AIBO database '04 [13] | Natural: children and robot interaction | 110 dialogues, 29200 words | A | Category: joyful, emphatic, surprised, ironic, helpless, touchy, angry, bored, motherese, reprimanding, rest | Listeners' judgment | N |
| Chen-Huang (CH) '00 [21] | Posed | 100 adults, 9900 visual and AV expressions | AV | Category: 6 basic emotions, and 4 cognitive states (interest, puzzle, bore, frustration) | N/A | N |
| Adult Attachment Interview (AAI)'04[111] | Natural: subjects were interviewed to describe the childhood experience | 60 adults Each interview last 30-60min | AV | Category: 6 basic emotions, embarrassment, contempt, shame, general positive and negative. | FACS | N |
| RU-FACS (RU) '05 [10] | Natural: subjects were tried to convince the interviewers they were telling the truth | 100 adults | AV | Category: 33 AUs | FACS | N |
| SAL '05[7] | Induced: subjects interacted with artificial listener with different personalities | 24 adults 10h | AV | Dimensional labeling/categorical labeling | FEEL-TRACE | Y |
| Belfast database (BE) '03 [38] | Natural: clips taken from television and realistic interviews with research team | 125 subjects. 209 sequences from TV, 30 from interview | AV | Dimensional labeling/categorical labeling | FEEL-TRACE | Y |

Vision Lab

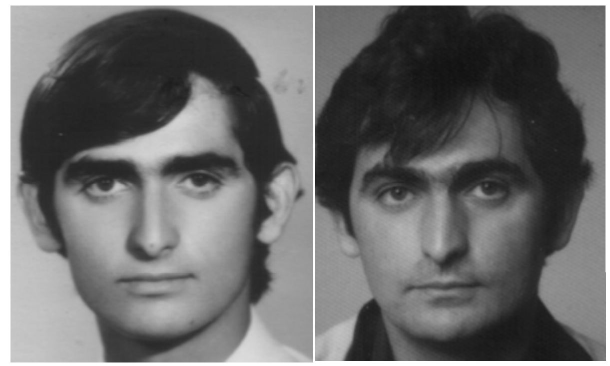
# Face actions databases

| References | Elicitation method | Size | A/V | Emotion description | Labeling | Accessibility |
|---|---|---|---|---|---|---|
| Cohn-Kanade (CK) '00 [71] | Posed | 210 adults, 3 races; Available: 480 videos | V | Category: 6 basic emotions, and AUs | FACS | Y |
| Sebe et al. (SD) '04 [123] | Natural: Subjects watched emotion-inducing videos | 28 adults | V | Category: Neutral, happy, surprise, disgust | Self-report | N |
| MMI '05[3] [106], [98] | Posed: static images, videos recorded simultaneously in frontal and profile view; Natural: Children interacted with a comedian. Adults watched emotion-inducing videos | Posed: 61 adults Natural: 11 children and 18 adults. Overall: 3 races Available: 1250 videos, 600 static images | V | Category: 6 basic emotions, single AU and multiple AUs activation | FACS, Observers' judgment | Y |
| UT Dallas '06 [95] | Natural: Subjects watched emotion-inducing videos | 229 adults | V | Category: 6 basic emotions, puzzle, laughter, boredom, disbelief | Observers' judgment | Y |
| BU-3DFE (BU)'06 [148] | Posed: 3D range data by using 3DMD digitizer. | 100 adults Mixed races | V | Category: 6 basic emotions. Four levels of intensity | N/A | Y |
| FABO face and body gesture [63] | Posed: two cameras to record facial expressions and body gestures respectively | 23 adults Mixed races Available: 210 videos | V | Category: 6 basic emotions, neutral, uncertainty, anxiety, boredom | N/A | Y |
| Chen-Huang (CH) '00 [21] | Posed | 100 adults, 9900 visual and AV expressions | AV | Category: 6 basic emotions, and 4 cognitive states (interest, puzzle, bore, frustration) | N/A | N |
| Adult Attachment Interview (AAI)'04[111] | Natural: subjects were interviewed to describe the childhood experience | 60 adults Each interview last 30-60min | AV | Category: 6 basic emotions, embarrassment, contempt, shame, general positive and negative. | FACS | N |
| RU-FACS (RU) '05 [10] | Natural: subjects were tried to convince the interviewers they were telling the truth | 100 adults | AV | Category: 33 AUs | FACS | N |
| SAL '05[7] | Induced: subjects interacted with artificial listener with different personalities | 24 adults 10h | AV | Dimensional labeling/categorical labeling | FEEL-TRACE | Y |
| Belfast database (BE) '03 [38] | Natural: clips taken from television and realistic interviews with research team | 125 subjects. 209 sequences from TV, 30 from interview | AV | Dimensional labeling/categorical labeling | FEEL-TRACE | Y |

