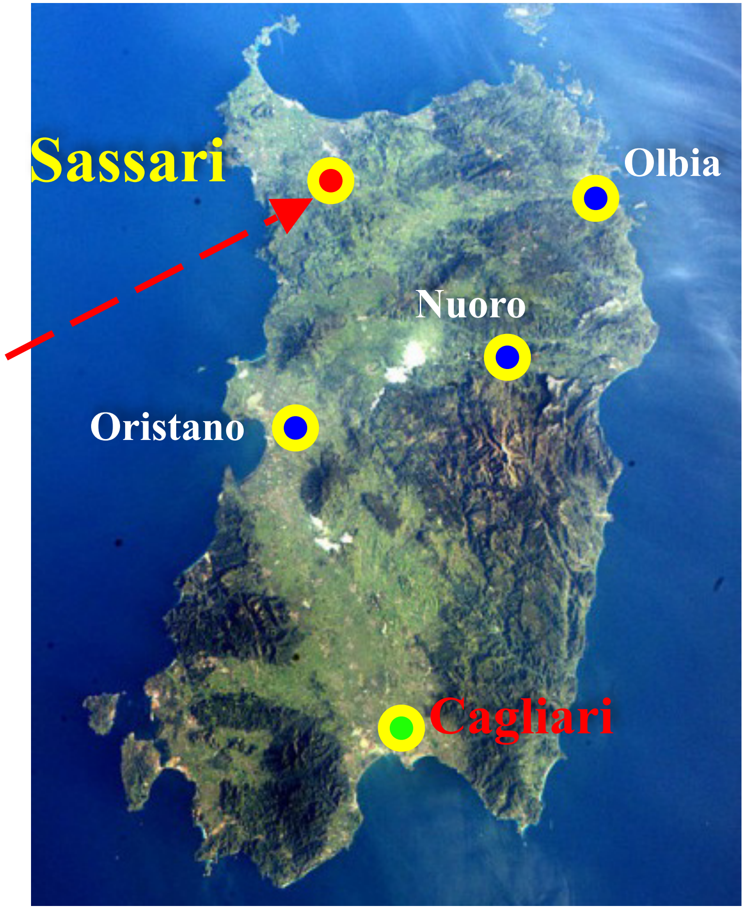
Vision Lab

Face Recognition: Past, Present and Future

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

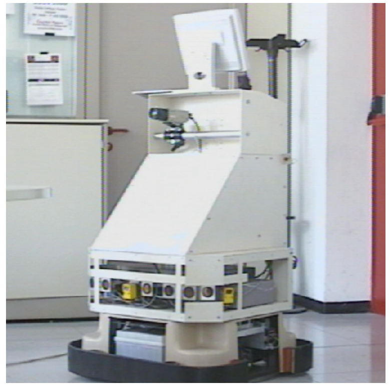
The University of Sassari

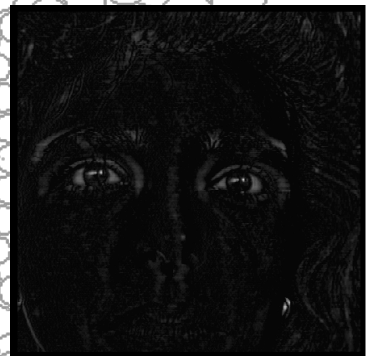
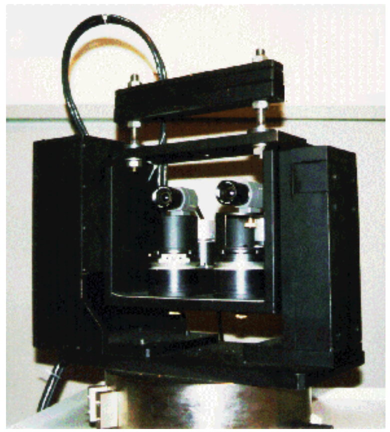
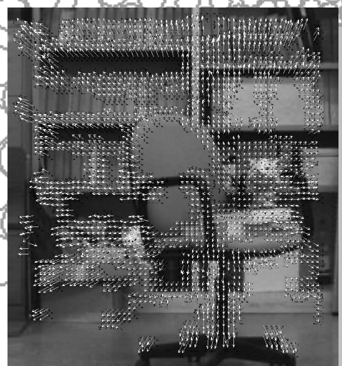
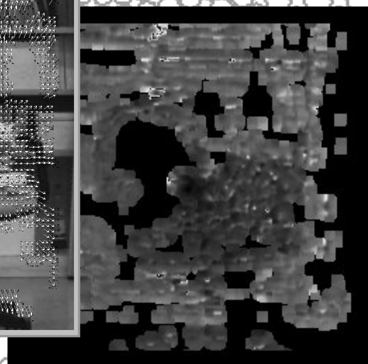
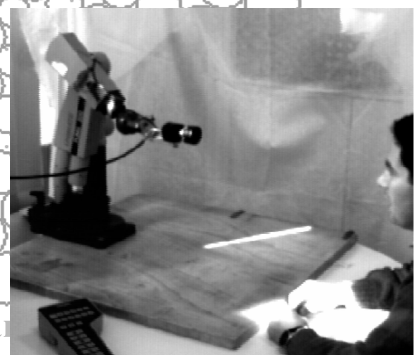


The Computer Vision Lab

Stems from a well established core group with a strong background in:

- Motion and Stereo vision / fusion
- Non standard image geometries
- Sensory-motor coordination
- Robotic sensing/navigation
- Active and dynamic vision
- Visual Recognition
- Biometrics



The Computer Vision Lab

Since 2003 hosting the Int.l Summer School on Biometrics

Credits

- From the laboratory staff:

  Linda Brodo
  Marinella Cadoni
  Filippo Casu
  Massimo Gessa
  Enrico Grosso
  Souad Khellat Khiel
  Andrea Lagorio
  Ludovica Lorusso
  Gianluca Masala
  Heydi Mendez (past visiting)
  Norman Poh (past visiting)
  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

Vision Lab

…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
Chang-Tsun Li – University of Warwick
Jonathon Phillips – NIST
Norman Poh – University of Surrey

IC1106 - Integrating Biometrics and Forensics for the Digital Age

Multi-Foresee

IDENTITY

Computer Vision Enabled Multimedia Forensics and People Identification

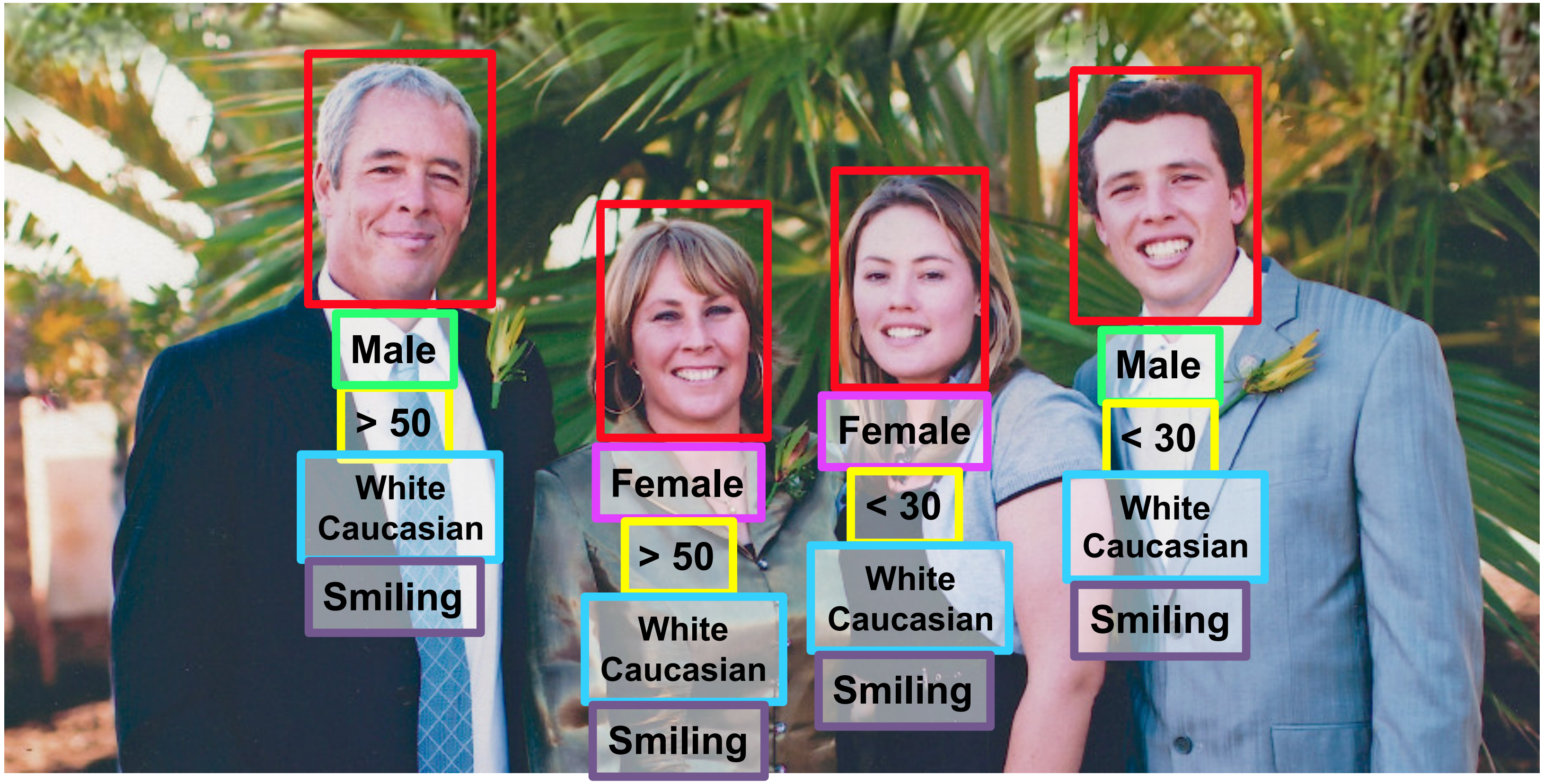
Face recognition

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

II. (**PRESENT**) **What can we learn?**

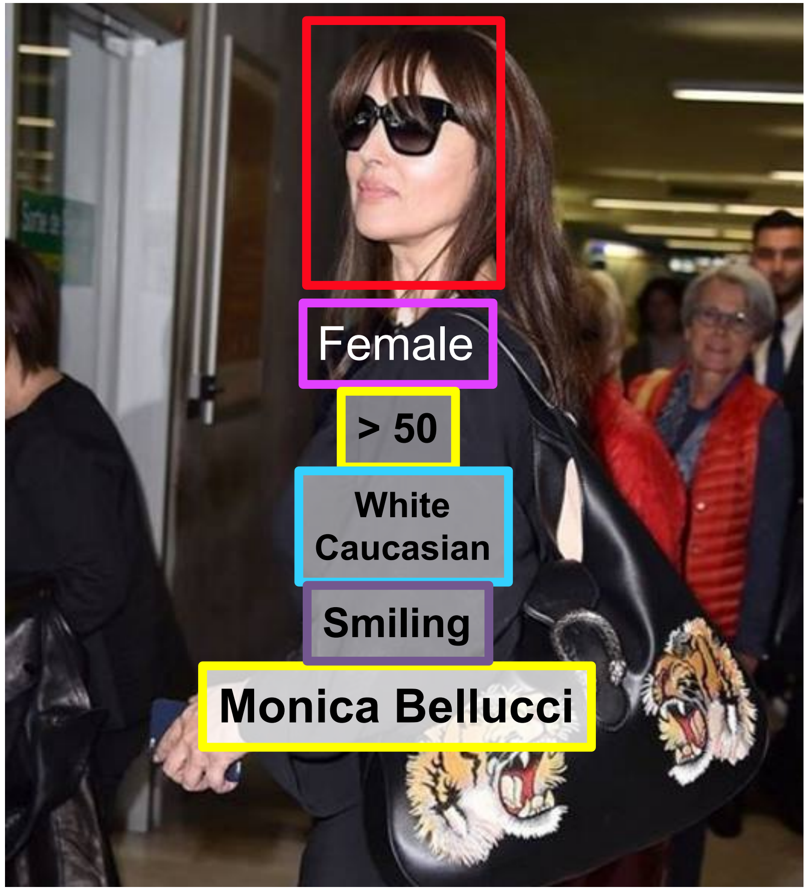
III. (**FUTURE**) **What is still to be done?**