# Face aging databases

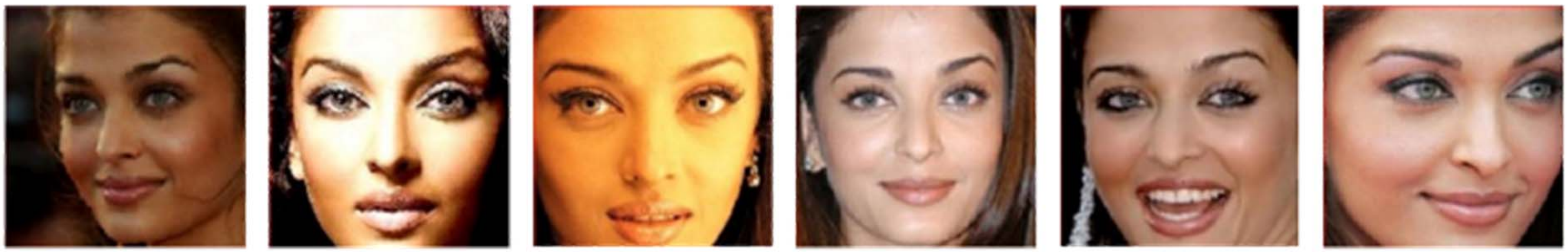


**FG-NET dataset:**
**images of about 100 individuals with ages varying from 5 to 69 years of age**



**MORPH dataset:**
**images of 13,000 individuals collected over four years**

# Face aging databases



**IIIT-Delhi face aging database:2,618 images from 49 female and 53 male Indian celebrities**

# 3D Face Databases

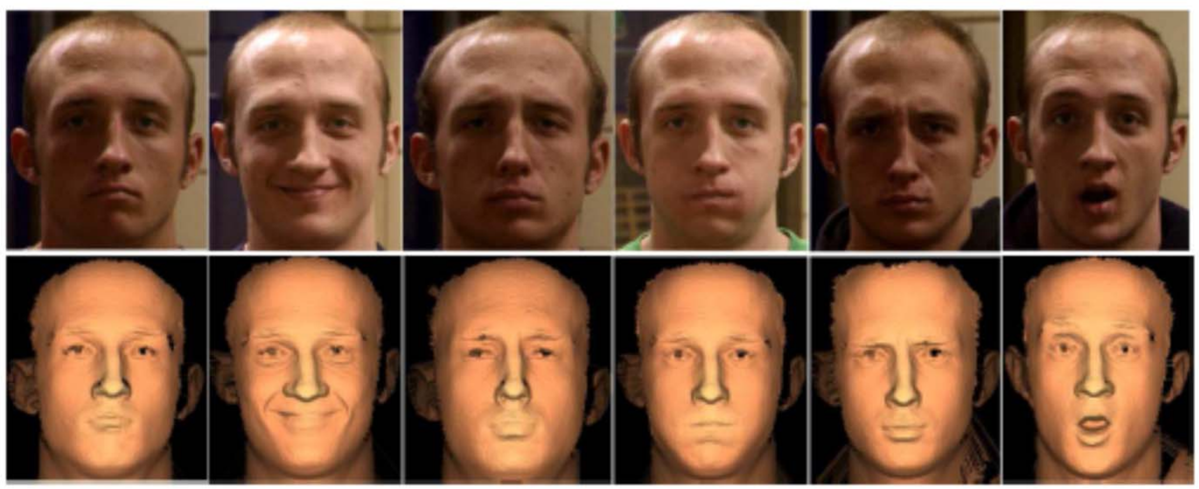
- Features:
  - Full 3D / 2.5D
  - Single view / Multi-view
  - Illumination changes
  - Expressions changes
  - Only shape or shape + texture
  - Pre-processing
  - Quality of data (sensors)

# 3D Face Databases

- UND database (University of Notre Dame)
  - 953 facial scans (277 subjects)
  - frontal scans (neutral expression)
  - 2.5D shape + texture

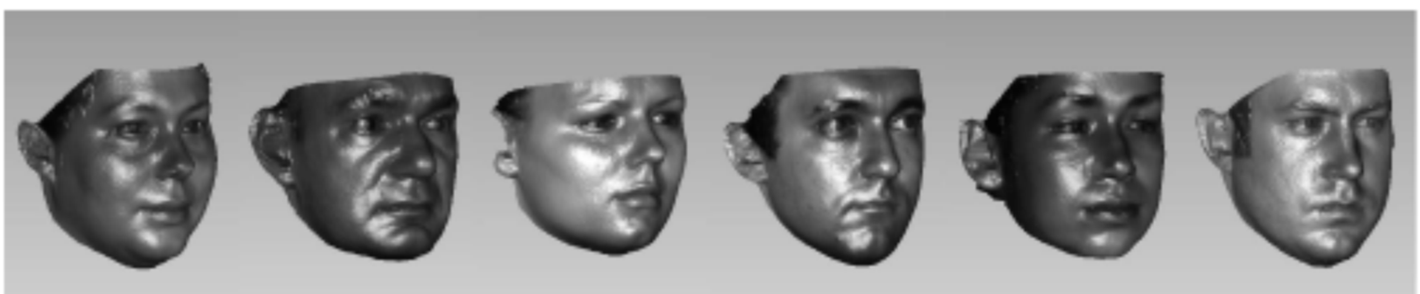
# 3D Face Databases

- FRGC database (NIST)
  - 4007 scans (465 subjects)
  - nearly frontal
  - different expressions
  - 2.5D shape + texture



# 3D Face Databases

- USF database (University of South Florida)
  - 100 scans
  - full-view models
  - neutral expression

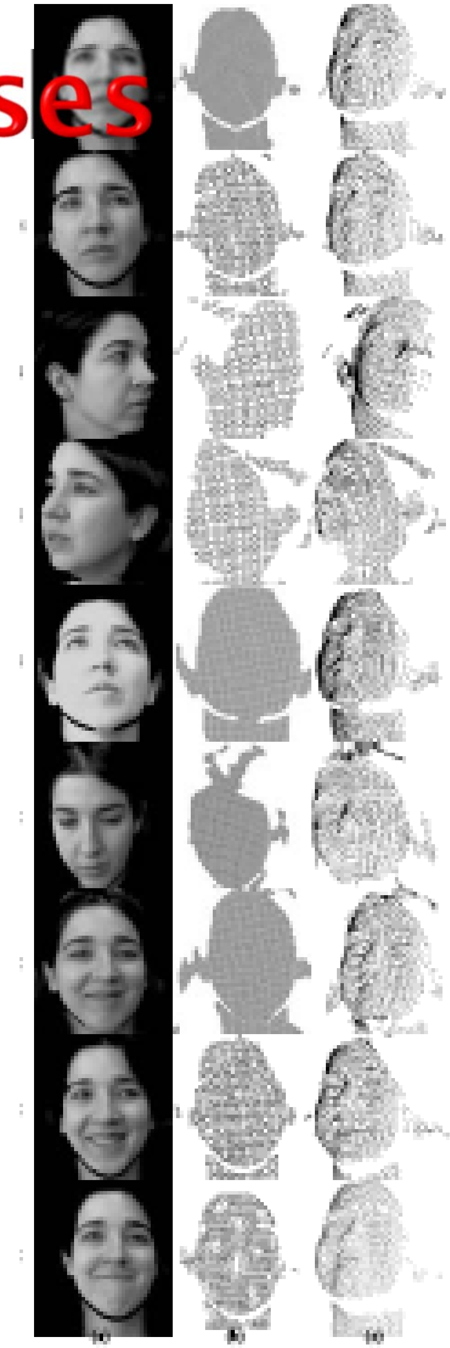


# 3D Face Databases

- 3D_RMA database (Royal Military Academy of Belgium)
  - 120 subjects (2 sessions, 3 scans each)
  - different (but limited) orientations
  - 3D points

# 3D Face Databases

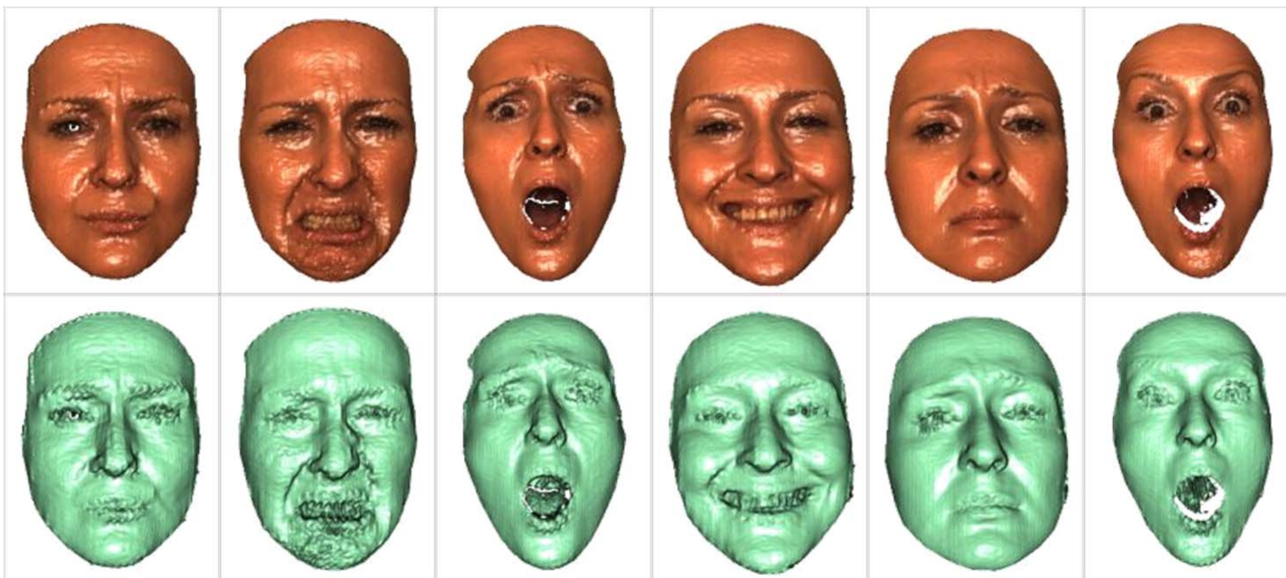
- GavabDB database (University Rey Juan Carlos – Madrid)
  - different orientations and expressions
  - 61 subjects (9 scans)
  - Shape only



# 3D Face Databases

**The Bosphorus database contains scans of 105 individuals: 61 male; 44 female**

**About 50 scans/subject. Each scan either presents a diffent facial expression (anger, happiness, disgust), or a head rotation along different axes.**



A.Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun,
Bosphorus Database for 3D Face Analysis, The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008) Roskilde University, Denmark, May 2008.

# 3D Mask Attack Dataset

**76500 frames of 17 persons, recorded using Kinect for both real-access and spoofing attacks. https://www.idiap.ch/dataset/3dmad**

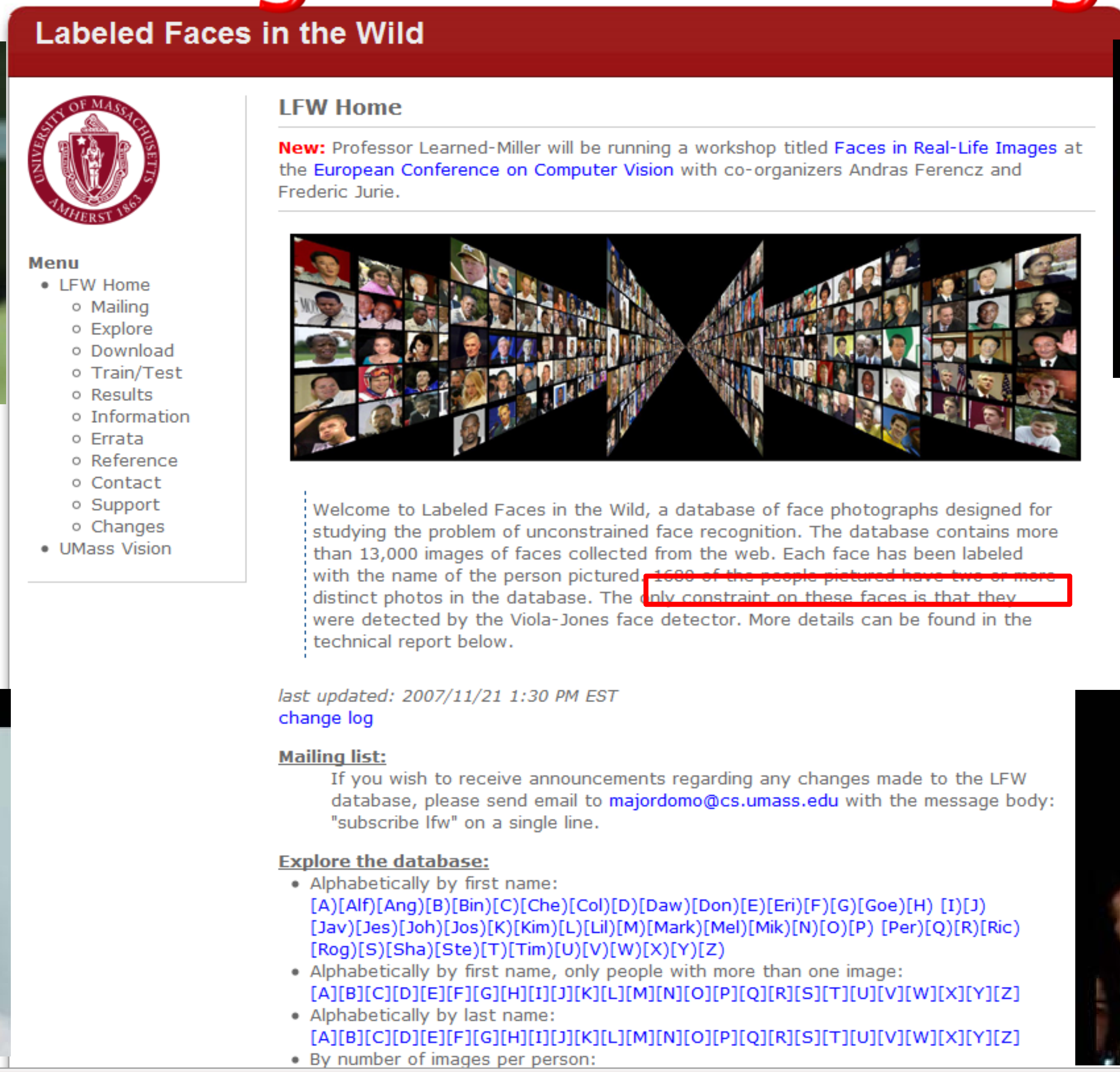
# Face recognition challenges



**After**

- **FERET**
- **FRVT**
- **FRGC**
- **…**

# Face recognition challenges



Labeled Faces in the Wild

## LFW Home

**New:** Professor Learned-Miller will be running a workshop titled Faces in Real-Life Images at the European Conference on Computer Vision with co-organizers Andras Ferencz and Frederic Jurie.

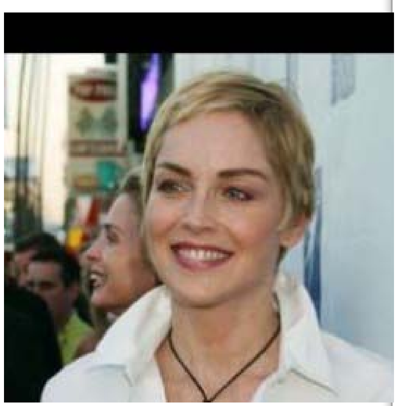
**Menu**
- LFW Home
  - Mailing
  - Explore
  - Download
  - Train/Test
  - Results
  - Information
  - Errata
  - Reference
  - Contact
  - Support
  - Changes
- UMass Vision

Welcome to Labeled Faces in the Wild, a database of face photographs designed for studying the problem of unconstrained face recognition. The database contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the database. The only constraint on these faces is that they were detected by the Viola-Jones face detector. More details can be found in the technical report below.