Biometric Face Analysis



Information from a single face image

Biometric Face Analysis

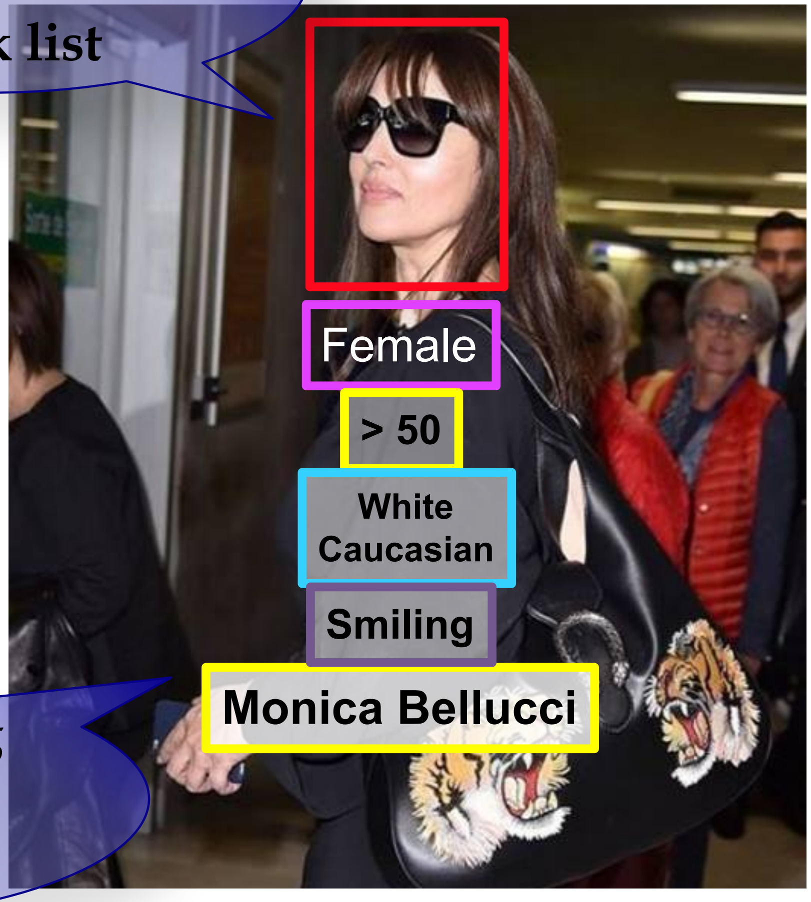


Information from two face images

Biometric Face Analysis

Monica Bellucci

Not found in China's terrorist black list



Not found among Shenzhen University staff

Information from many face images

CSI Fiction

Media Fiction

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 …*

Elements of Face Biometrics

Image pre-processing

- Face detection
- Normalization

Feature-level processing

- Features extraction
- Features selection

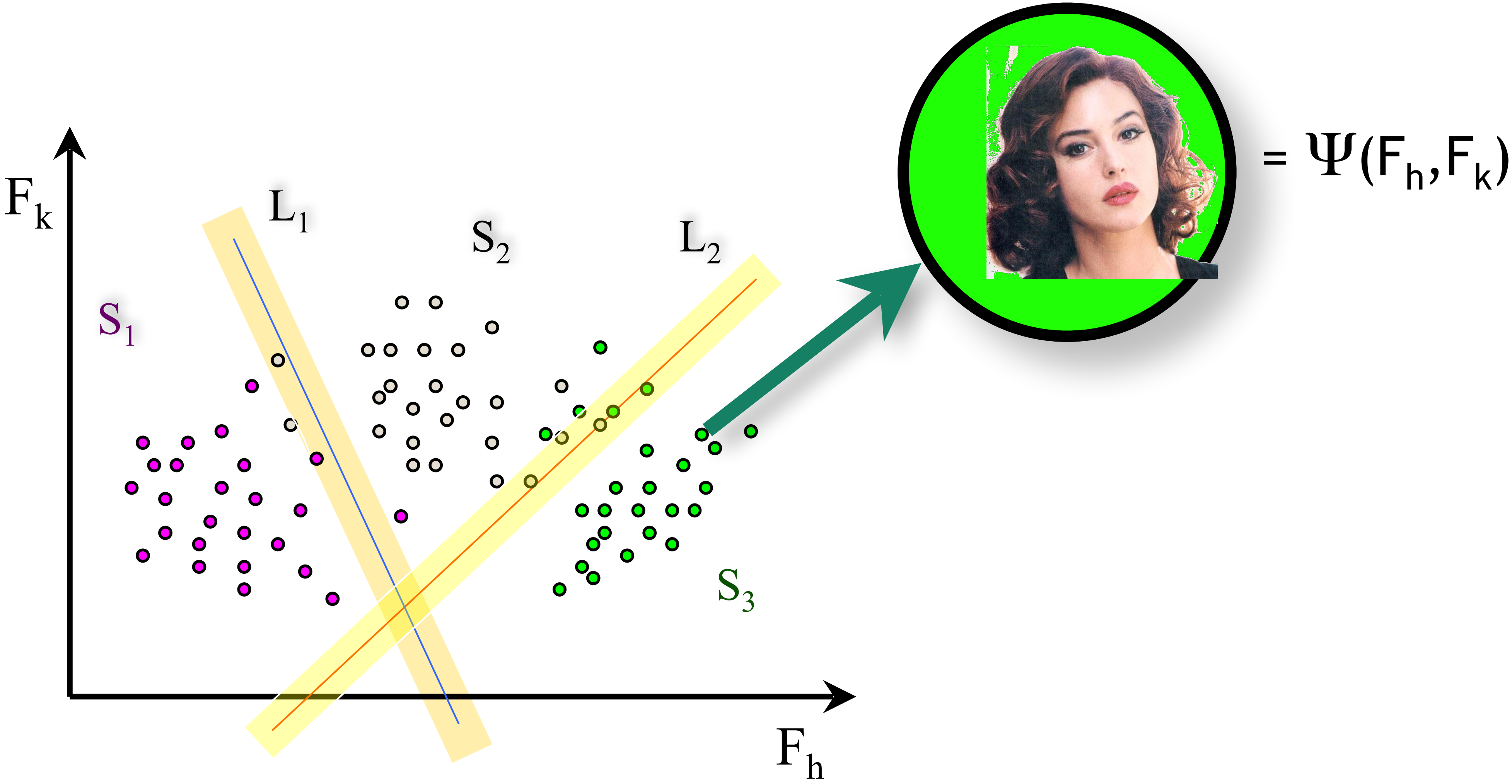
Matching

- Classification
- Recognition

Face Identification

Vision Lab

A class (identity) separation problem



Biometric uniqueness

Vision Lab

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

Inter-class *similarity*

Two different people with very similar appearance

FALSE <u>MATCH</u>

www.marykateandashley.com

Twins

news.bbc.co.uk/hi/english/in_depth/americas/2000/us_elections

Father and son

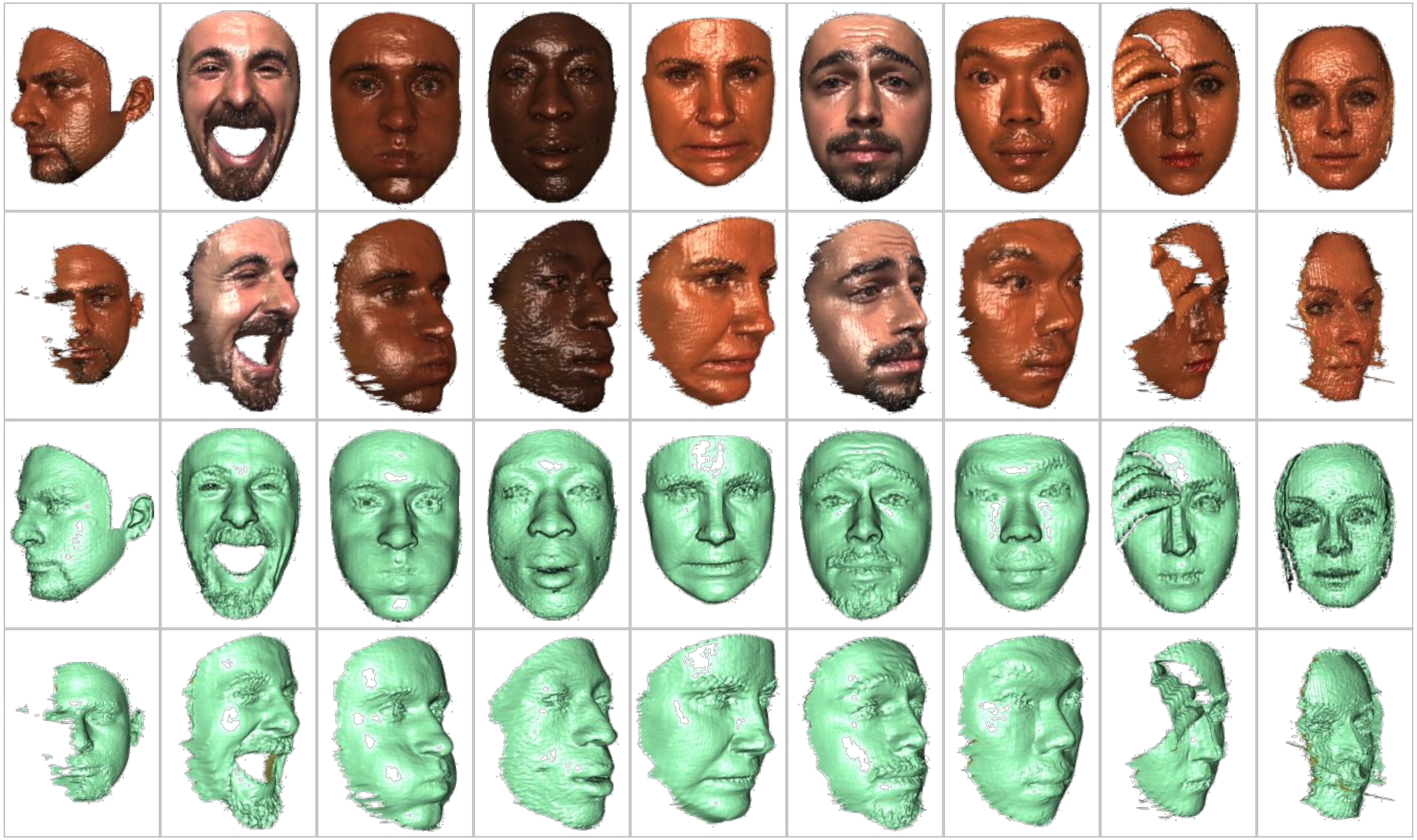
Intra-class *variability*

The same person may present very different biometric samples

FALSE <u>NON-MATCH</u>

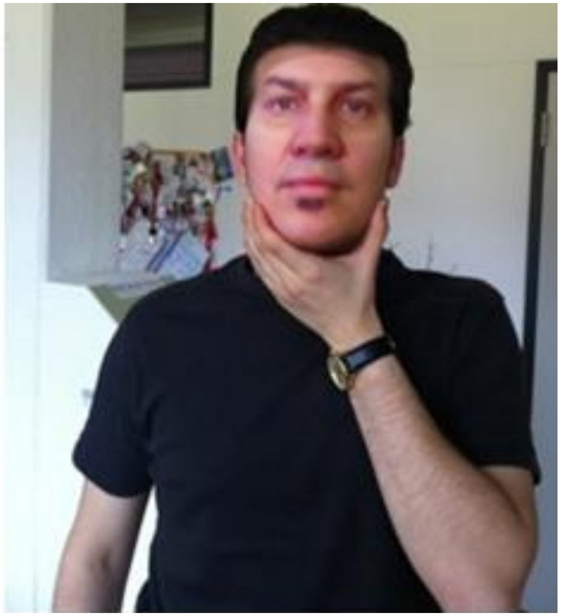
Monica Bellucci

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 spoofing and impersonation



Face surgery

Face makeup

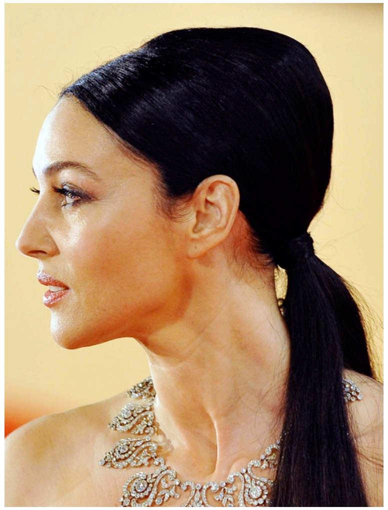
Four Big Problems

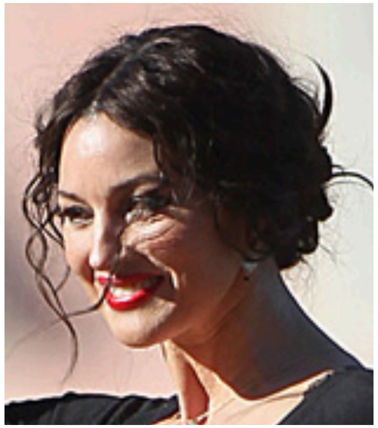
A – Aging

P – Pose

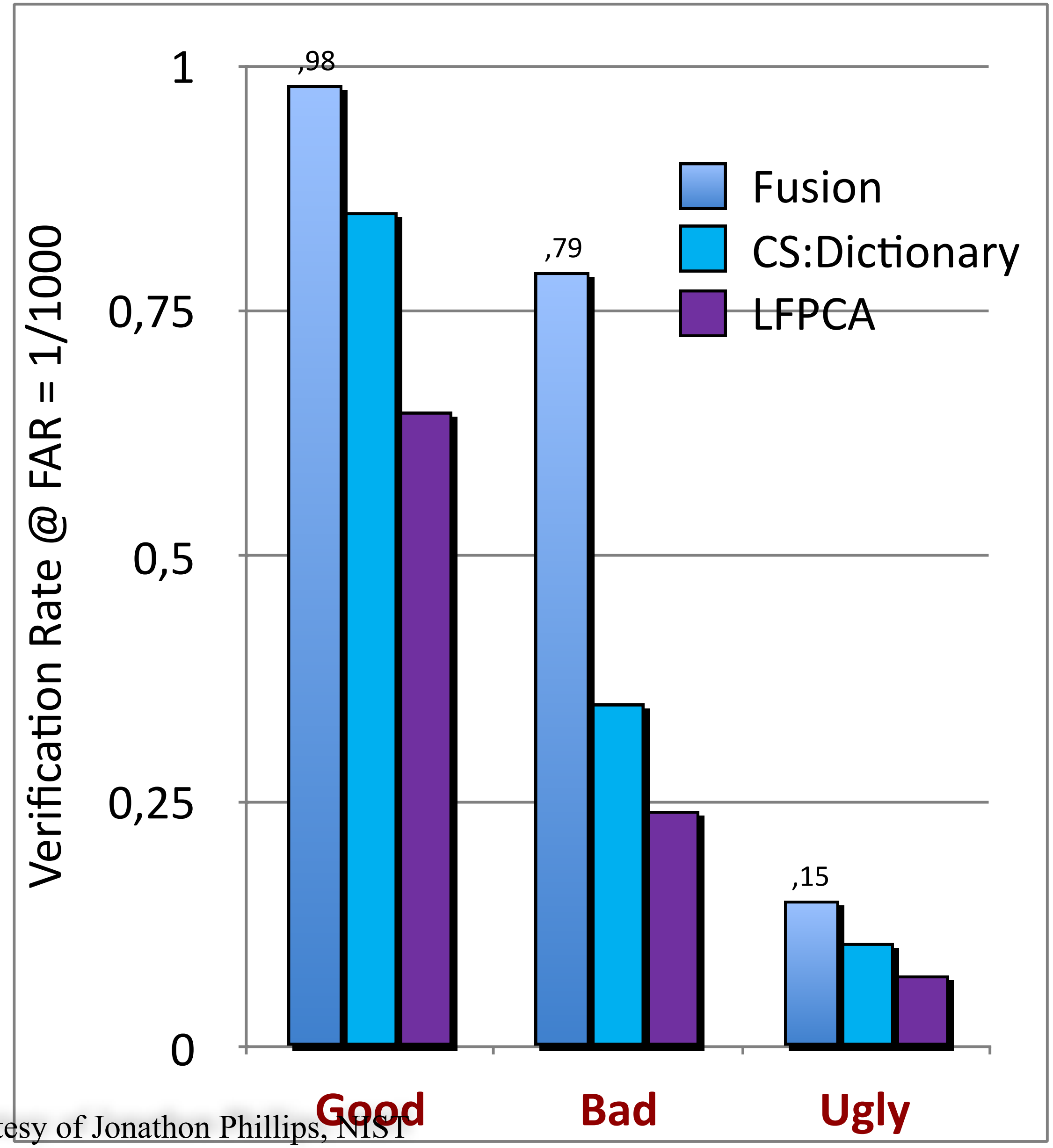
I – Illumination

E - Expression

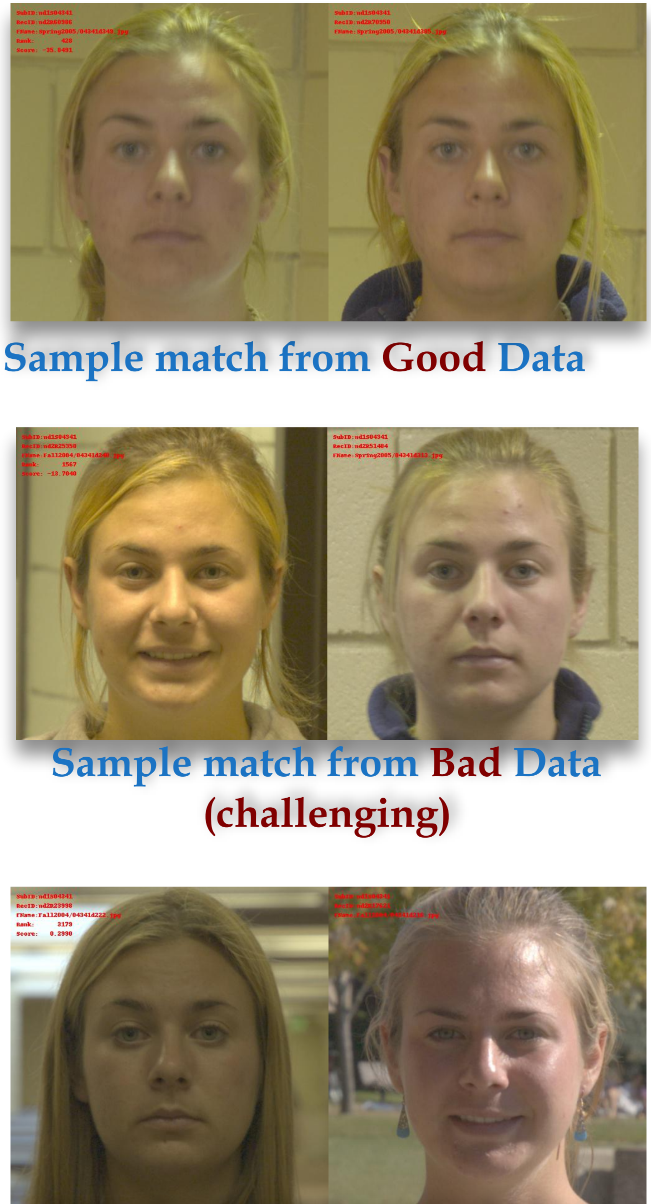




The “Good, Bad, Ugly”



Courtesy of Jonathon Phillips, NIST



Sample match from Good Data

Sample match from Bad Data (challenging)

Sample match from Ugly Data (very challenging)

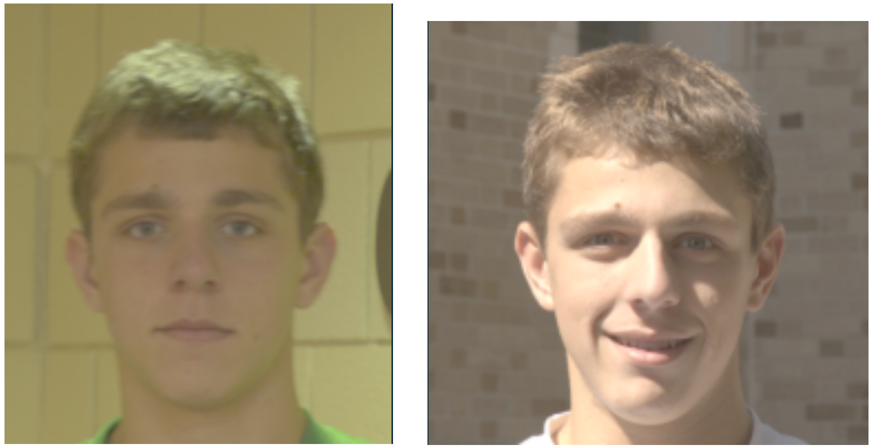
Lighting & Expression

Same lighting, Same expression

Different lighting, Same expression

Same lighting, Different expression

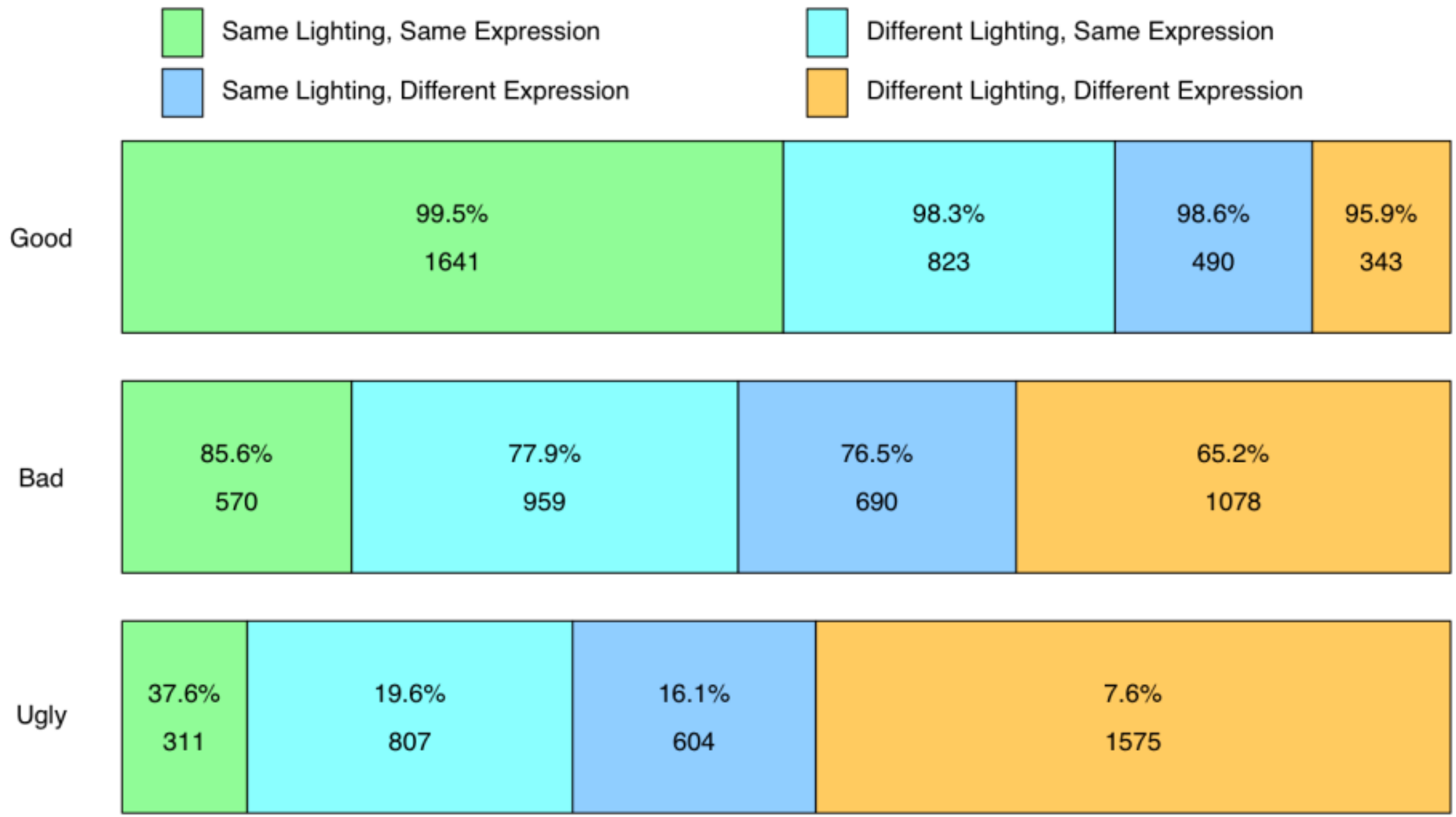
Different lighting, Different expression