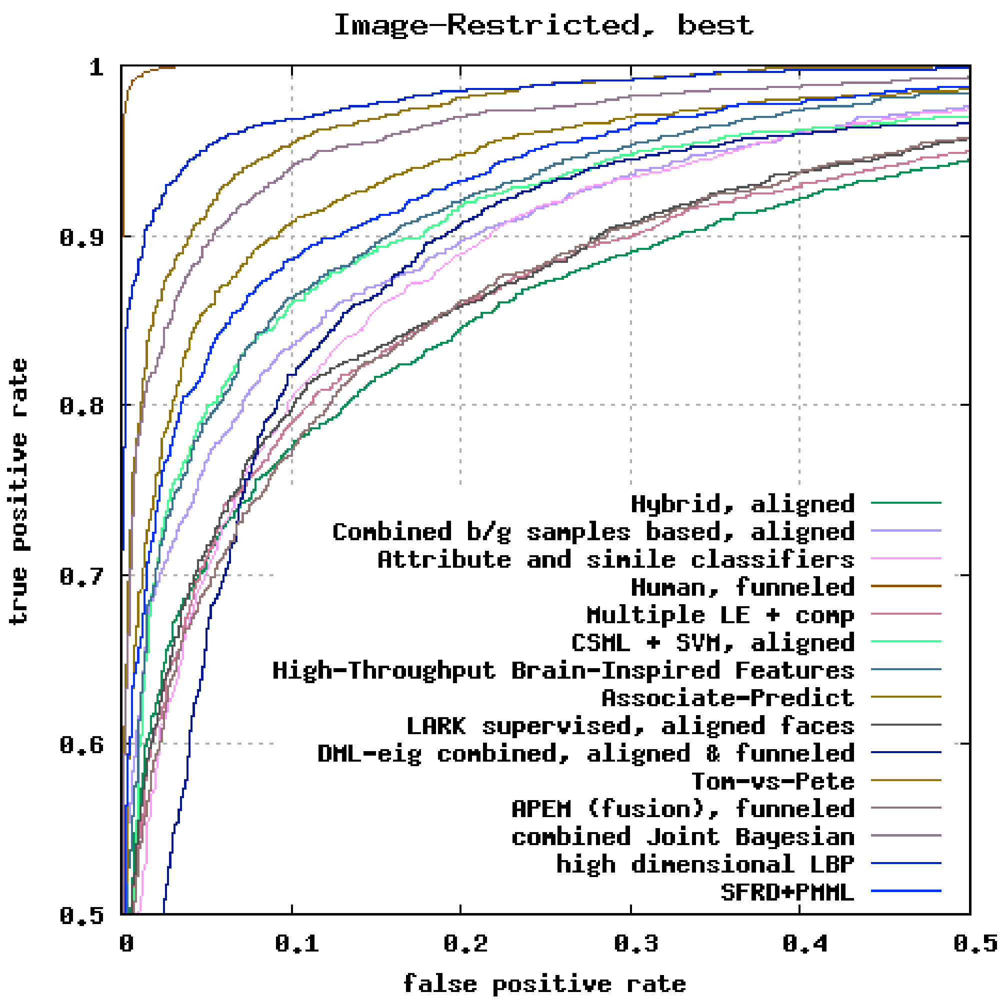
*last updated: 2007/11/21 1:30 PM EST*
change log

**Mailing list:**
If you wish to receive announcements regarding any changes made to the LFW database, please send email to majordomo@cs.umass.edu with the message body: "subscribe lfw" on a single line.

**Explore the database:**
- Alphabetically by first name:
  [A][Alf][Ang][B][Bin][C][Che][Col][D][Daw][Don][E][Eri][F][G][Goe][H] [I][J] [Jav][Jes][Joh][Jos][K][Kim][L][Lil][M][Mark][Mel][Mik][N][O][P] [Per][Q][R][Ric] [Rog][S][Sha][Ste][T][Tim][U][V][W][X][Y][Z]
- Alphabetically by first name, only people with more than one image:
  [A][B][C][D][E][F][G][H][I][J][K][L][M][N][O][P][Q][R][S][T][U][V][W][X][Y][Z]
- Alphabetically by last name:
  [A][B][C][D][E][F][G][H][I][J][K][L][M][N][O][P][Q][R][S][T][U][V][W][X][Y][Z]
- By number of images per person:



# Face recognition challenges




http://vis-www.cs.umass.edu/lfw/results.html

# Biometrics challenges

Vision Lab

2-5 June 2009 University of Sassari, Italy

ICB 2009

The 3rd IAPR/IEEE International Conference on Biometrics

- Home
- Call for Papers
- Committee
- Competitions
- Keynote Speakers
- Paper Submission
- Demo Submission
- Awards
- Call for Tutorials
- Conference Venue
- Sponsors
- Travel Information
- Poster

## ICB2009 Competitions

IAPR

IEEE Biometrics Council

### Competitions Chairs

| | |
|---|---|
| Bernadette Dorizzi | Biosecure Foundation, France |
| Jonathon Phillips | NIST, USA |

### Face recognition from stills and video

This competition is performed under the supervision of Norman Poh from the University of Surrey

### Fingerprint

This competition is performed under the supervision of Raffaele Cappelli from the University of Bologna

### Multimodal Biometric Feature Selection Challenge

This competition is performed under the supervision of Krzysztof Kryszczuk from the Ecole Politechnique Federale de Lausanne

### Multiple Biometrics Grand Challenge

The Multiple Biometrics Grand Challenge is organized and supported by the National Institute of Standards and Technology (NIST). The MBGC is sponsored by multiple U.S. Government Agencies. Dr Jonathon Phillips is the responsible for the NIST MBGC evaluation. Within the framework of ICB program, submissions are encouraged to the MBGC evaluation. The results of the MBGC, together with the other competitions, will be presented at a special conference session.

### Signature verification

This competition is performed under the supervision of Sonia Garcia-Salicetti from the institute TELECOM SudParis

# Biometrics challenges

Vision Lab

## ICB-2013

**The 6th IAPR International Conference on Biometrics**

**June 4 - 7, 2013 Madrid, Spain**

- Home
- Call for Papers
- Paper Submission
- Camera-Ready Instructions
- Presentation Instructions
- Program / Schedule
- Keynote Speakers
- Organizing Committee
- Program Committee
- **Competitions**
- Doctoral Consortium
- Tutorials
- Registration
- Visa Information
- Accommodation
- Related Events
- Conference Venue
- Social Program

### Competitions

The availability of common benchmark databases, together with evaluation protocols has been partly responsible for the significant gains made in biometrics in recent years. We believe that such evaluations should be continued. Databases and, more importantly, unbiased evaluation mechanisms should be spread across the scientific community, making it possible for scientists to evaluate their progress.

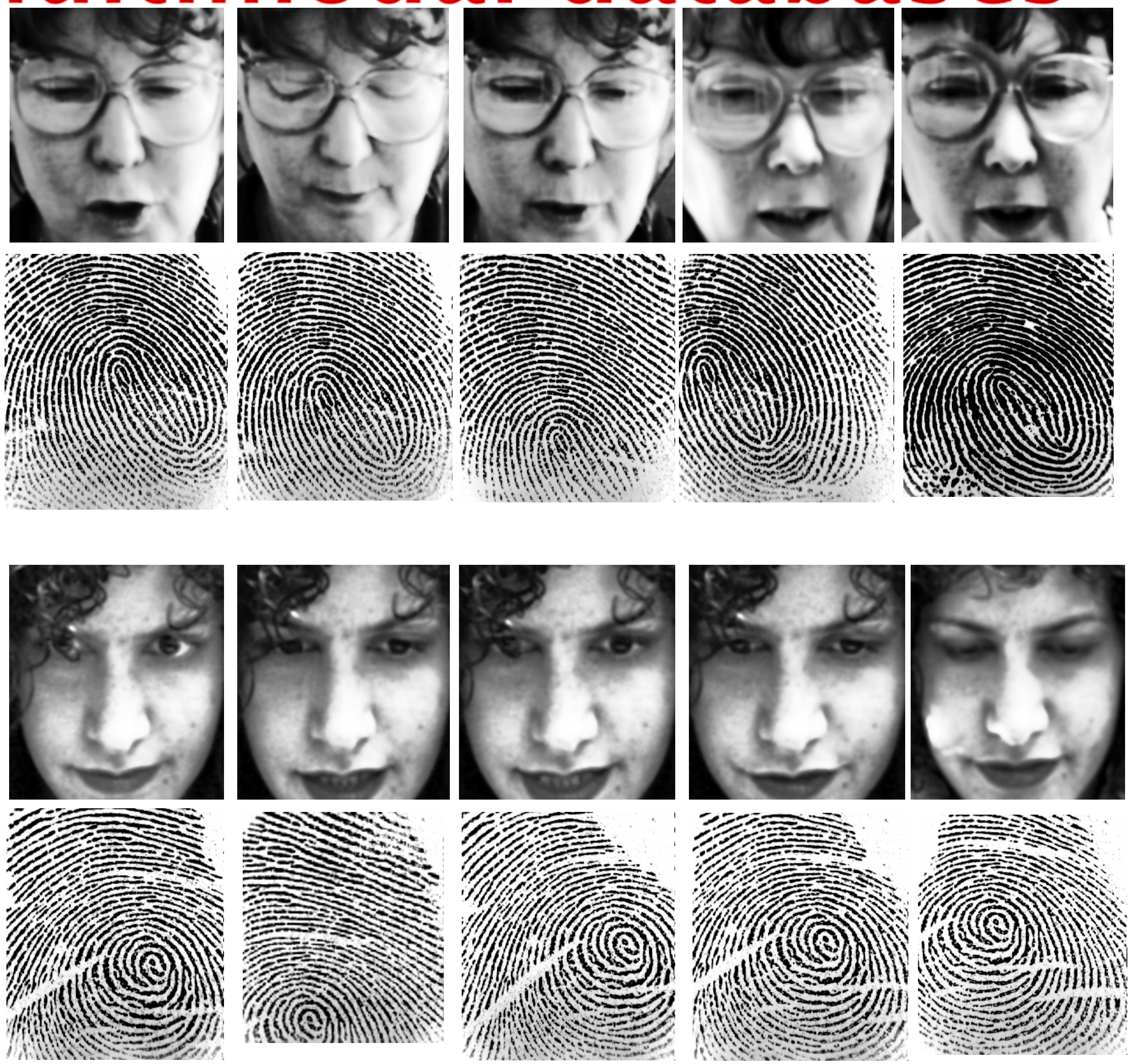
The 6th IAPR International Conference on Biometrics (ICB 2013) is supporting the organization of the following 8 evaluations:

» The 2nd competition on counter measures to 2D facial spoofing attacks
» Competition on face recognition in mobile environment using the MOBIO database
» Competition on speaker recognition in mobile environment using the MOBIO database
» The First ICB Competition on Face Recognition (ICFR2013)
» The First ICB Competition on Iris Recognition (ICIR2013)
» Competition on Secure Template Fingerprint Verification (STFV@ICB-2013)
» Competition on Fingerprint Indexing (FIDX@ICB-2013)
» Competition on Fingerprint Liveness Detection

Competitions will be running from January 7, 2013 to March 22, 2013, but database and instructions will be available in late 2012. Each competition will have the opportunity to submit for review a competition summary paper for possible publication into the official proceedings.

Together with these 8 evaluations, an on-site spoofing challenge intended to evaluate operational vulnerabilities of various biometric systems will be conducted:
» TABULA RASA Spoofing Challenge. [NEW!]

# Multimodal databases



# Quality in Face and Iris Research Ensemble (Q-FIRE)

- **Release 1**
  - **4TB of visible face and NIR face/iris video for 90 subjects**
  - **Multiple distance up to 8.3 meters with controlled quality degradation**
- **Release 2**
  - **An additional 83 subjects (*currently sequestered*)**
- **Available by Request to CITeR** ***http://www.clarkson.edu/citer/research/collections***

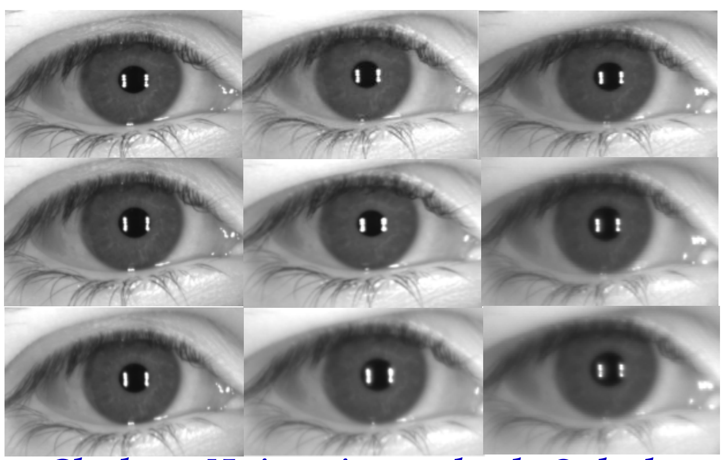
## Out-of Focus Blur

5 ft | 15 ft | 25ft

Sharp No blur

Low blur

High blur



5 ft | 15 ft | 25ft

Sharp No blur

Low blur

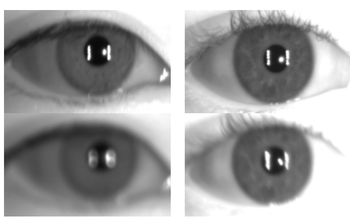
High blur

*S. Schuckers, Clarkson University, sschucke@clarkson.edu*

# Quality in Face and Iris Research Ensemble (Q-FIRE)

- **Release 1**
  - **4TB of visible face and NIR face/iris video for 90 subjects**
  - **Multiple distance up to 8.3 meters with controlled quality degradation**
- **Release 2**
  - **An additional 83 subjects (*currently sequestered*)**
- **Available by Request to CITeR** ***http://www.clarkson.edu/citer/research/collections***

**Motion Blur**



**Illumination**

**Angle**

**Multiple Faces**

*S. Schuckers, Clarkson University, sschucke@clarkson.edu*

# Multimodal databases

# VIDEO DATASETS

# Natural Viewing Environment Video

Courtesy of Jonathon Phillips, NIST

# Point & Shoot Challenge—Video

# Multiple Biometric Grand Challenge MBGC

- Walking footage: Subject walks towards camera.
- Activity/Conversation footage: Non-frontal footage of subject performing an A/C.

U of Texas at Dallas & U of Notre Dame

# Examples PaSC—Stills

# PaSC

- Problem Defined
  - 9,376 stills; 293 subjects
  - 2,802 videos; 265 subjects
  - Video-to-video
  - Still-to-video
  - Still-to-still
- http://face.nist.gov
- Release 11 June 2013

- Design: Colorado State (CSU) and NIST
- Data collection: Notre Dame, NIST, and CSU

# FOCUS UT-Dallas Dataset

FOCS videos: 510 walking (frontal face) and 506 activity (profile face) videos

J. O'Toole, P. J. Phillips, S. Weimer, D. A. Roark, J. Ayyad, R. Barwick, and J.Dunlop, “Recognizing people fromdynamicandstatic faces and bodies:Dissectingidentity with a fusion approach,” *Vision Research, vol.51, no.1, pp.74-83, 2011.*

National Institute of Standards and Technology, "Face and Ocular Challenge Series (FOCS)”.