Courtesy of Jonathon Phillips, NIST

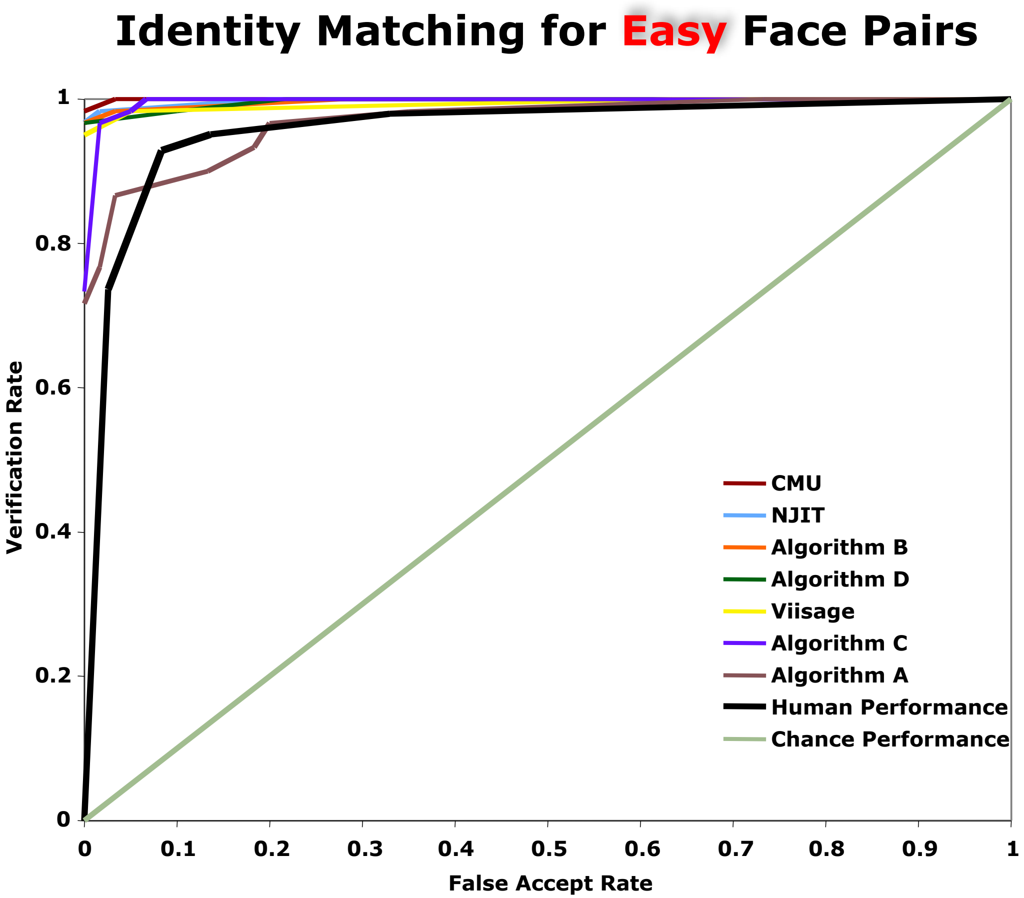
Lighting & Expression

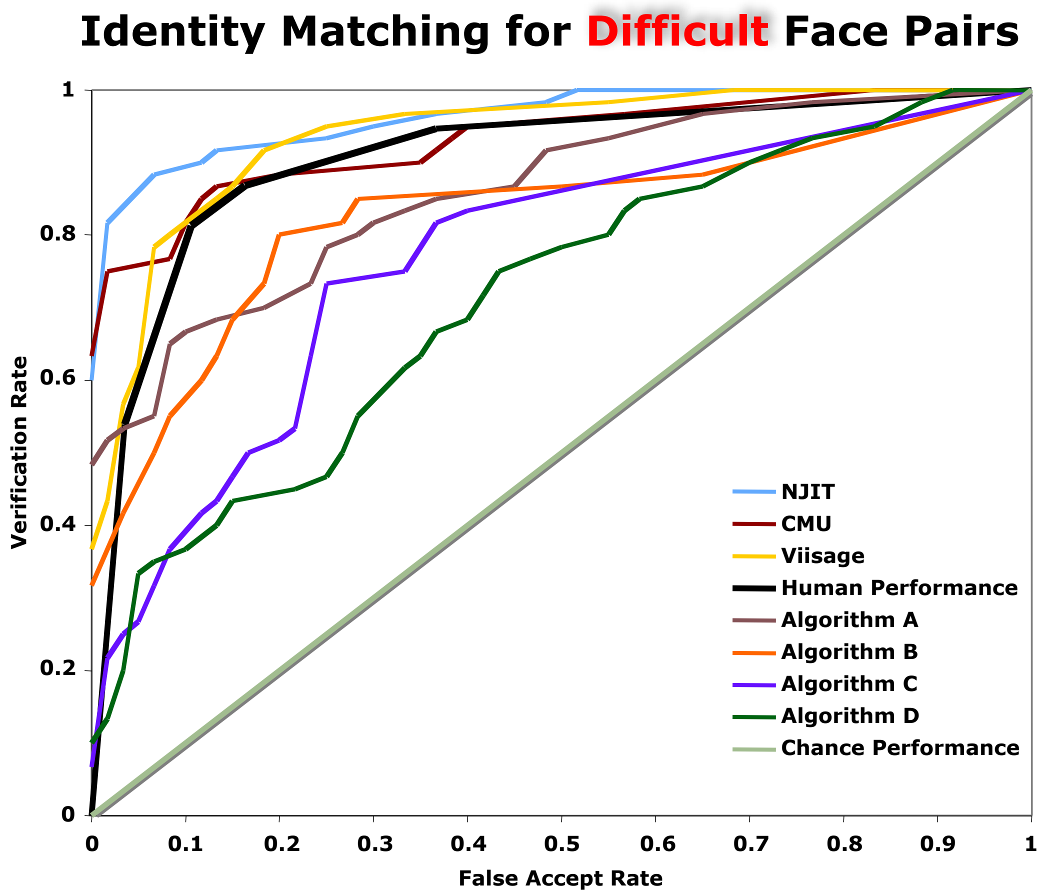
Verification rate @ FAR = 0.1%



Courtesy of Jonathon Phillips, NIST

Human vs Machine performances





A. O'Toole, J. Phillips, F. Jiang, Ayyad, Pénard & Abdi, "Benchmarking Algorithms Against Humans" *IEEE:T-PAMI, 2007*

An ill-posed problem

A problem is said to be *well-posed* in the sense of Hadamard when (1) a *unique* solution exists and (2) it depends *continuously* upon the data.

J. Hadamard, Sur les problemes aux derivees partielles et leur signification physique. In: Princeton University Bulletin, 1902, 49–52.

An ill-posed problem

Two adverse conditions:

1. Noise in the data (many sources)
2. Dimensionality of the data (from 4D to 2D)

Solution: Regularization $\boxed{\mathbf{T}(\boldsymbol{q},\boldsymbol{f}_\delta,\alpha) = ||A\boldsymbol{q} - \boldsymbol{f}_\delta||^2 + \alpha\, \boldsymbol{\Omega}(\boldsymbol{q} - \boldsymbol{q}^0)}$

A.N. Tikhonov, On the stability of inverse problems. Doklady Acad. Sci. USSR 39 (1943), 176–179.

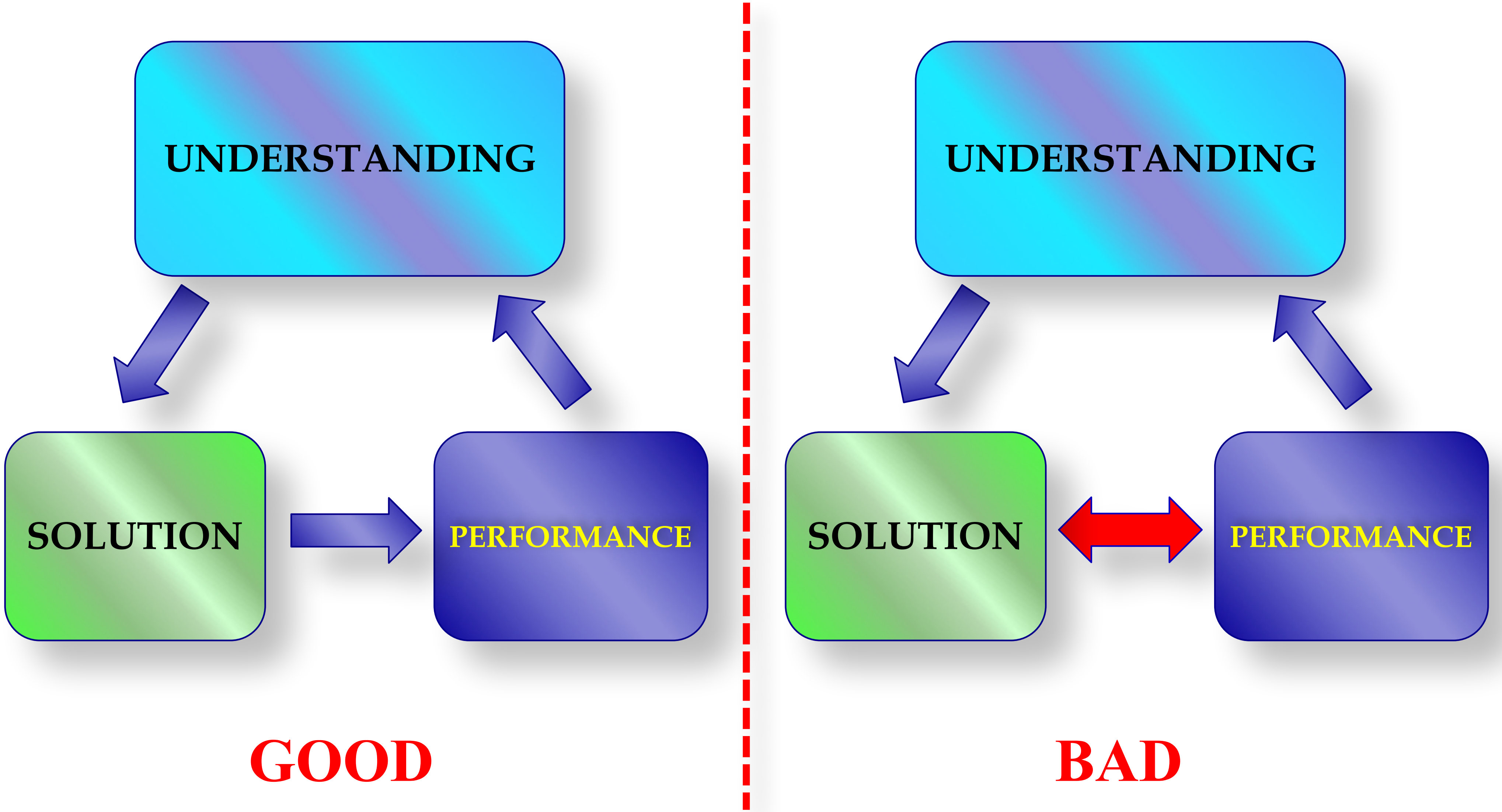
A.N. Tikhonov, On the solution of ill-posed problems and the method of regularization. Dokl. Akad. Nauk SSSR 151, 3 (1963), 501–504 (in Russian).

A.N. Tikhonov, On the regularization of ill-posed problems. Dokl. Akad. Nauk SSSR 153, 1 (1963), 49–52 (in Russian).

A. N. Tikhonov and V. Ya. Arsenin, Solutions of Ill-Posed Problems. Wiley, New York, 1977.

FACE RECOGNITION RESEARCH AND TECHNOLOGIES

Good research or bad research?



Common mistakes

1. Start programming before thinking;
2. Building a system blindly combining a number of already available algorithms;
3. Performing blind tests with already available tools;
4. Performing blind tests on available datasets;
5. Twickling the parameters until you obtain the desired performance;
6. Arbitrarily selecting the data from the available dataset on which perform the testing;
7. Making strong statements without a solid proof;
8. Making unrealistic assumptions.

Addressing the problem

1. Analyze the problem, the available data and the constraints;
2. Make some bibliographical search (don't try to re-invent the wheel…), on an up-to-date search engine;
3. Define a model describing the physics of the event;
4. Find a mathematical framework which may help to find a solution;
5. Carefully design an experimental set-up;
6. Collect or acquire a statistically meaningful dataset;
7. Start programming;
8. Perform an evaluation test to define the parameters space;
9. Start testing and collecting results, especially the failing modes;
10. Perform a comparative analysis of the results with other systems.

Face recognition

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

II. **(PRESENT) What can we learn?**

III. **(FUTURE) What is still to be done?**

Taxonomy of face recognition

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 . . .

- 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

Deep Learning…

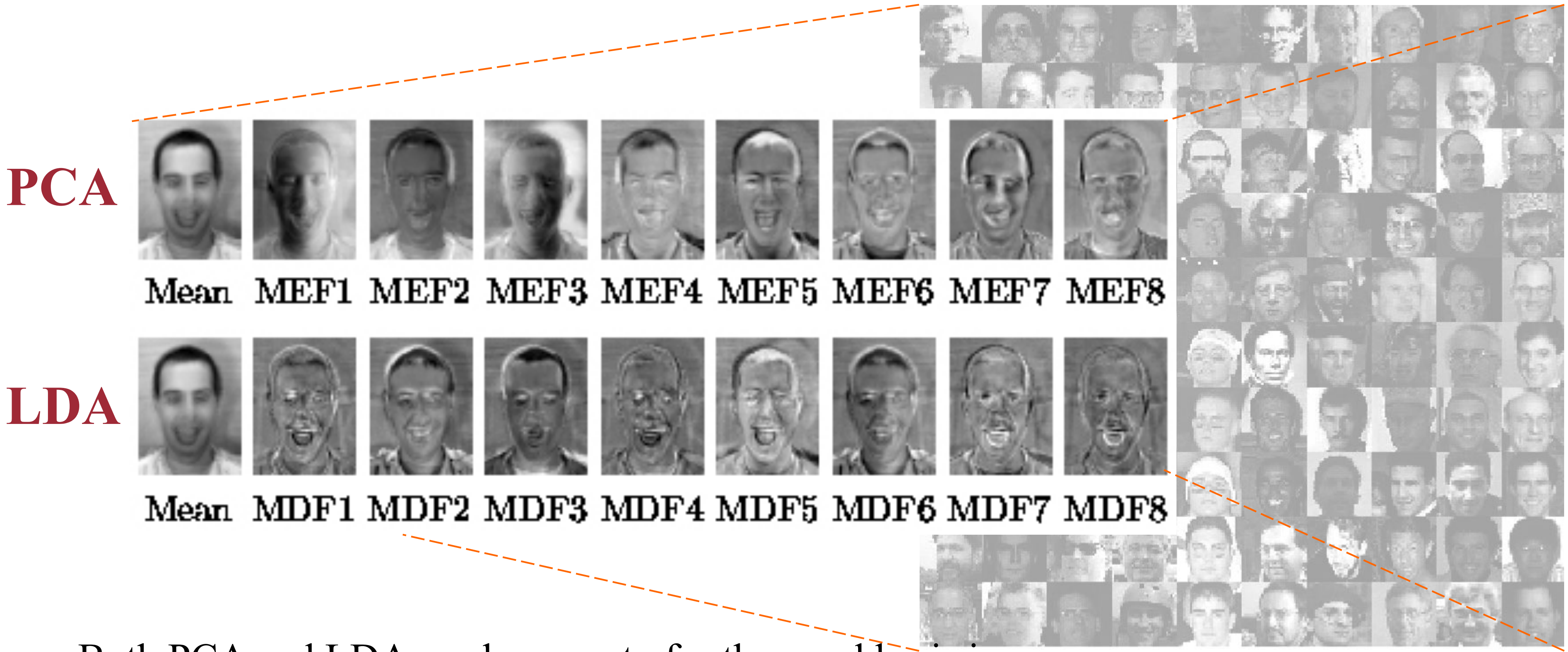
Holistic face recognition

- 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 σ_i define the "relevance" of each eigenvector (*eigenface*)

Holistic face recognition



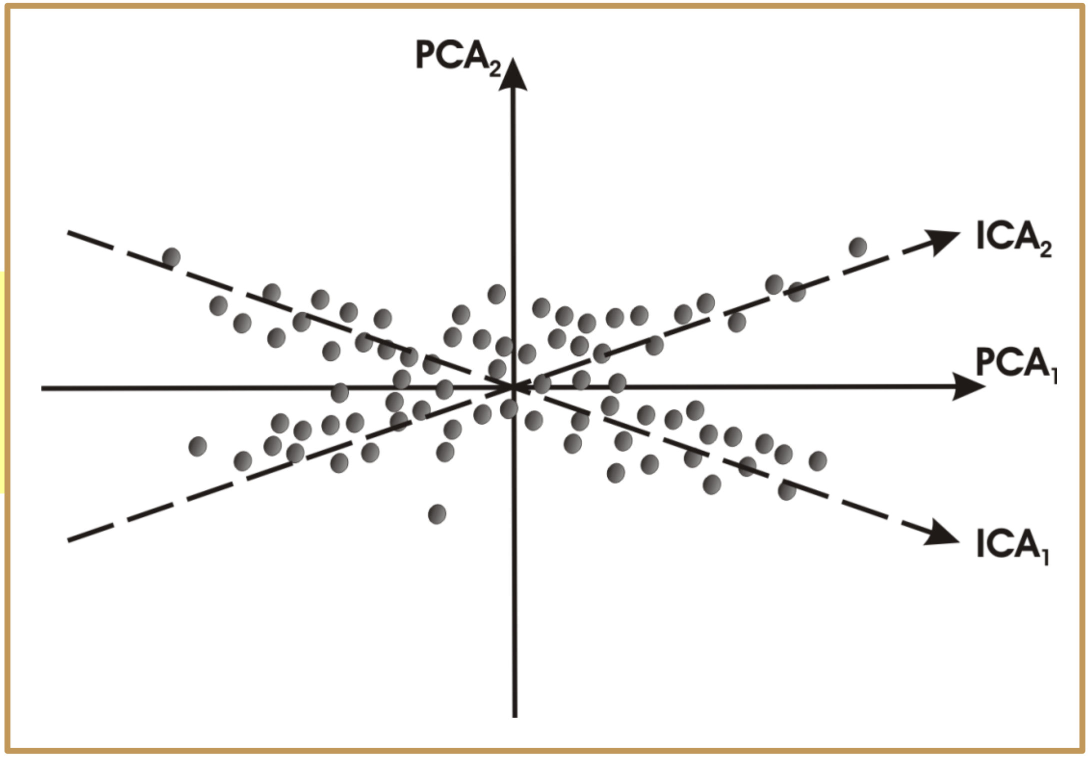
- Both PCA and LDA produce a set of orthogonal basis images.
- Both provide a compact and global representation of face images.
- LDA explicitly attempts to model the difference between the classes of data.
- PCA does not take into account any difference in class.

Holistic face recognition

Vision Lab

ICA

Orthonormality
vs
Maximal variance



PCA

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

Vision Lab

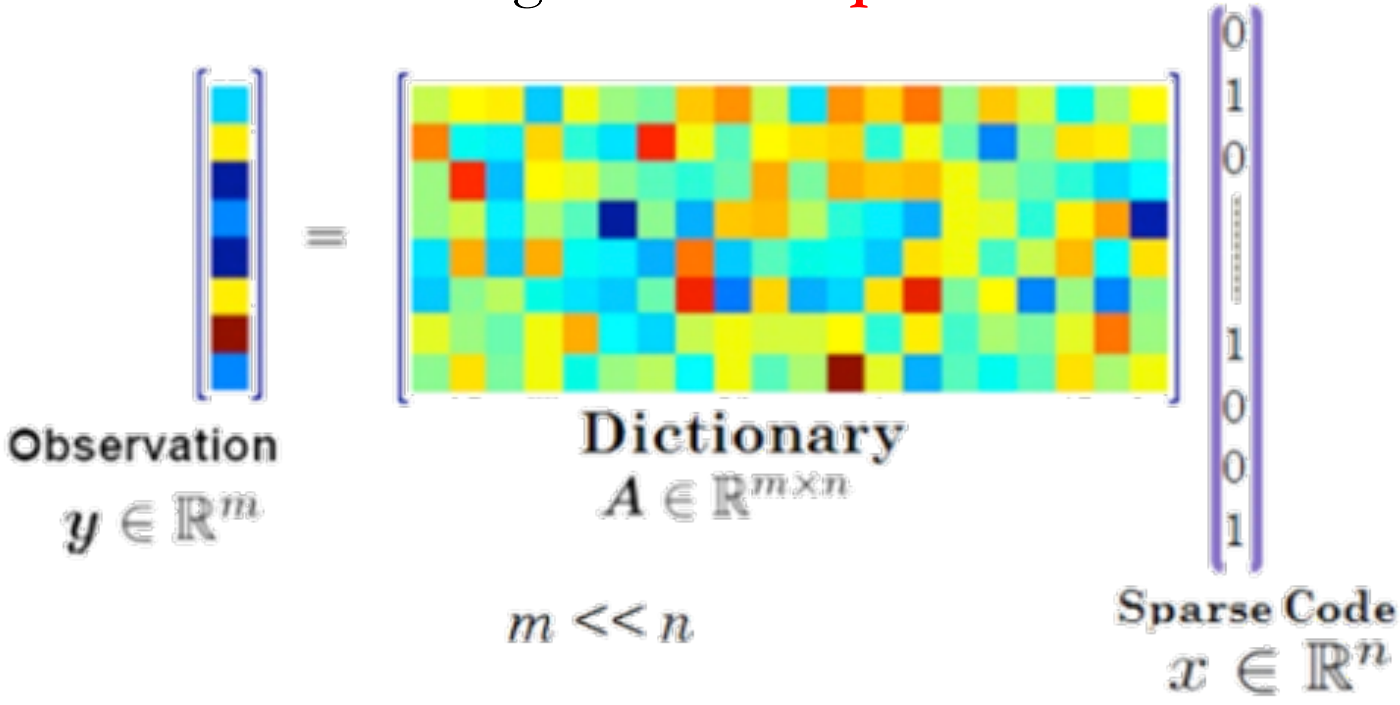
- Automatic face recognition algorithm robust to occlusion, expressions and disguise.
  - Represent the test face as a *sparse linear combination* of the training faces.
  - Estimate the class of the test image from the sparse coefficients.
- Can identify and reject "non face" images.
- …Performance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

Sparse Representations

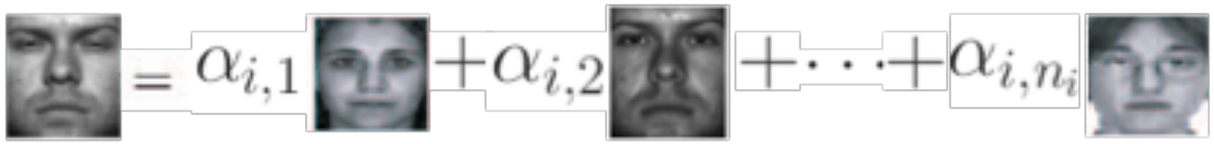
John Wright et al. PAMI 2009

Represent the test face as a **sparse linear combination** of the training faces.

Estimate the class of the test image from the **sparse coefficients**.



Dictionary matrix A

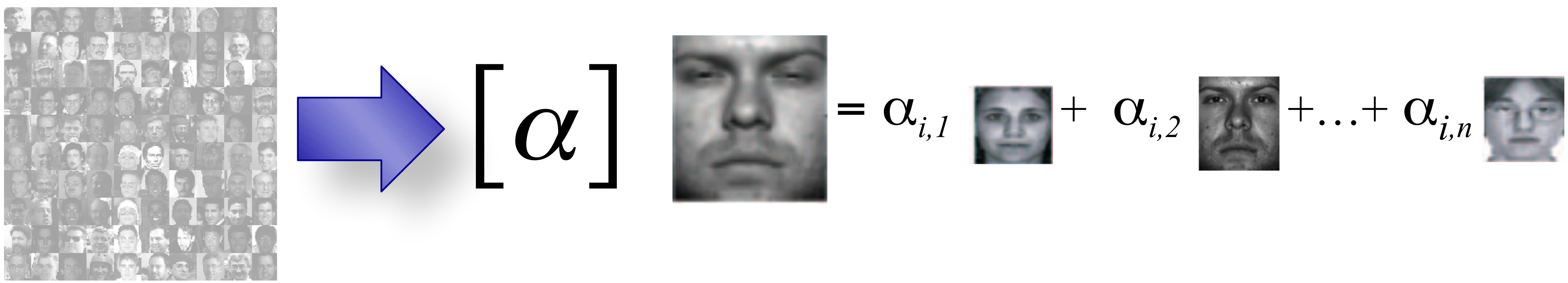


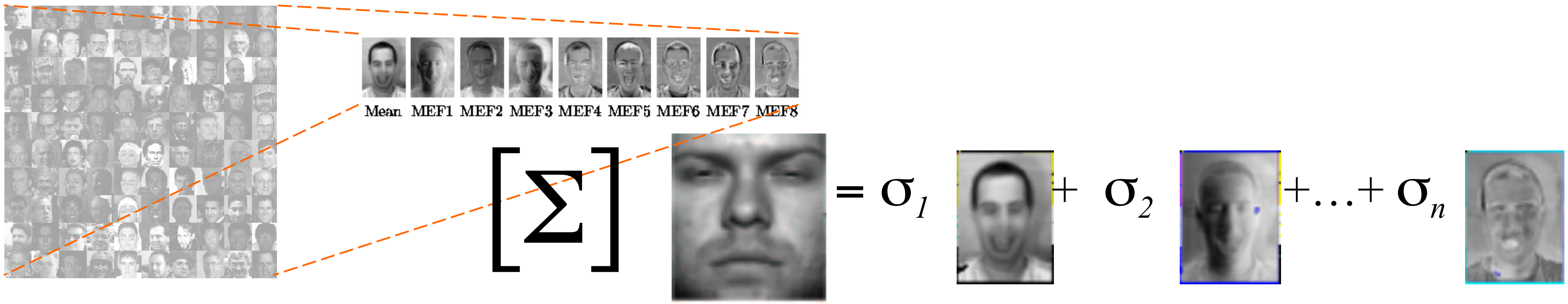
$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \cdots + \alpha_{i,n_i} v_{i,n_i}$$

Holistic face recognition

Sparse Representation *vs* Principal Component Analysis

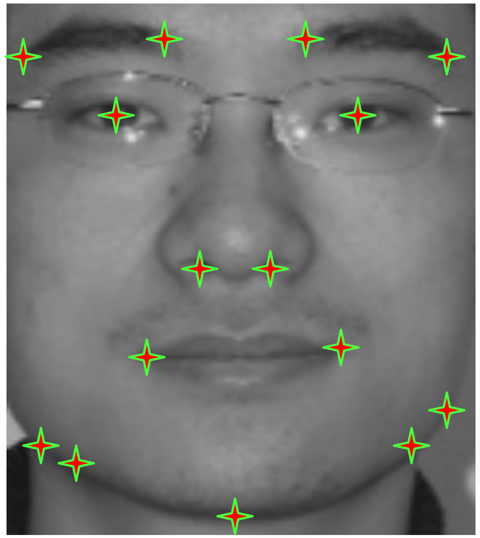
Similar formulation, different objectives

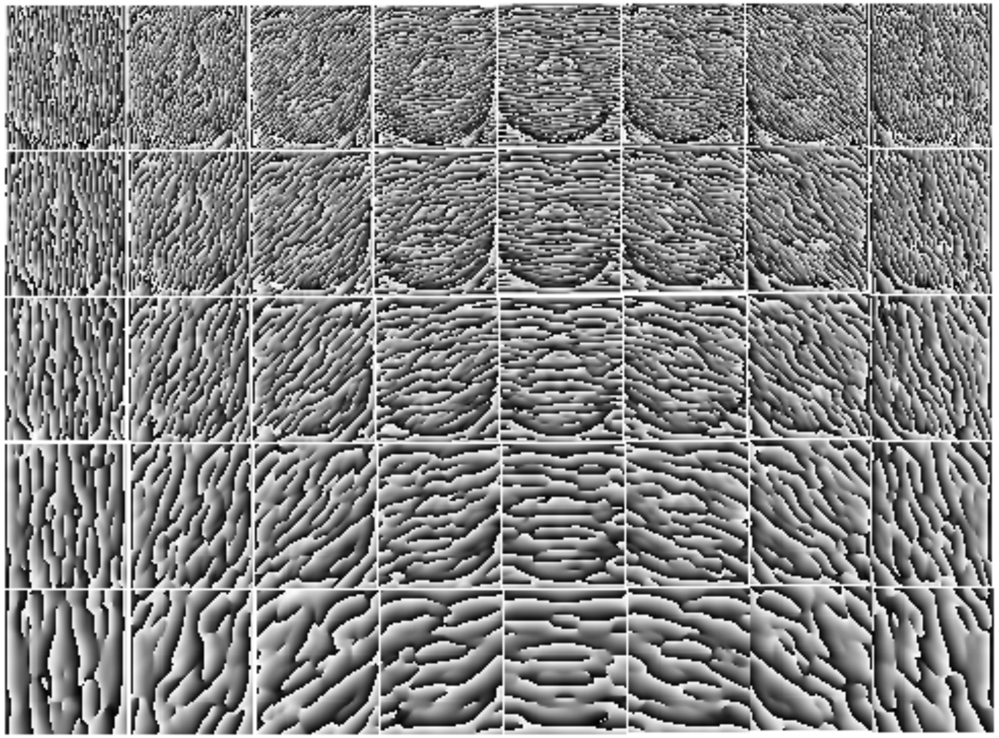




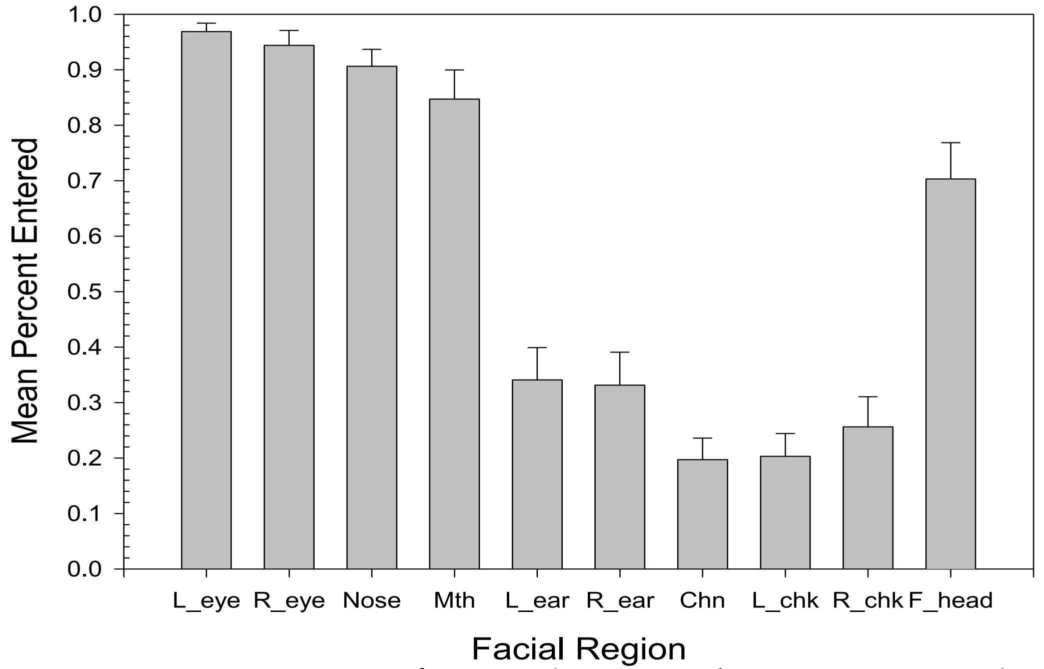
Facial features

- Physical Landmarks
- Gray level oriented patterns/photometric properties

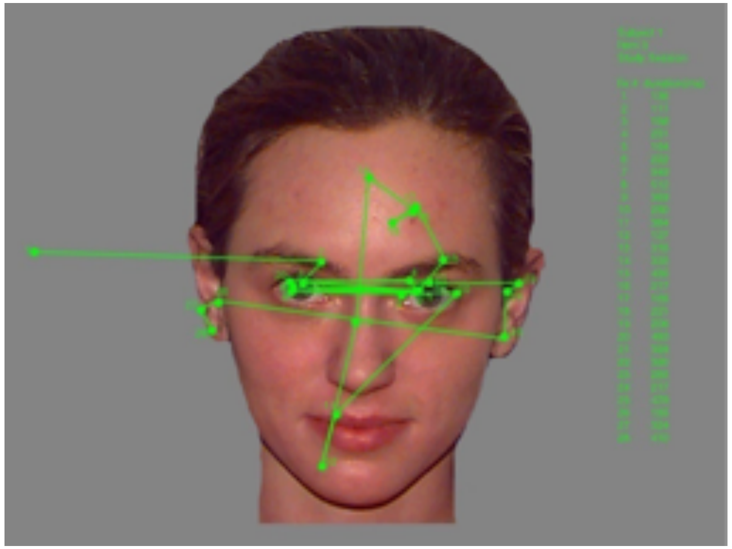


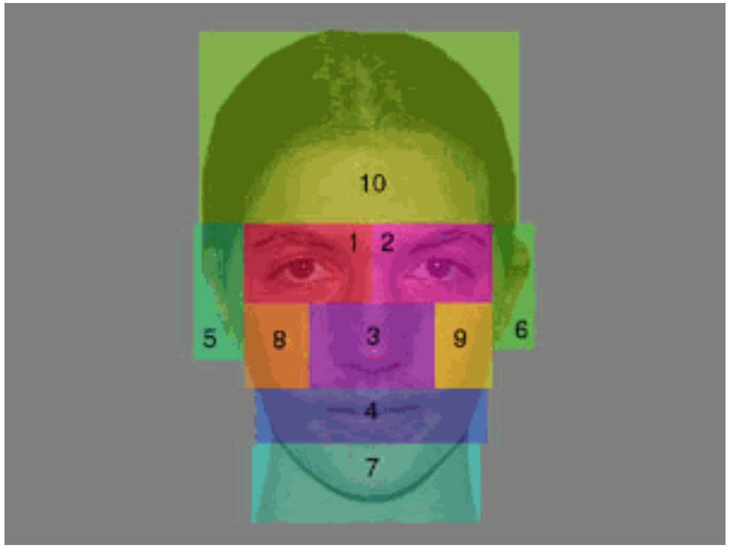


Facial features as landmarks



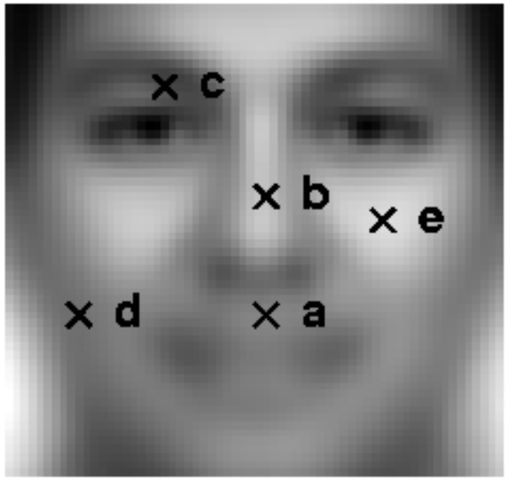
Mean percentage of times (averaged across viewers) that each facial region was fixated at least once.



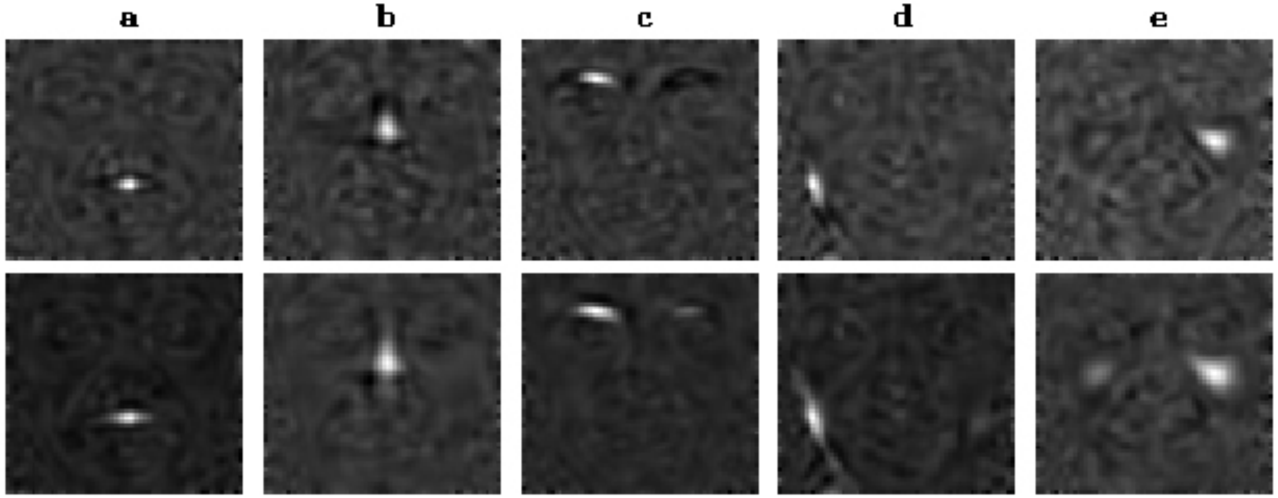


J.H. Henderson et al. "Gaze Control for Face Learning and Recognition by Humans and Machine"; in T. Shipley and P. Kellman (Eds.), From Fragments to Objects: Segmentation and Grouping in Vision

Facial features as 2D *landmarks*



Marked average face image



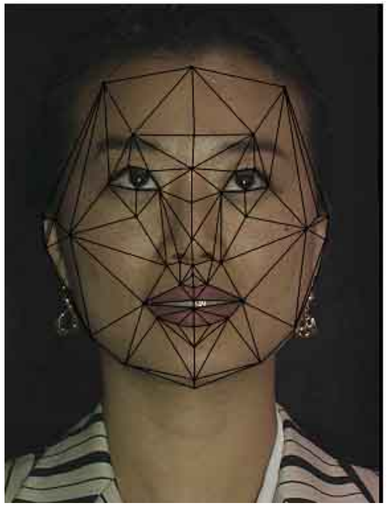
Top: five topographic kernels

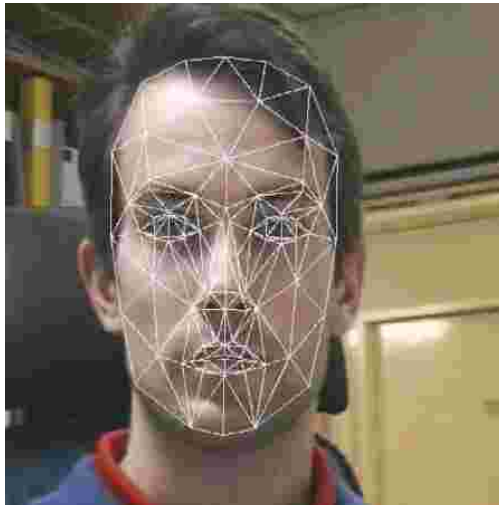
Bottom: five corresponding residual correlations

***Local Feature Analyisis* (LFA)**: localized kernels are built from PCA-based eigenvectors of topographic facial features.

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.

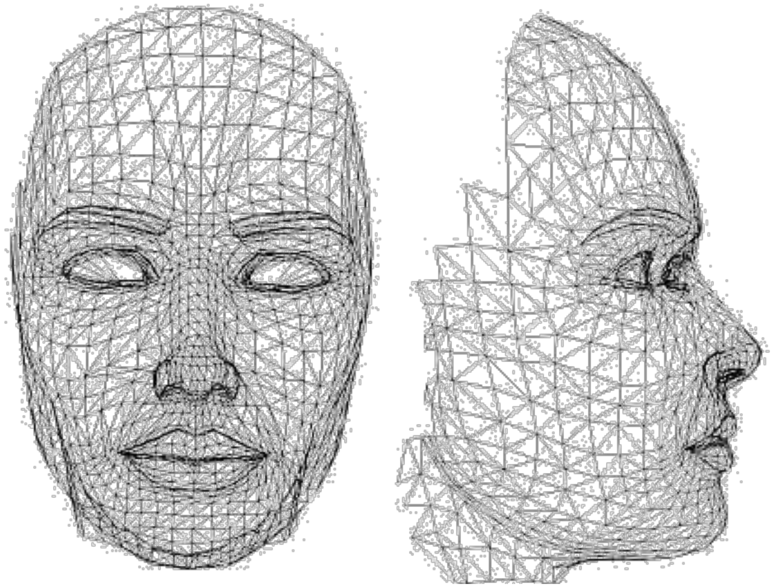
Facial features as 2D/3D *landmarks*





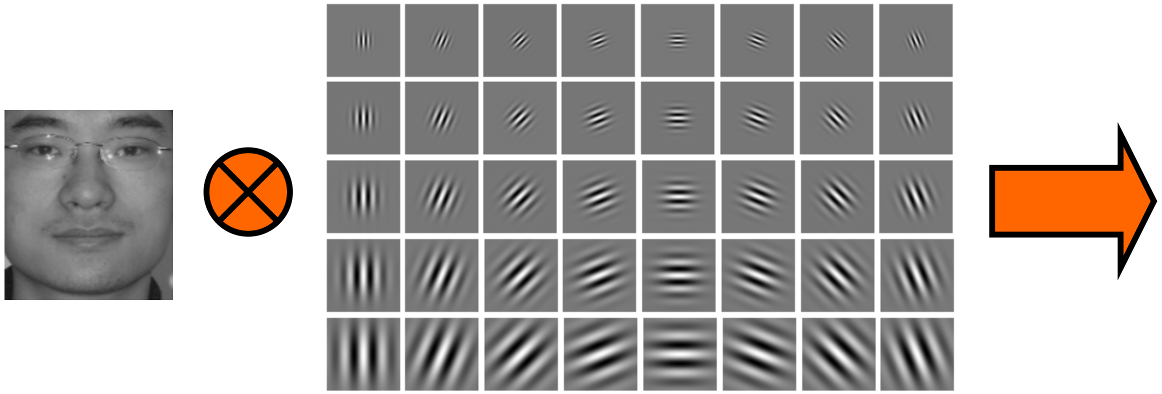
- 2D landmarks can be defined and tracked on face images