# Face in Video Evaluation

Vision Lab



F I V E

## Face In Video Evaluation (FIVE)

**2014-07-16 Program Announcement**

**Scope:** The Face in Video Evaluation (FIVE) is being conducted to assess the capability of face recognition algorithms to correctly identify or ignore persons appearing in video sequences – i.e. the open-set identification problem. Both comparative and absolute accuracy measures are of interest, given the goals to determine which algorithms are most effective and whether any are viable for the following primary operational use-cases: 1. High volume screening of persons in the crowded spaces (e.g. an airport); 2. Low volume forensic examination of footage from a crime scene (e.g. a convenience store); 3. Persons in business meetings (e.g. for video-conferencing); and 4. Persons appearing in television footage. These applications differ in their tolerance of false positives, whether a human examiner will review outputs, the prior probabilities of mate vs. non-mate presence, and the cost of recognition errors.

**Out of scope:** Gait, iris and voice recognition; Recognition across multiple views (e.g. via stereoscopic techniques); Tracking across sequential cameras (re-identification); anomaly detection; detection of evasion.

**Relationship to FRVT:** The Face Recognition Vendor Tests of 2000, 2002, 2006, 2010, and 2013 gave quantitative statements of accuracy and speed of mostly still-image face recognition algorithms. The last test included a video track (FRVT class V) – results from that work are being provided to participants. Our new FIVE program supersedes the FRVT work but proceeds in an almost identical manner.

**Test progression:** Software submitted to NIST will be evaluated on sequestered sets to quantify accuracy and speed. Algorithms must be implemented behind the formal C++ API to be published by NIST. This will be very similar to the API used in the prior FRVT evaluation. The test will be conducted over at least three iterative cooperative test-report-test phases engaging algorithm developers.This process will culminate in the publication of reports on this website and in the open literature.

Select Language

Powered by Google Translate

SHARE

**Other Links**

- FRVT Homepage
- face.nist.gov
- Image Group Publications Archive Homepage
- Fingerprint Homepage
- Image Group Homepage
- Biometrics Evaluations Homepage
- ITL Biometrics Overview
- iris.nist.gov

*http://www.nist.gov/itl/iad/ig/five.cfm*

# ...and Beyond

- Female makeup Datasets
- Plastic Surgery Face Database
- YouTube Faces Database
- ChokePoint
- SCface - Surveillance Cameras Face Database
- Long Distance Heterogeneous Face Database (LDHF-DB)
- VADANA: Vims Appearance Dataset for facial ANAlysis
- MOBIO - Mobile Biometry Face and Speech Database

# APPENDIX 2

# KINSHIP

Lorusso L., Brelstaff G., Brodo L., Lagorio A., Grosso E., “Visual judgments of kinship: an alternative perspective”, Perception 40(11):1282-9, 2011

# Faces and Kinship

# Faces and Kinship

## 1 Kings 3:16-27


**Raphael's oil painting *The Judgement of Solomon* – 1518**

# Face-based Kinship analysis: the literature

Vision Lab

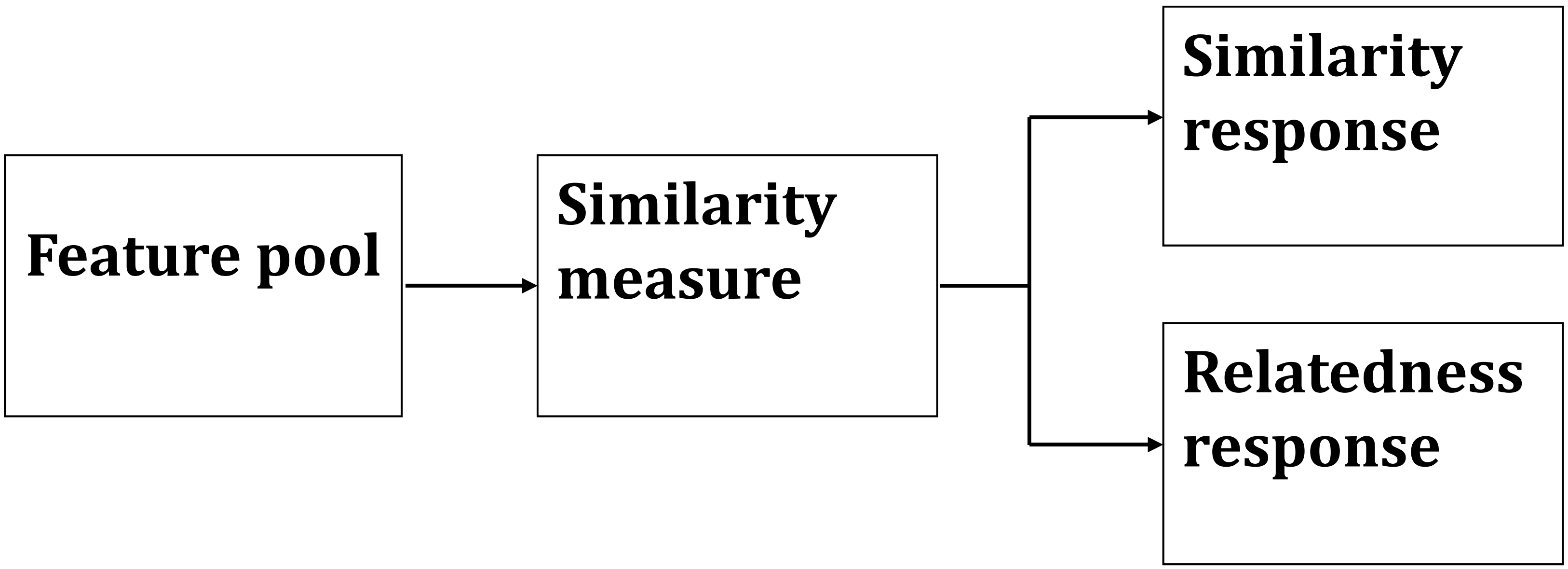
Studies of kin recognition investigated the relationship between **KINSHIP** and **SIMILARITY**:

- Maloney & Dal Martello , 2006, "Kin recognition and the perceived facial similarity of children" *Journal of Vision* 6, pp. 1047-1056.
- Dal Martello & Maloney, 2006, "Where are kin recognition signals in the human face?", *Journal of Vision* 6, pp. 1356-1366.
- DeBruine et al., 2009, "Kin recognition signals in adult faces" *Vision Research* 49, pp. 38-43.
- Alvergne et al., 2010, "Are parents' perceptions of offspring facial resemblance consistent with actual resemblance? Effects on parental investment" *Evolution and Human Behaviour* 31, pp.7-15.

# Face-based Kinship analysis: the literature

Maloney and Dal Martello (2006) have introduced a model for kin recognition - ***Thresholded Similarity Observer (TSO) model*** - in which kin recognition is reduced to a similarity measure of similarity cues.

# The TSO Model




The flow of visual information (visual pathway) in similarity and kinship judgments
(Maloney and Dal Martello 2006).

# The TSO Assumptions

Vision Lab

1. Genetically-related people manifest in their faces "Kin signals".
2. "Kin signals" are represented by similarity features.
3. Kin recognition is a signal detection task.
4. Perception of similarity and kin recognition are based on a common measure of similarity of features.
5. Kinship and similarity judgments are based on a common flow of visual information from the stimulus to the observer.
6. The model of perceived similarity between faces is a "generalized" (*vs* an "individualized") model.