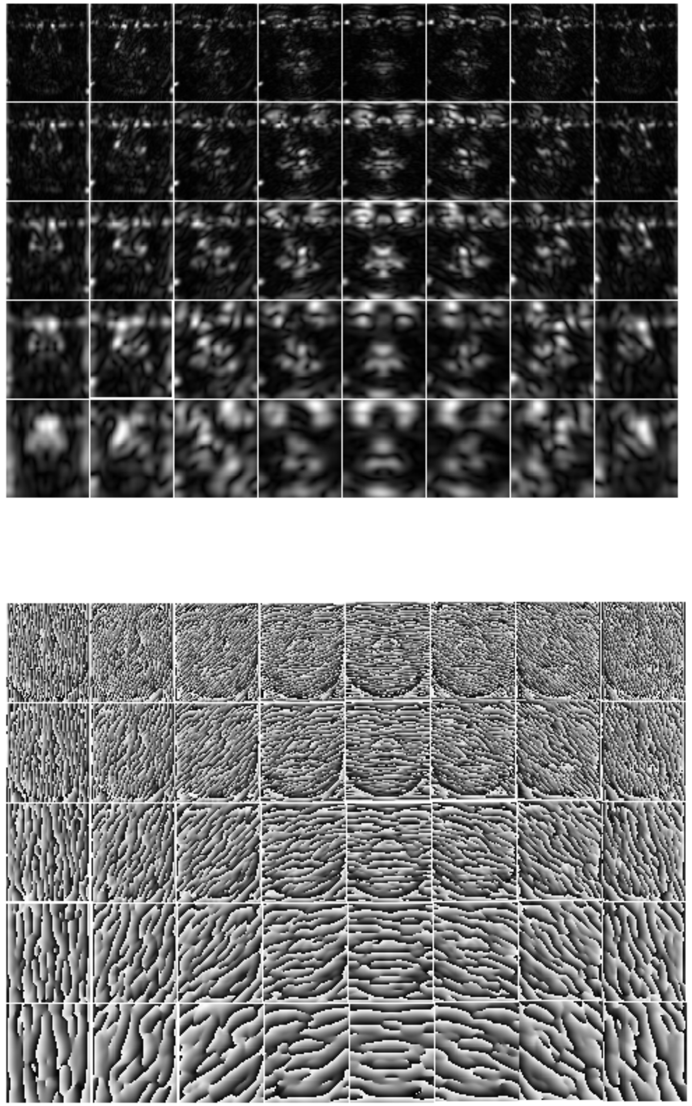
- Simple 2D vs complex 3D representations



Gabor wavelets

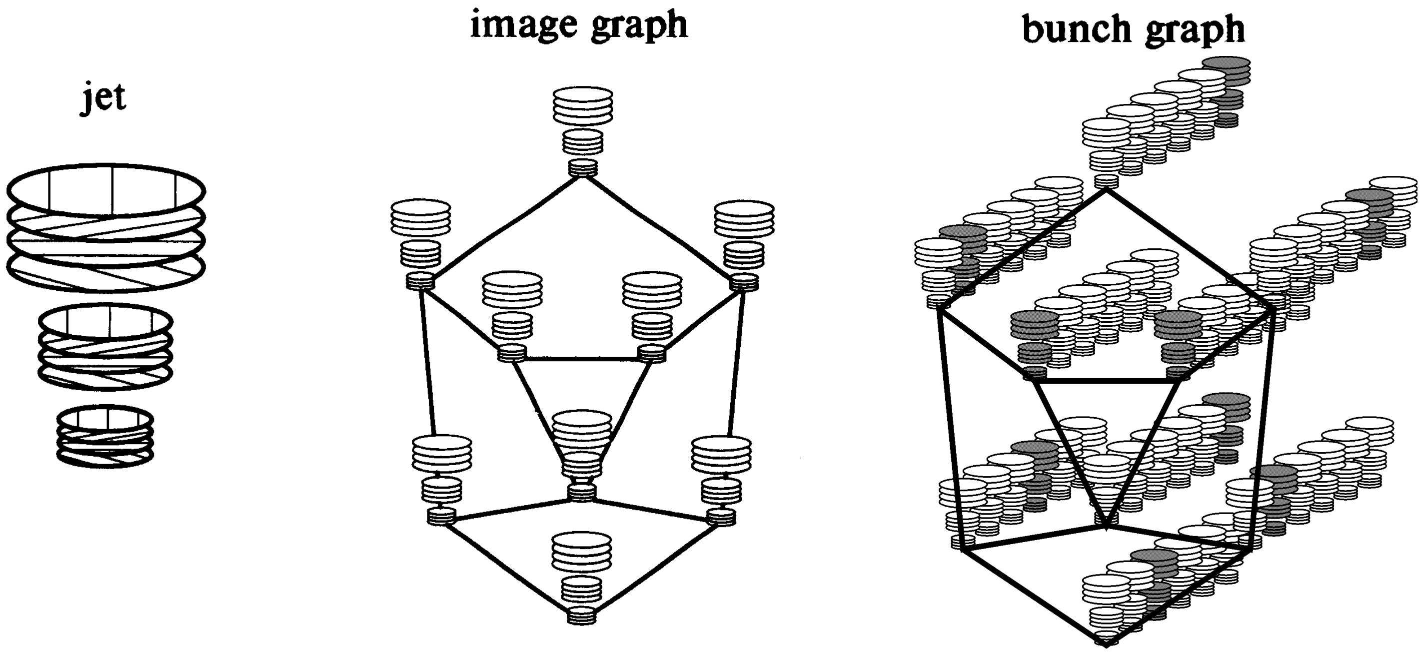
- Provide a vector description of the local structure of the facial patterns

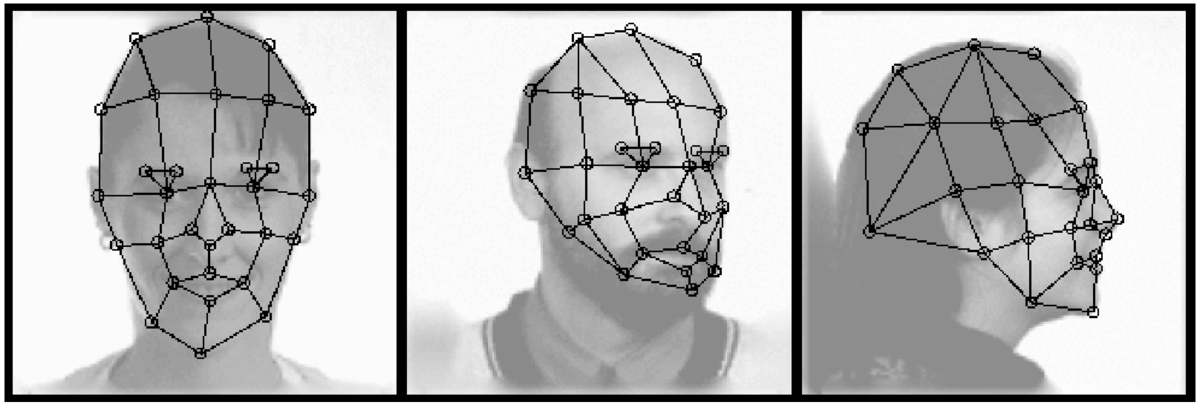




- Convolution with a bank of frequency-tuned filters

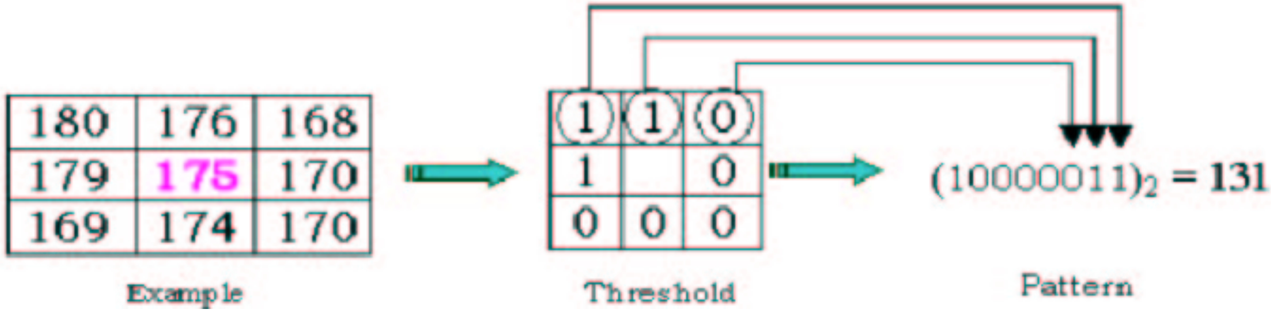
Gabor Jets

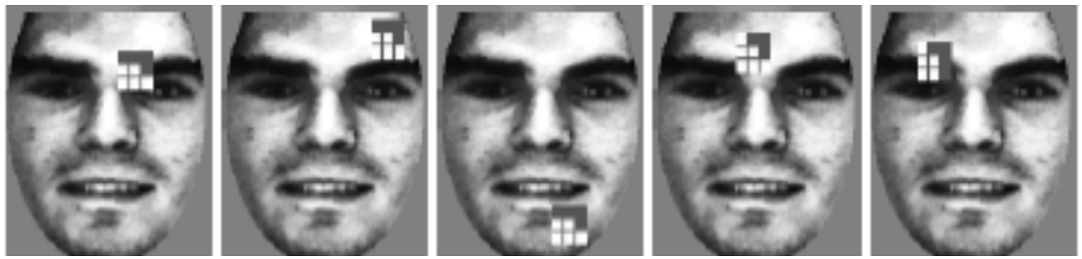




L. Wiskott, J-M Fellous, N. Krüger, and C.von der Malsburg "Face Recognition by Elastic Bunch Graph Matching". IEEE Transactions on PAMI 19(7):775-779, July 1997.

Local Binary Patterns (1)





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

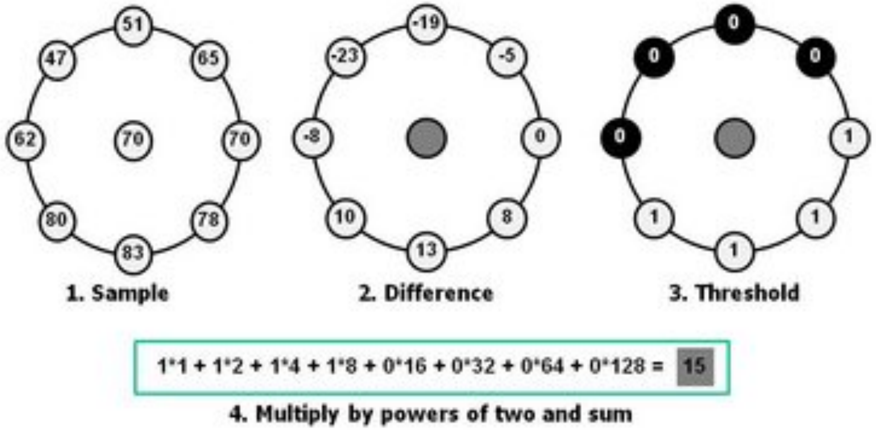
The histogram of the labels is used as a texture descriptor.

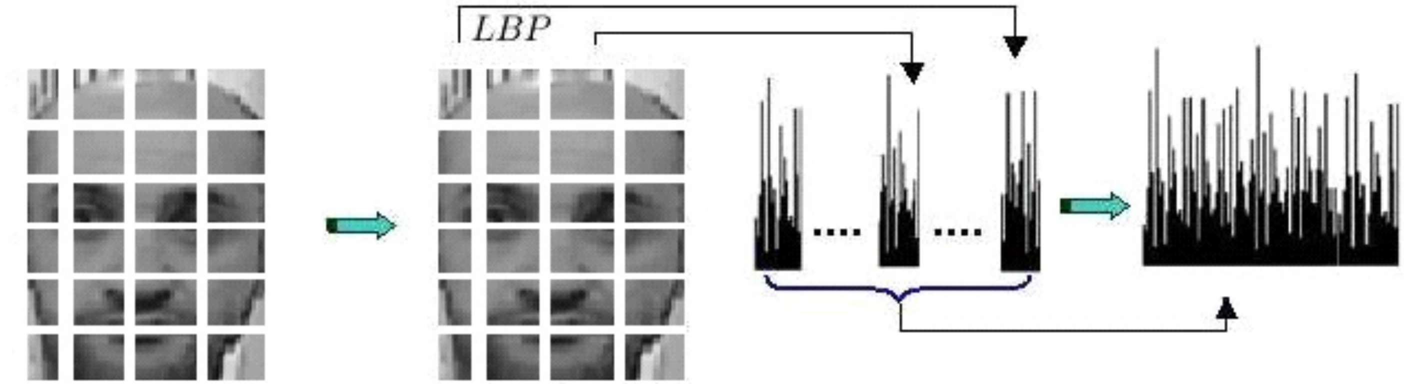
T. Ahonen et al. "Face Description with Local Binary Patterns: Application to Face Recognition" IEEE Transactions on PAMI 28(12):2037-2041, 2006.

Local Binary Patterns (2)

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

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



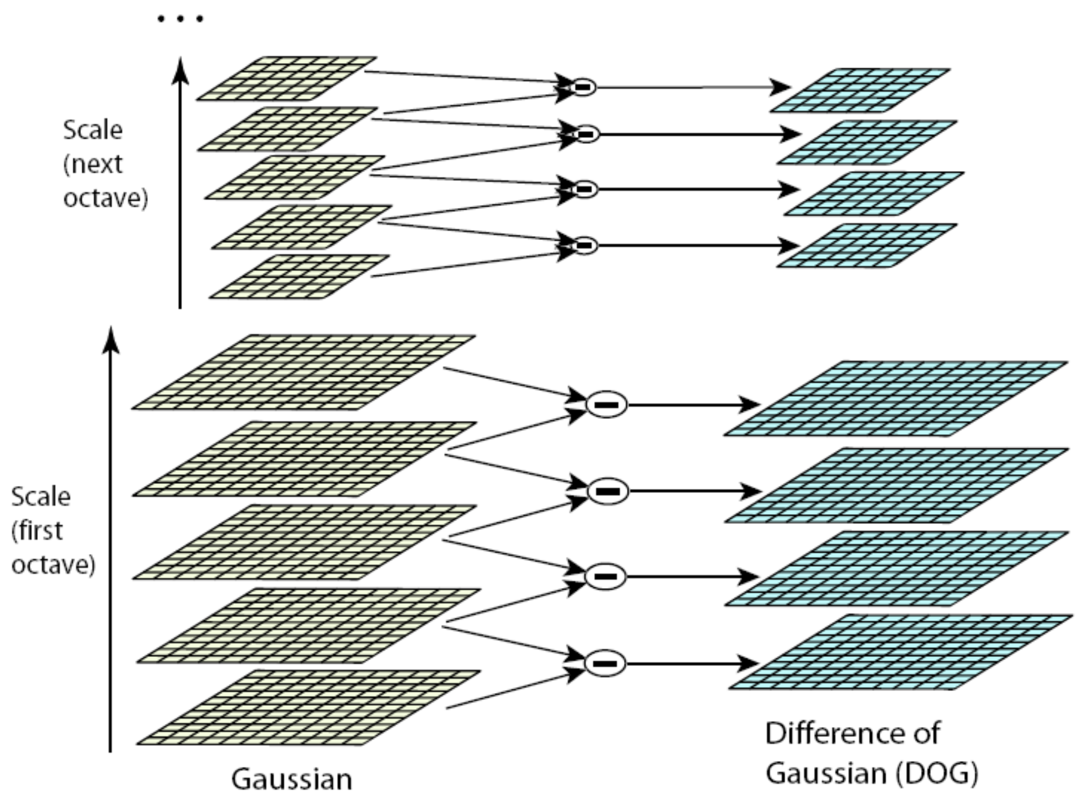


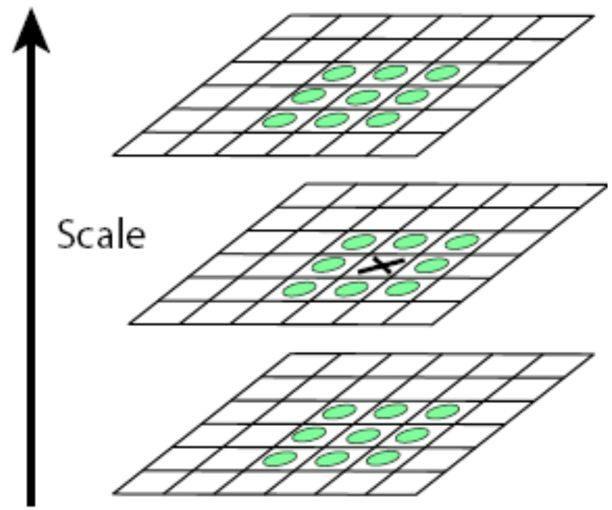
A face image (144x112 pixels) | The image is divided into 24 blocks of 24*28 pixels | LBP histogram from each block | Feature histogram

Scale Invariant Features (1)

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

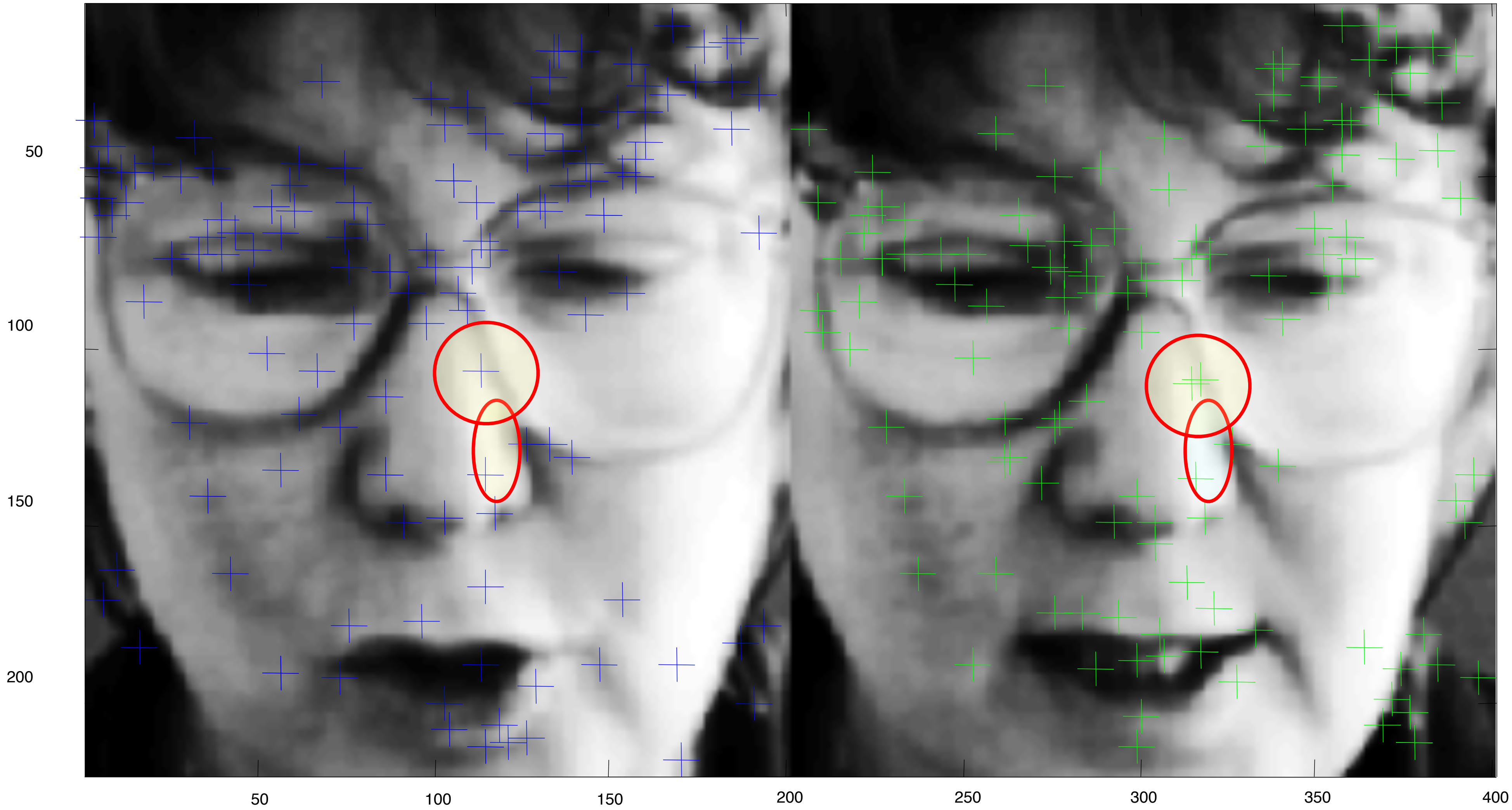
$$\mathbf{D}(x, y, \sigma, k) = \mathbf{L}(x, y, k\sigma) - \mathbf{L}(x, y, \sigma)$$