# The TSO Model Expanded

# The TSO Assumptions

- From such assumptions it follows that judgments of similarity and kinship do not require any kind of cognitive modulation, but a visual flow of similarity information is sufficient for the observer to make the judgments.

- In other words, **bottom-up mechanisms are sufficient** to make kinship judgments.

# Not considered in TSO

Vision Lab

- **Top-down mechanisms**: task-driven and context-dependent strategies. Different tasks and visual contexts may induce the observer to follow different observational strategies.

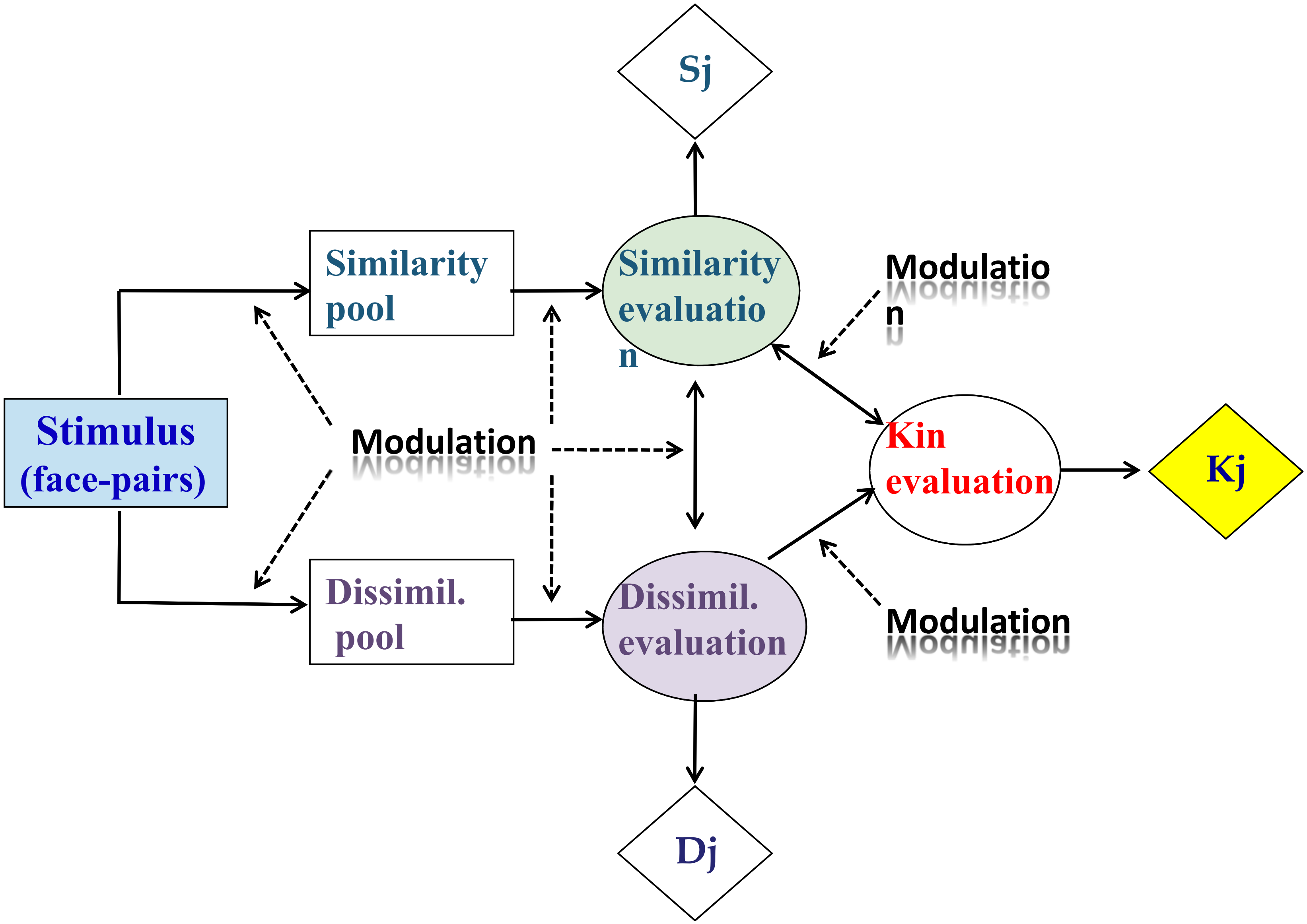
- **Relations between concepts,** may enhance or undermine the belief that two persons are ***similar*** or ***genetically related*** (i.e. *priming* effects of a judgment over another).

# Not considered in TSO

- Our past study on this topic (Lorusso et al., *Perception* 2011) investigates the possibility that a **judgment of kinship may not be just a judgment of similarity** and that different strategies may be involved in those judgments modulated by the ***task***, ***relations*** between concepts, and different ***stimulus contexts***.

Lorusso L., Brelstaff G., Brodo L., Lagorio A., Grosso E., "Visual judgments of kinship: an alternative perspective", *Perception* 40(11):1282-9, 2011

# The TSO Model Expanded



# APPENDIX 2

# INDUSTRIAL SYSTEMS

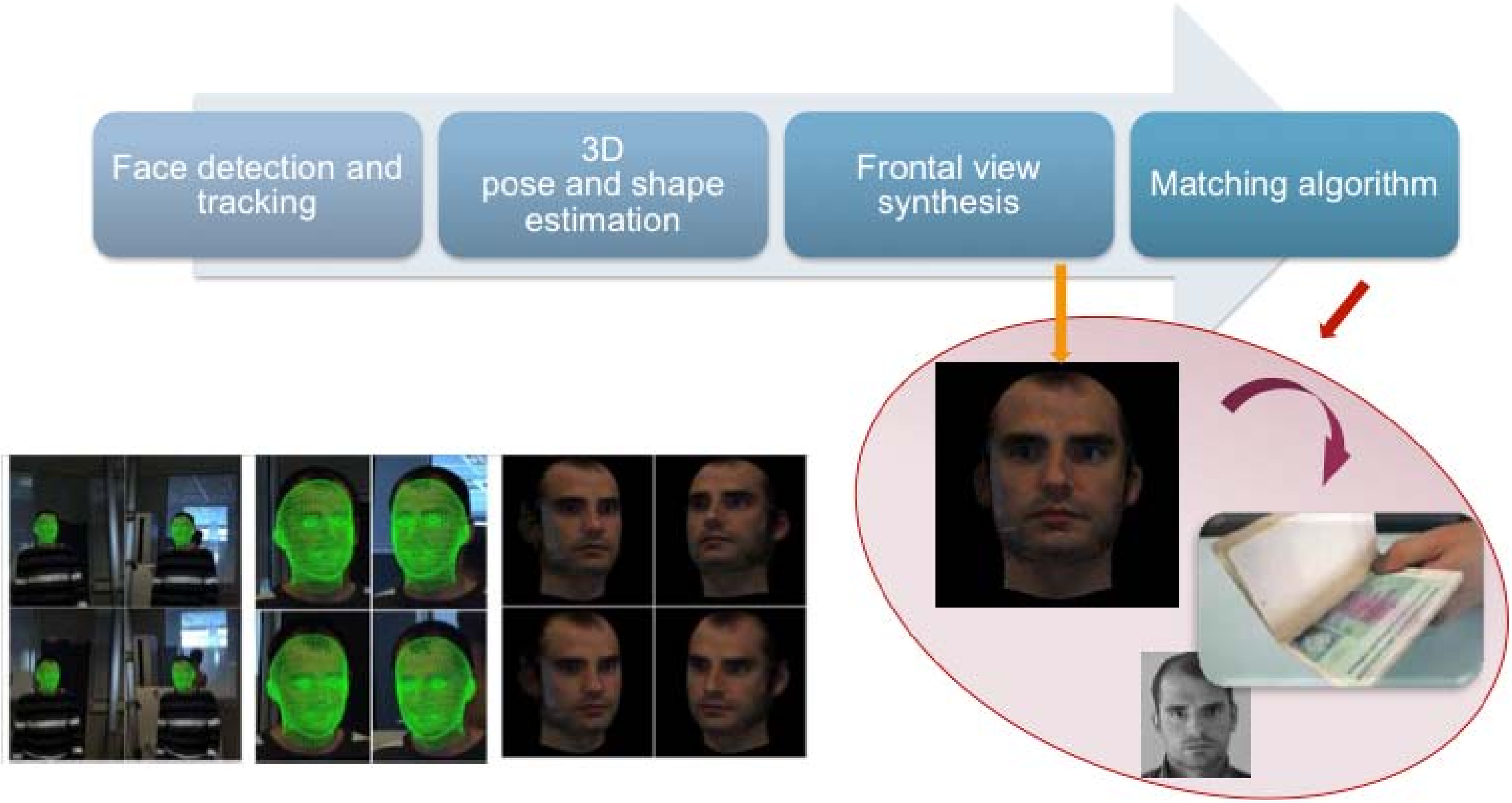
# Commercial FR systems

Vision Lab

**From: *www.face-rec.org***

- 3M/Cogent *(2D + 3D rectification)*
- A4Vision, Inc. *(3D scanner)*
- AcSys Biometrics Corp.
- Animetrics Inc. *(3D shape)*
- Ayonix Inc. *(2D ???)*
- Betaface *(2D + hair & variable features)*
- Cognitec Systems GmbH *(LFA)*
- Crossmatch Tech. *(2D face capture)*
- Cybula Ltd. *(3D shape/2D texture)*
- Face.com *(2D mugshots)*
- DreamMirh Co., Ltd. *(2D ???)*
- Geometrix, Inc. *(3D shape)*
- Iconquest *(2D Fractal-based ??? )*
- L-1/Identix Inc. / Safran-Morpho *(2D Templates +LFA)*
- KeeSquare Srl *(2D landmark-based)*
- Luxand *(2D facial landmarks)*
- NeuroTechnology *(2D ???)*
- Omniperception *(2D + Quality measurements)*
- PittPatt/Google *(Hi-Tech algorithms ???)*
- SensibleVision *(2D Template matching ??? )*

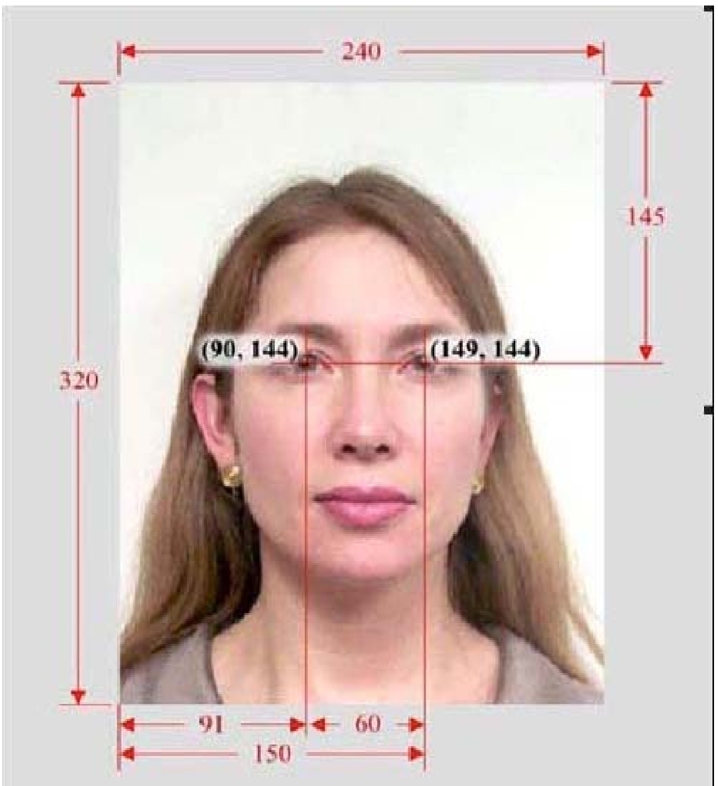
# MORPHOWAY V : WORKFLOW

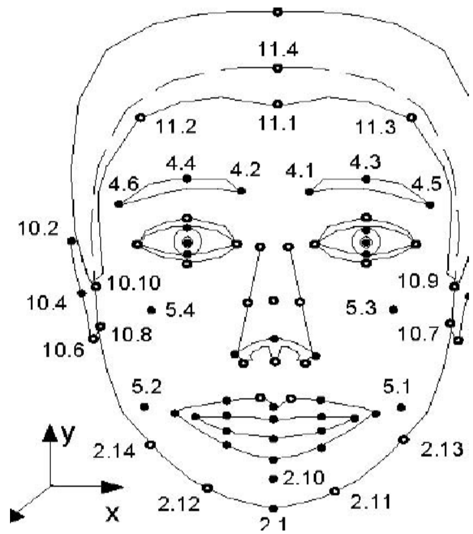


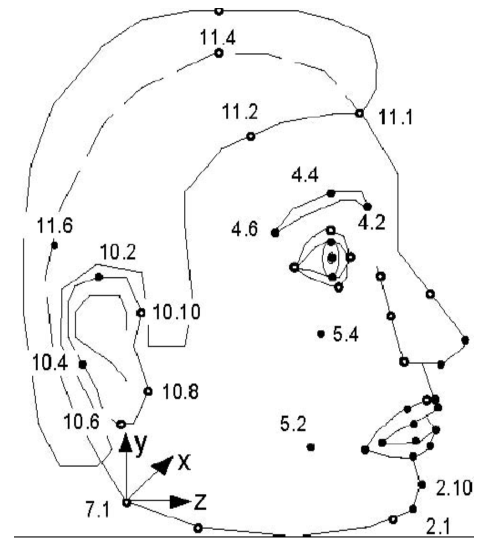



SAFRAN
Morpho

# The *standard* face




| Width(pixels) | Distance from Eye to Eye (Inclusive) |
|---|---|
| **240(min)** | **60** |
| 480 | 120 |
| 720 | 180 |

# Templates and standards

## Eurosmart white paper 2003

**Figure 2: Biometric Template Size**
*Source: Frost & Sullivan*

| Biometric | Bytes Required |
|---|---|
| Finger-scan | 300-1200 |
| Finger geometry | 14 |
| Hand geometry | 9 |
| Iris recognition | 512 |
| Voice verification | 1500 |
| Face recognition | 500-1000 |
| Signature verification | 500-1000 |
| Retina recognition | 96 |

- *It is generally anticipated that any single biometric template would fit within 10 Kbytes of data memory in the storage device (That includes the template itself, the signature or encryption overhead and any additional data required to fulfill the data file structure).*