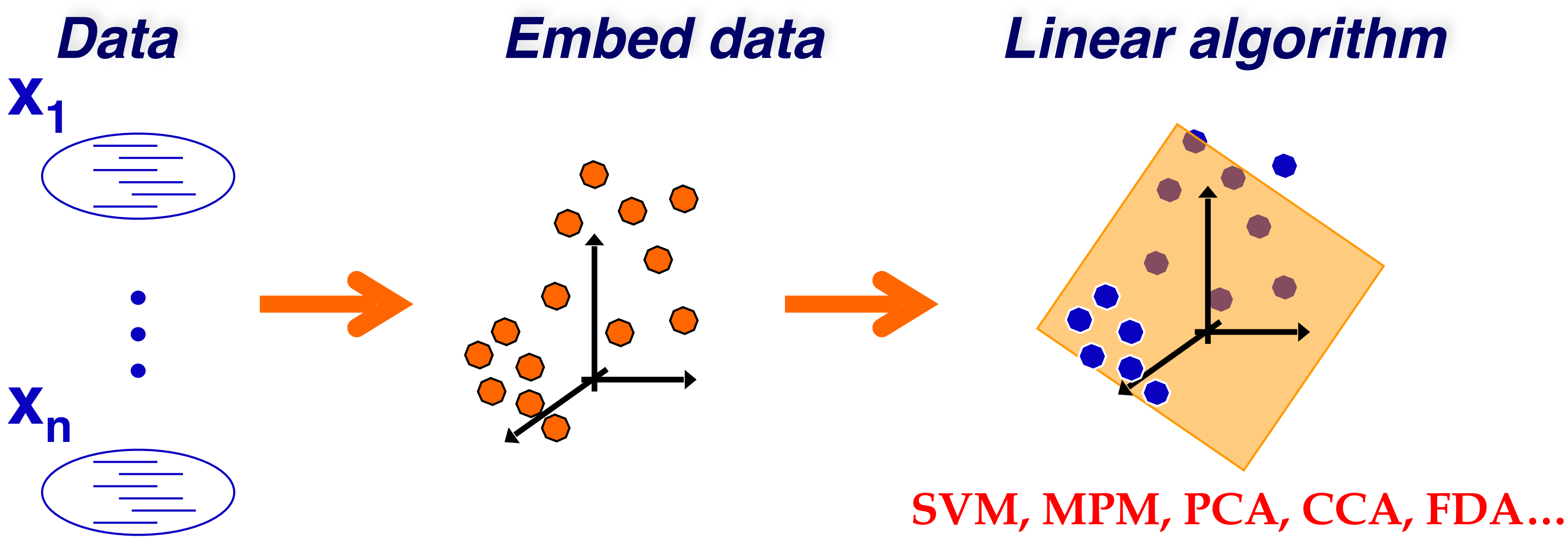
G. Lowe, "Object recognition from local scale invariant features", International Conference on Computer Vision , 1999.

Scale Invariant Features (2)



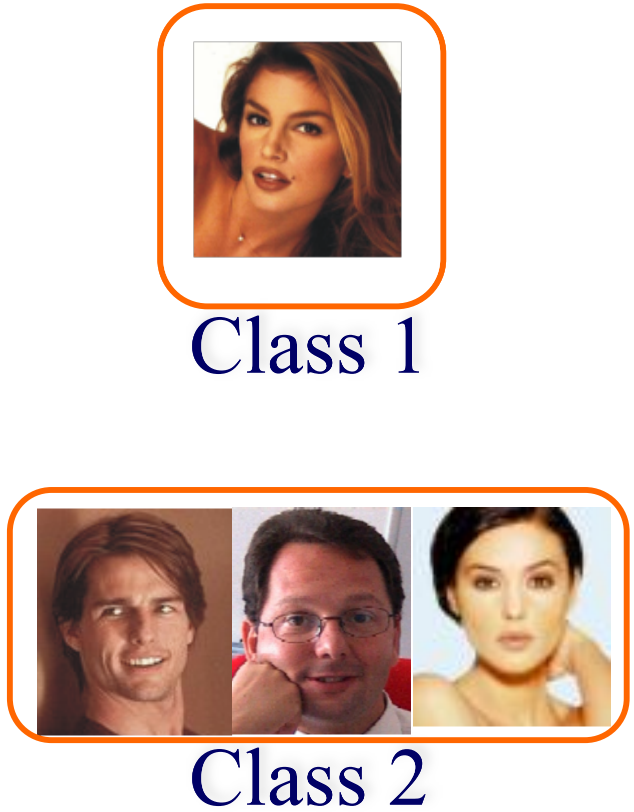
Kernel methods

- K-PCA; K-ICA; K-LDA… (B. Schölkopf et al. 1998)
- Are all variations of existing face-space representations. The transformation 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
- Related to the general *Learning Theory*



Vision Lab

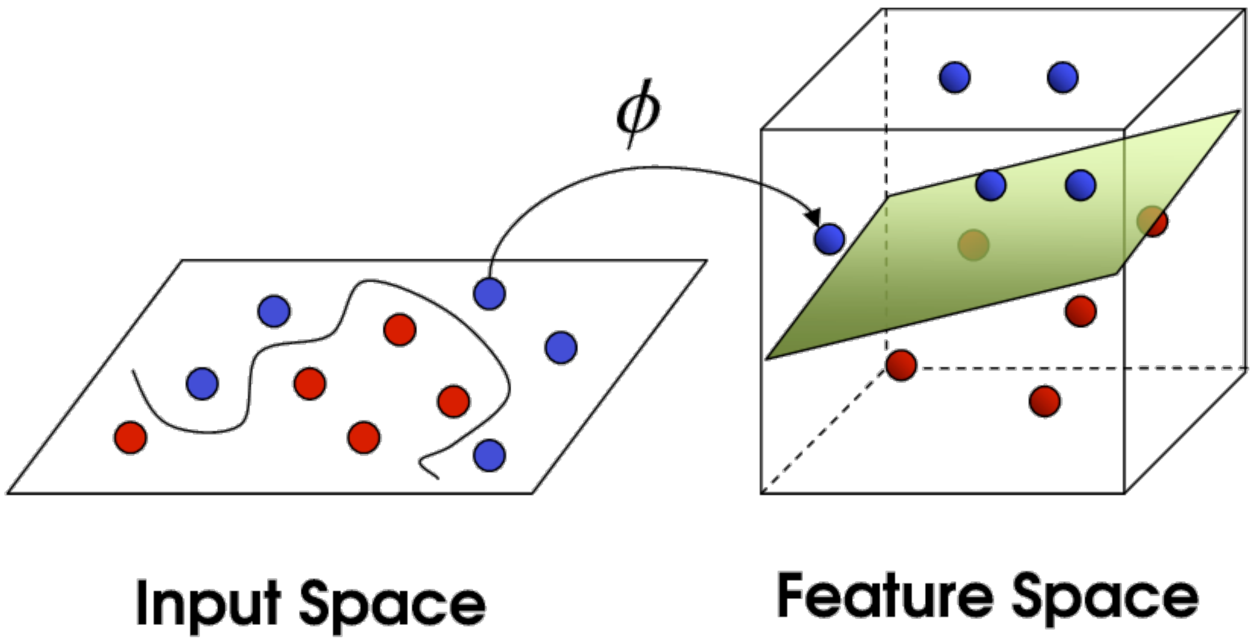
Support Vector Machines are binary classifiers



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

Support vectors (1)

- Solves linearly separable problems
  1. **Data projection**: Input data are transformed mapping into higher dimensions



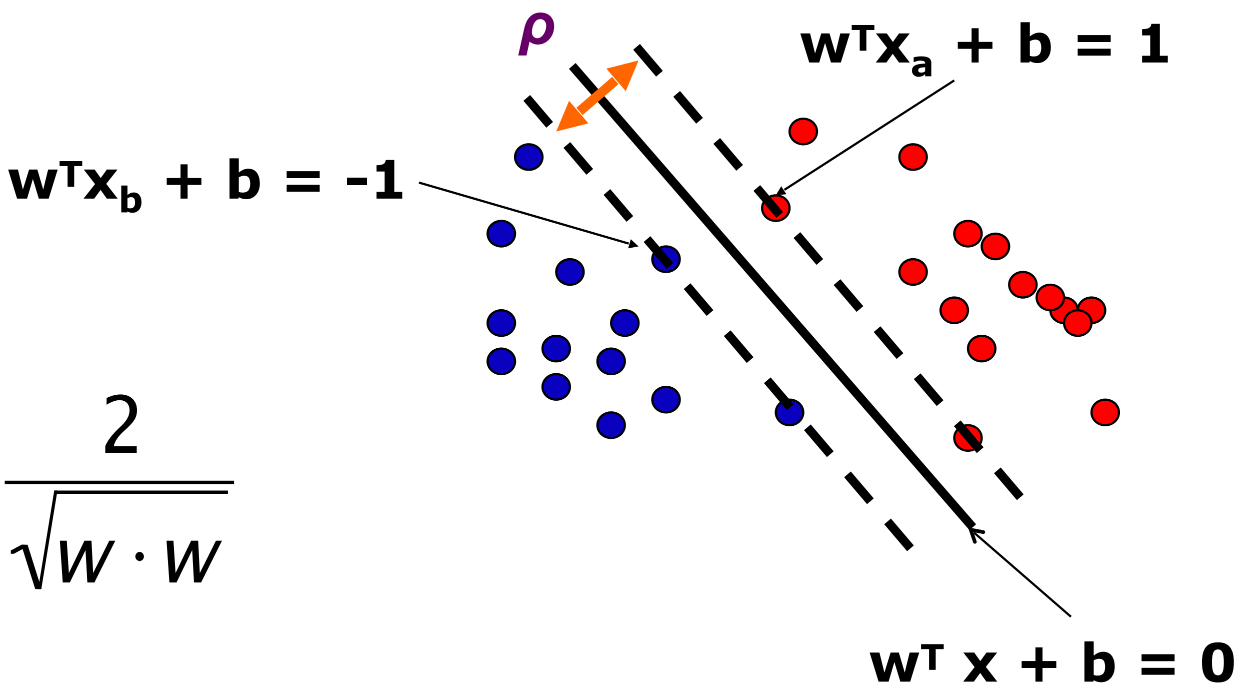
Vision Lab

Support vectors (2)

- Solves linearly separable problems

2. **Training**: find optimal hyperplane $\mathbf{w}^T\mathbf{x}_i + \mathbf{b} = 0$

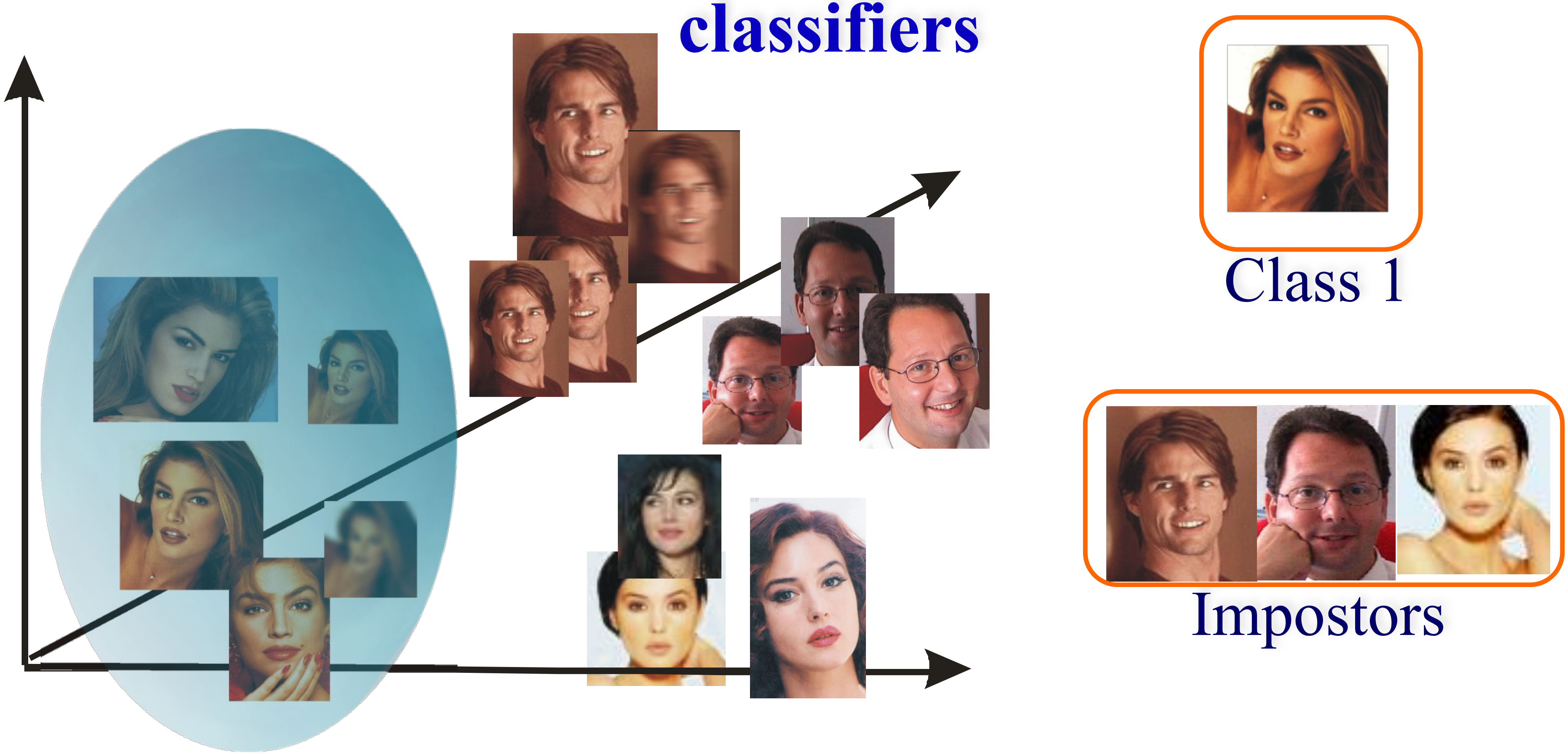
margin maximisation $\min_{i=1,\ldots,n} |\mathbf{w}^T\mathbf{x}_i + \mathbf{b}| = 1$



$$\rho(w) = \frac{2}{|w|} = \frac{2}{\sqrt{w \cdot w}}$$

One-Class Support Vector Machines (1)

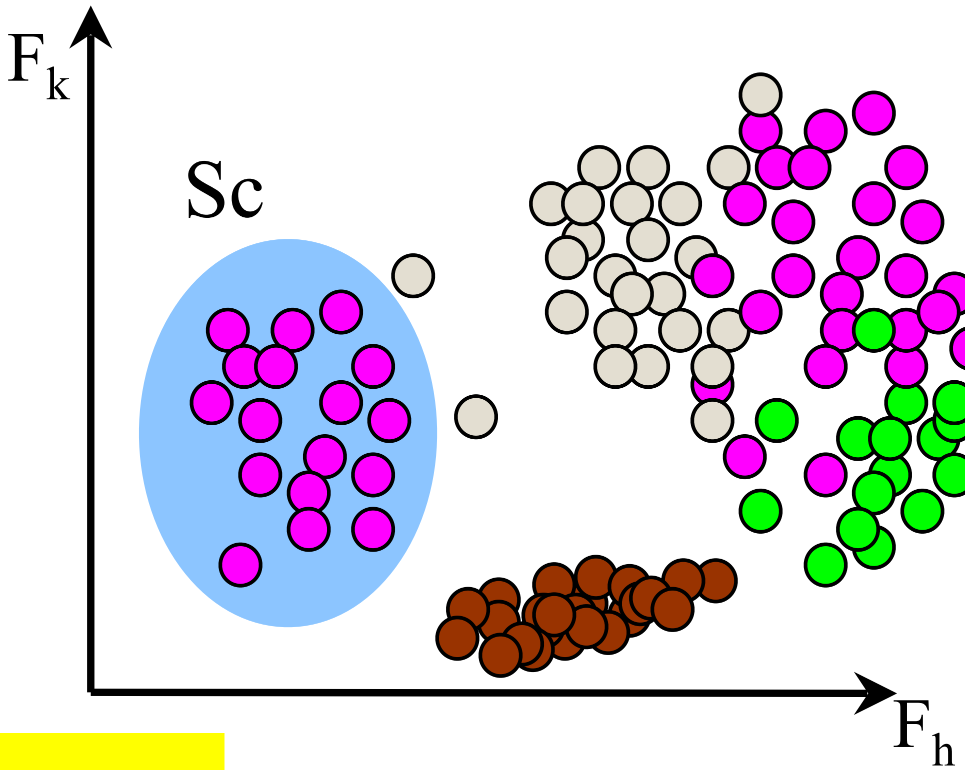
One-Class Support Vector Machines are <u>unary</u> classifiers



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 (2)

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



$$\|x_i - a\|^2 \leq R^2 \qquad i = 1..\ell$$

Parametric Morphable Models

3D shape

surface reflectance

Parametric Morphable Models

Statistically based analysis (PCA) of :

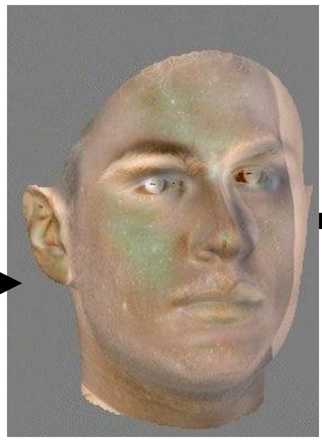
- laser scans - 3D shape (*x,y,z*) with reflectance (*r,g,b*) maps
- complete correspondence with average
- *Face = deformation from average* d*x*, d*y*, d*z*, d*r*, d*g*, d*b*

Input

Model Estimate

-

Difference δI



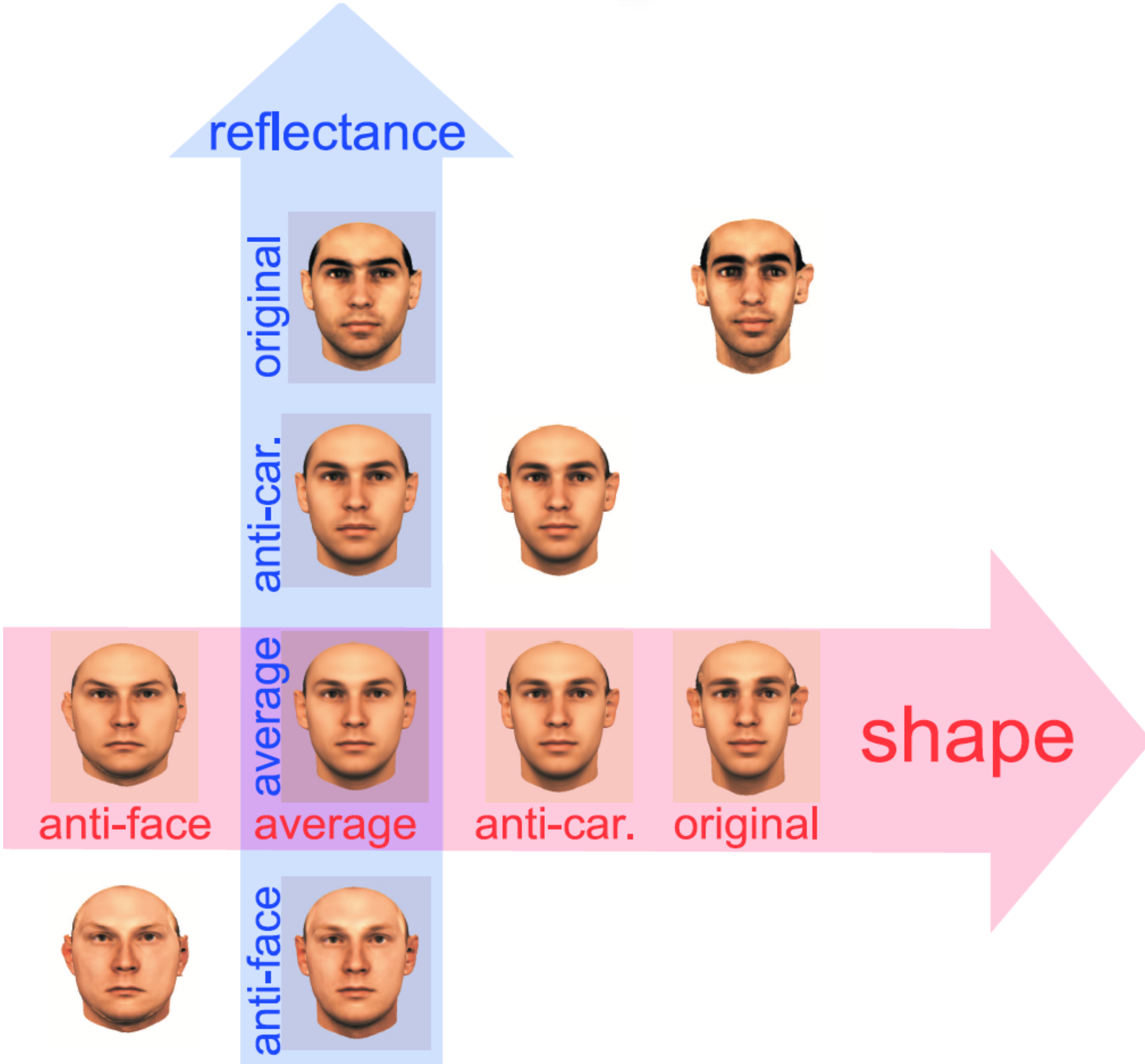
$$\delta\alpha = -\lambda \frac{\partial}{\partial\alpha}\left(\delta I^2\right)$$

$$\delta\beta = -\lambda \frac{\partial}{\partial\beta}\left(\delta I^2\right)$$

$$\delta\rho = -\lambda \frac{\partial}{\partial\rho}\left(\delta I^2\right)$$

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

Parametric Morphable Models



V. Blanz

Vision Lab

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.

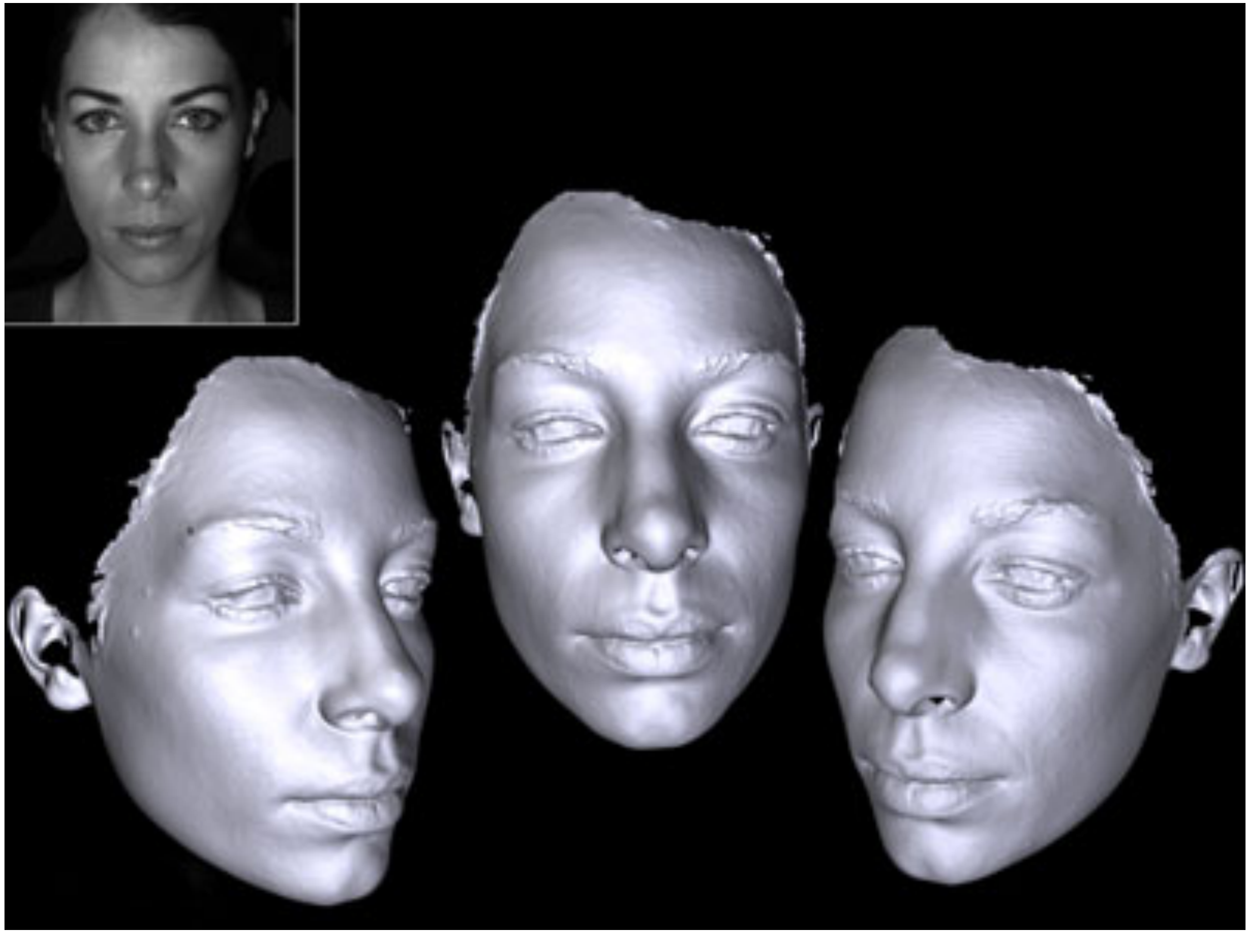
Cadoni M., Grosso E., Lagorio A., Tistarelli M., “Blending 2D and 3D Face Recognition”, T. Bourlai Ed. *Face Recognition Across the Imaging Spectrum*, pp. 305-331, Springer, 2016.

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.

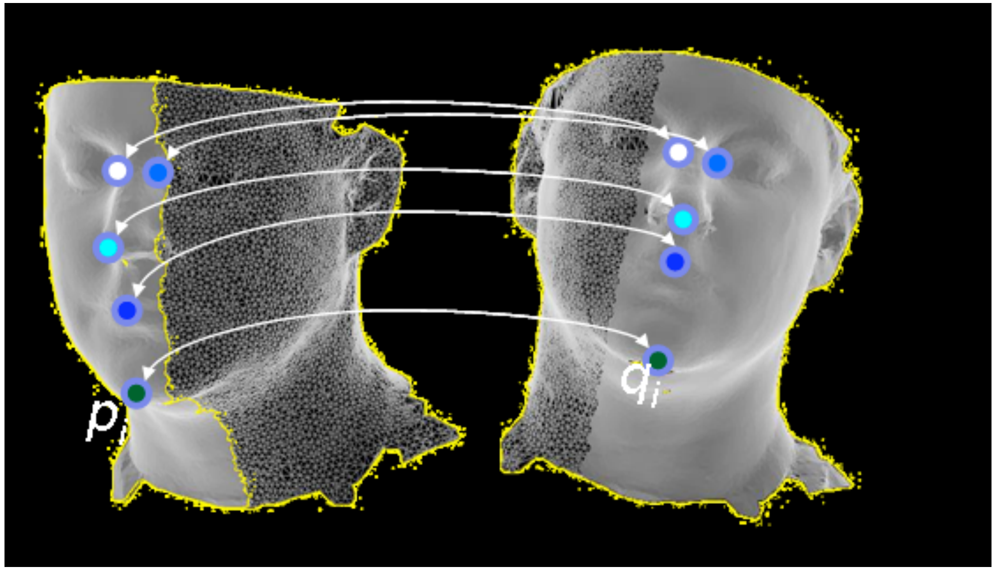
SACERENSIS • UNIVERSITATIS • TURRITANAE •

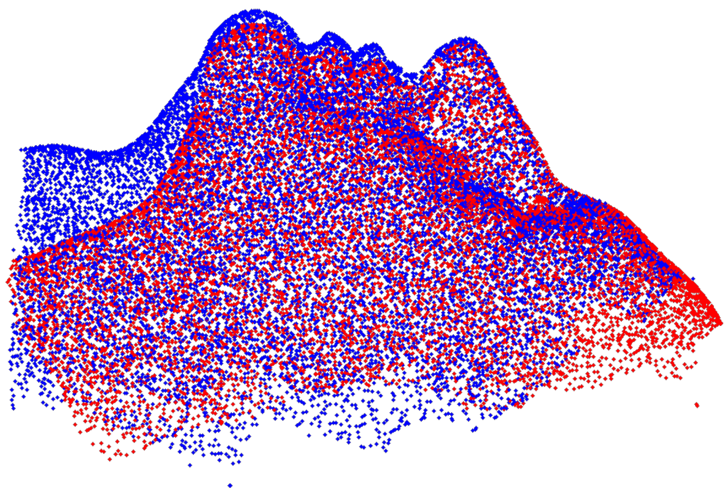
3D face recognition: from ill-posed to well-posed

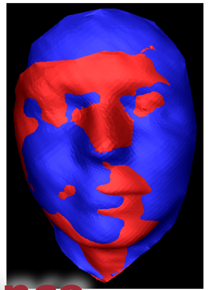


3D face matching

- 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

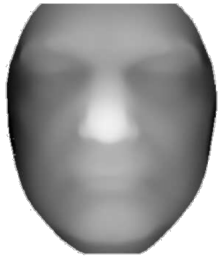
PC

Texture:
- Fusion with shape

TEX

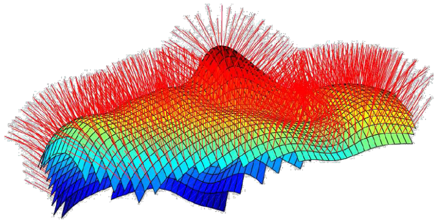
PCA

RI



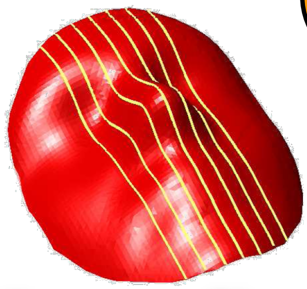
Range images:
- Benefit from 2D literature
- Easy to fuse with texture
- Applicable to 2.5D only

SN



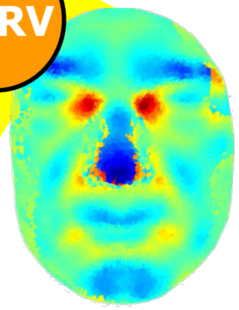
Surface normals:
- Too many …

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 invariants

- For each triplet (p_1, p_2, p_3) of feature points, a set of nine invariants is computed:

$$\left[\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3, \mathbf{J}_1, \mathbf{J}_2, \mathbf{J}_3, \tilde{\mathbf{J}}_1, \tilde{\mathbf{J}}_2, \tilde{\mathbf{J}}_3\right]$$

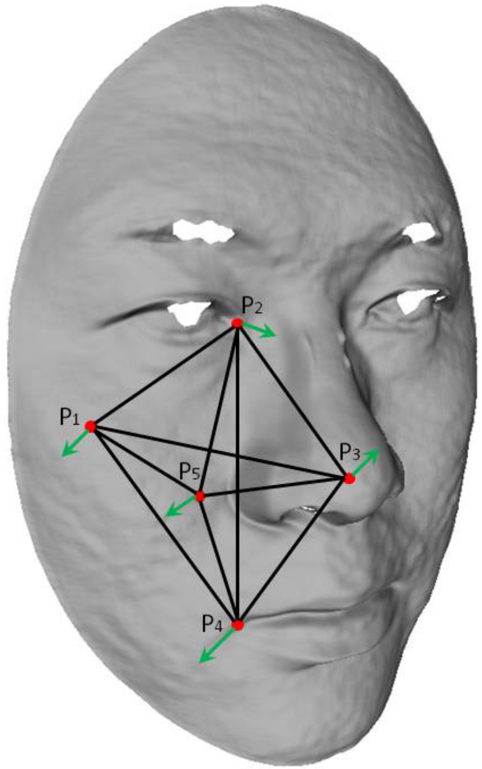
  - 3 of differential order zero:

$$\mathbf{I}_1 = \left\| p_2 - p_1 \right\|$$

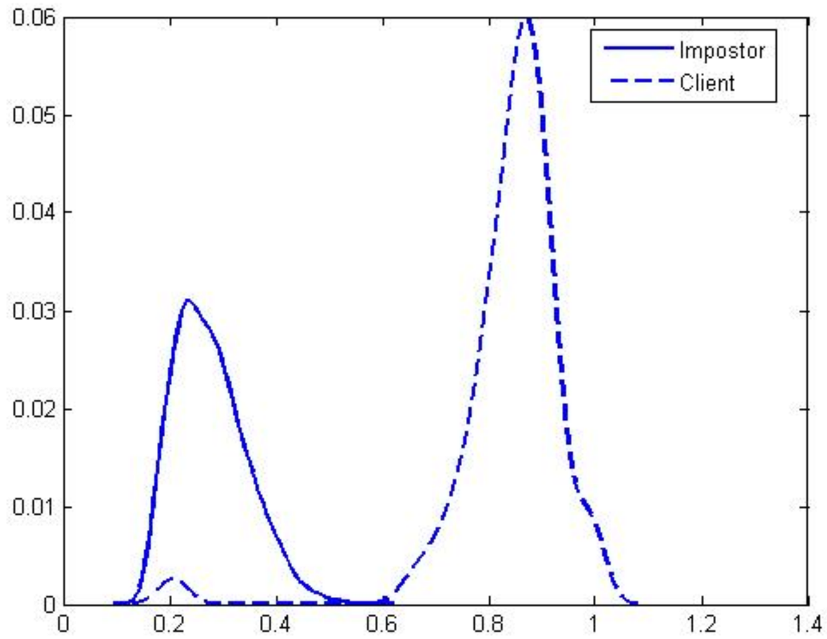
  - 6 of differential order one:

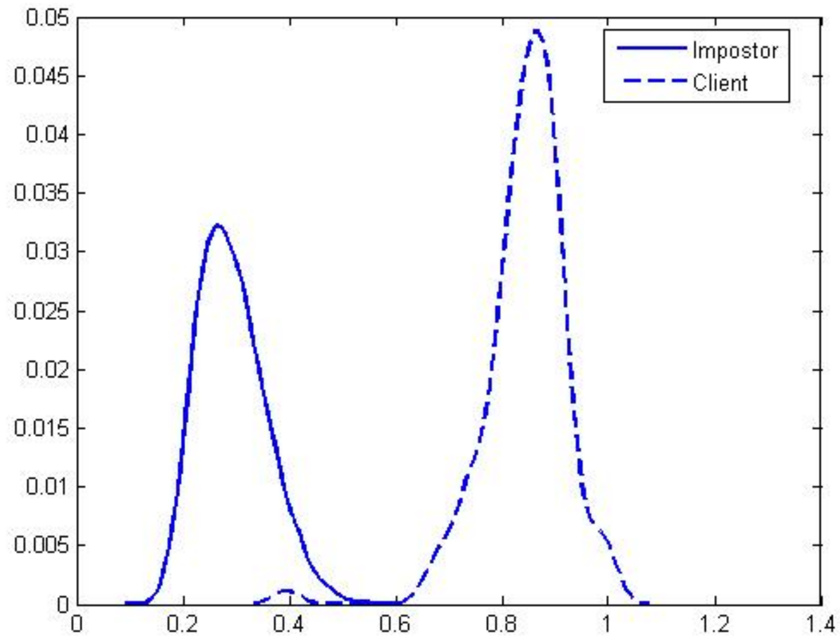
$$u = \frac{p_2 - p_1}{\left\| p_2 - p_1 \right\|} \qquad v_t = \frac{(p_2 - p_1) \wedge (p_3 - p_1)}{\left\| (p_2 - p_1) \wedge (p_3 - p_1) \right\|} \qquad \mathbf{J}_k = \frac{(v_t \wedge v) \cdot v_k}{v_t \cdot v_k}$$

$$\tilde{\mathbf{J}}_k = \frac{u \cdot v_k}{v_t \cdot v_k}$$



3D Recognition Results

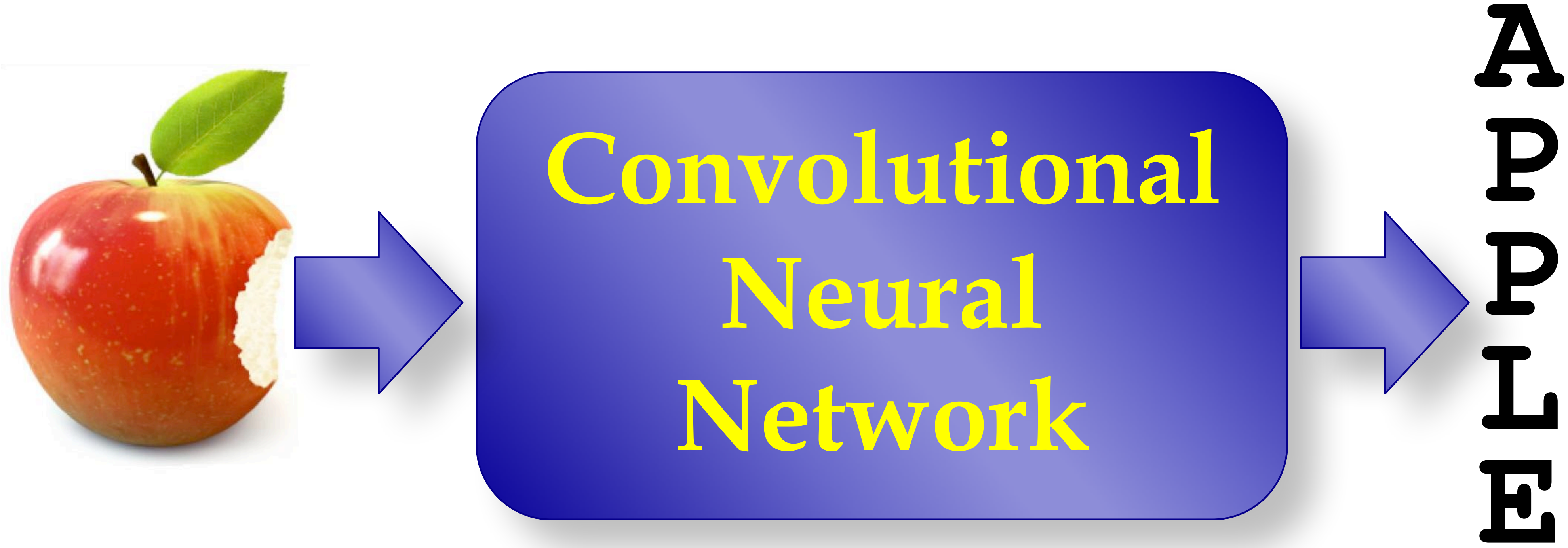




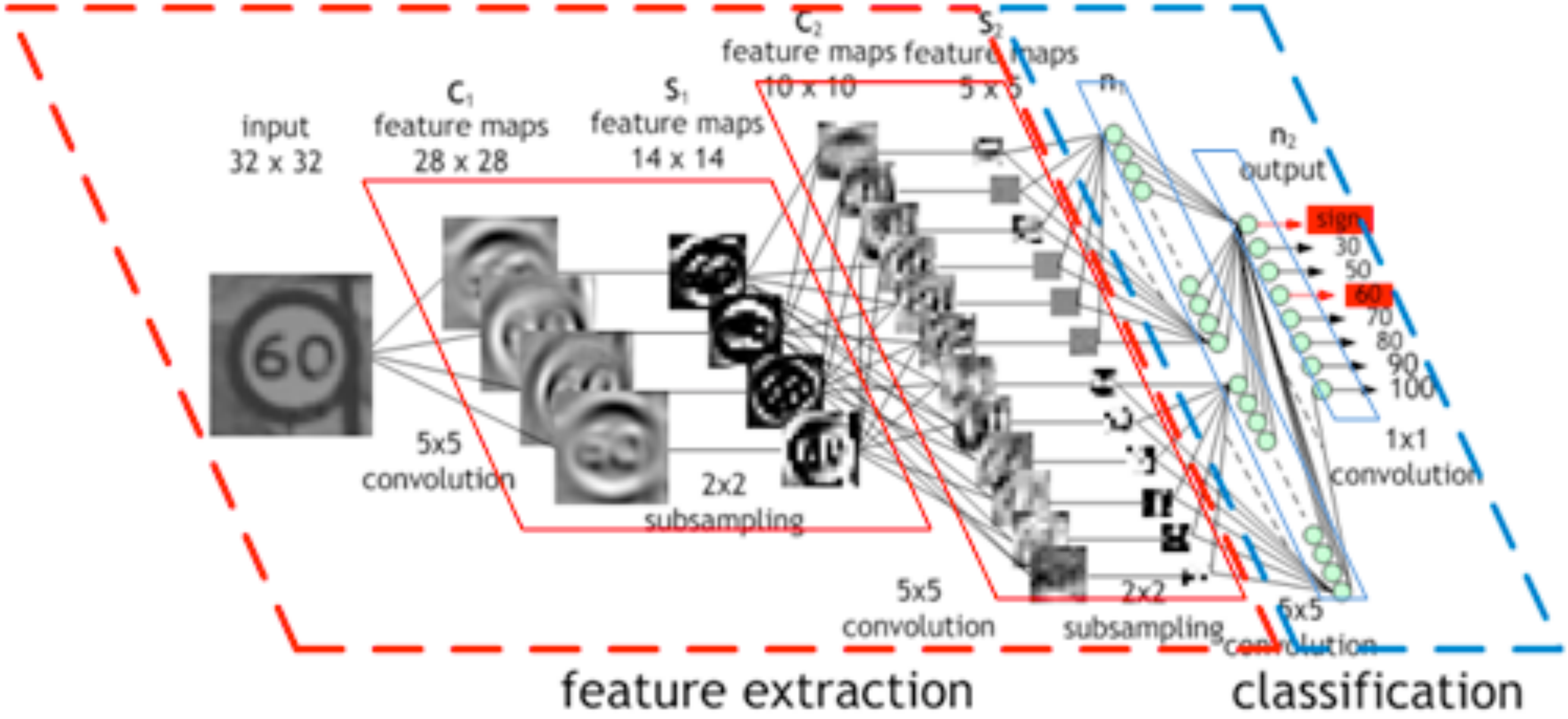
Impostor and client distributions for FRGC experiment **1** (left), and **3** (right)

- 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

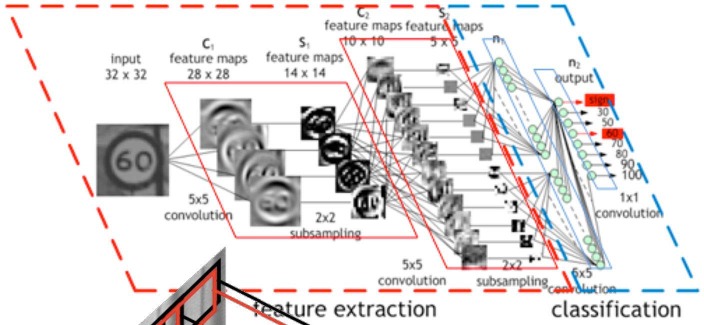
Convolutional Neural Networks

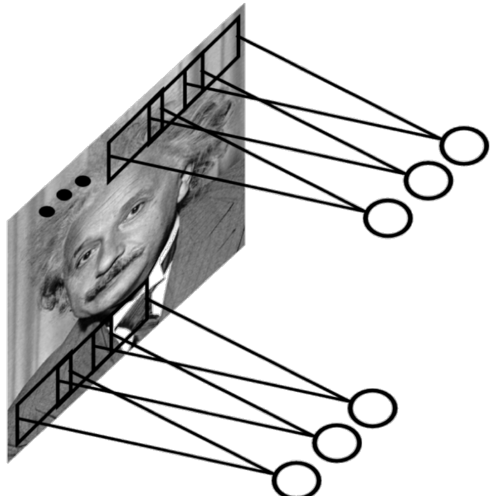


Convolutional Neural Networks

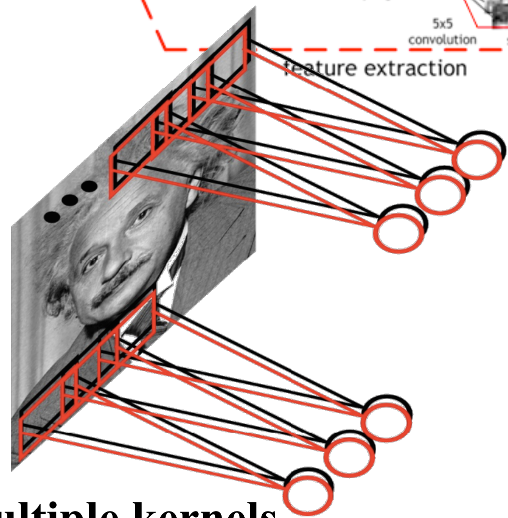


Convolutional Neural Networks

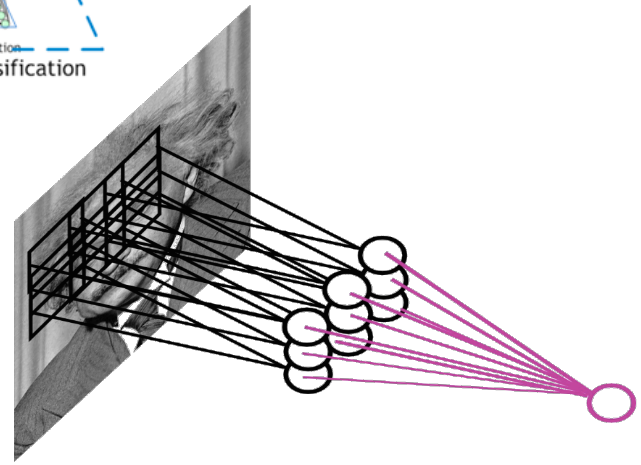




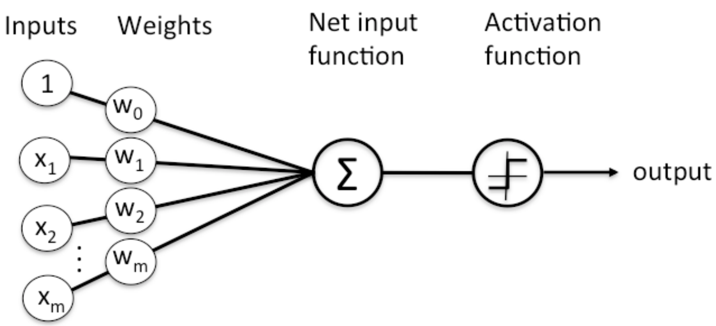
Single kernel Convolution



Multiple kernels Convolution



Spatial Pooling

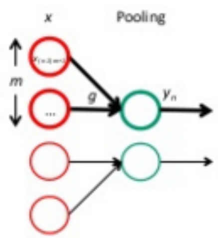


Let *m* be the size of pooling region, *x* be the input, and *y* be the output of the pooling layer. subsample(*f*, *g*)[*n*] denotes the *n*-th element of subsample(*f*, *g*).

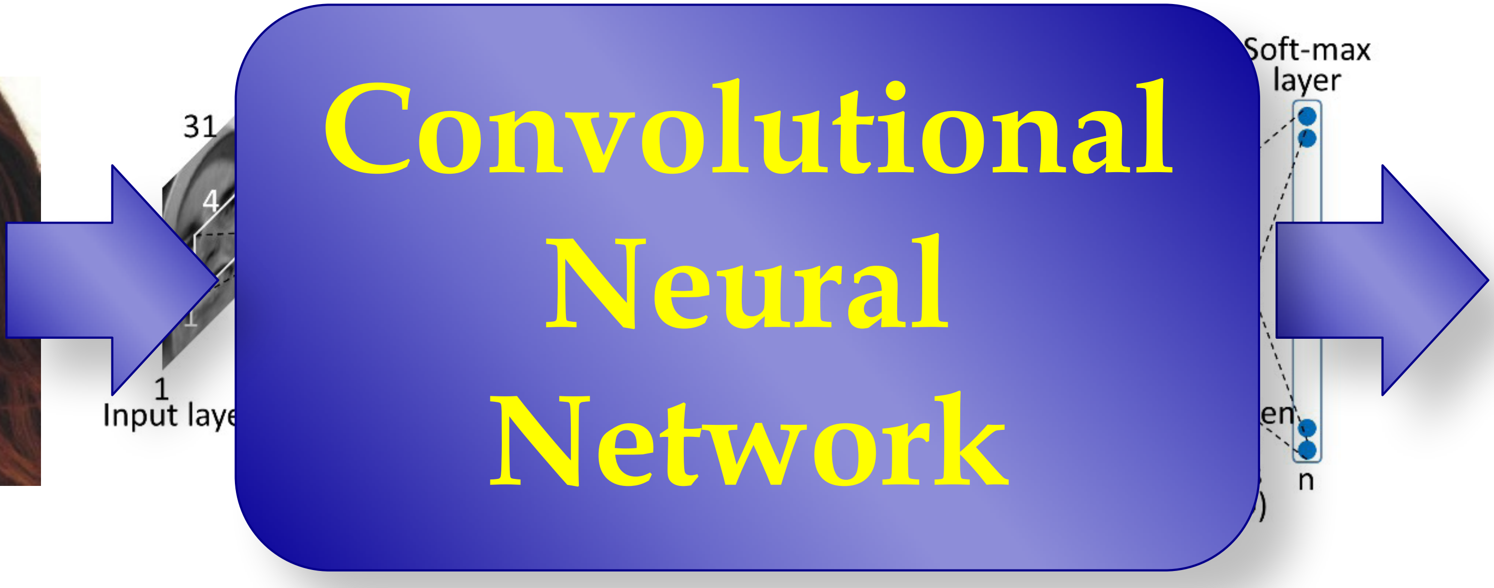
$$y_n = \mathrm{subsample}(x,g)[n] = g\left(x_{(n-1)m+1:nm}\right)$$

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

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

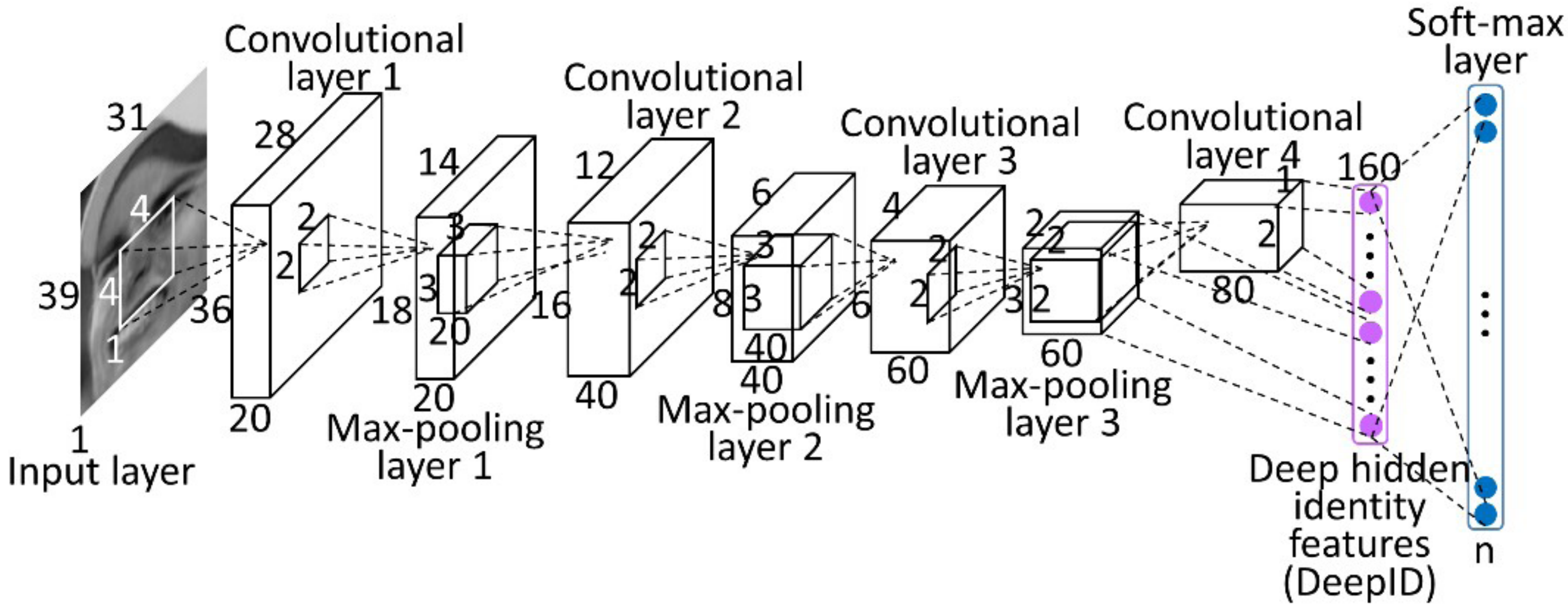


Convolutional Neural Networks



Cindy Crawford

Convolutional Neural Networks



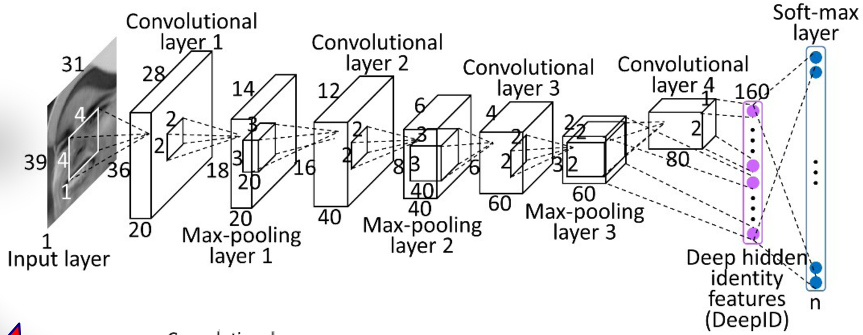
A deep CNN is used to extract a feature vector with relatively high dimension. The network can be supervised by multiclass loss and verification loss

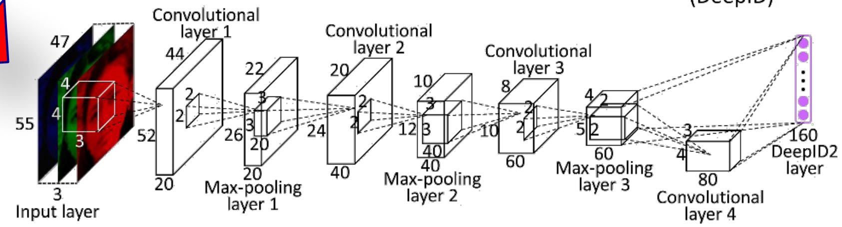
PCA, Joint Bayesian or metric-learning methods are used to learn a more efficient low dimensional representation

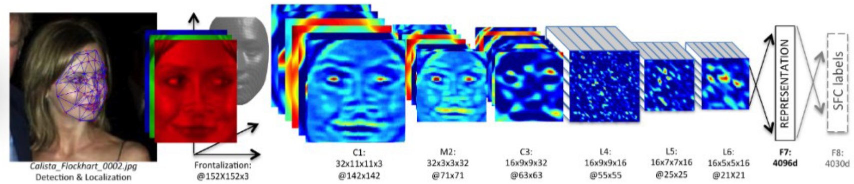
The amount of training data can range from **100K** up to **260M**

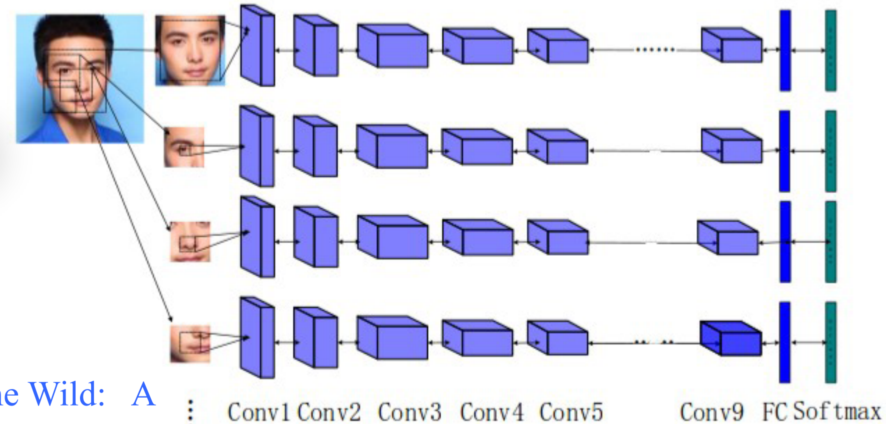
Convolutional Neural Networks

- 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?**



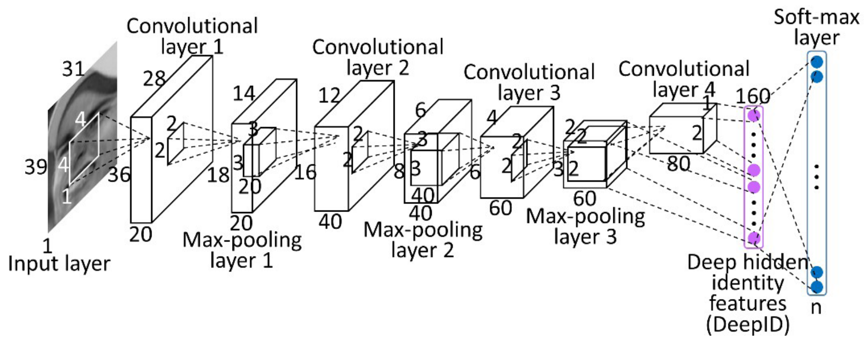






E. Learned-Miller, G. Huang, A. RoyChowdhury, H. Li, G. Hua, "Labeled Faces in the Wild: A Survey", Advances in Face Detection and Facial Image Analysis, pp 189-248, Springer 2016.

Convolutional Neural Networks

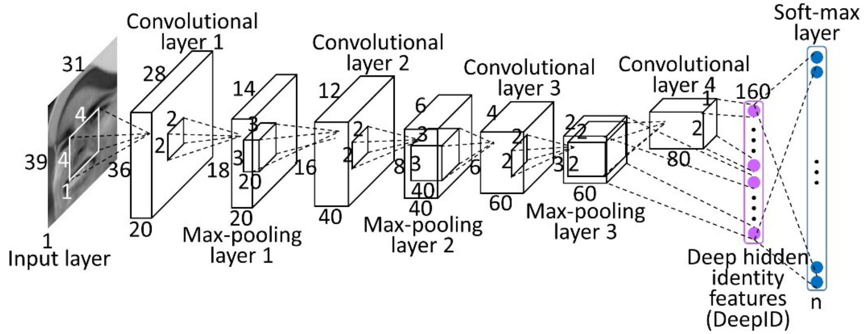


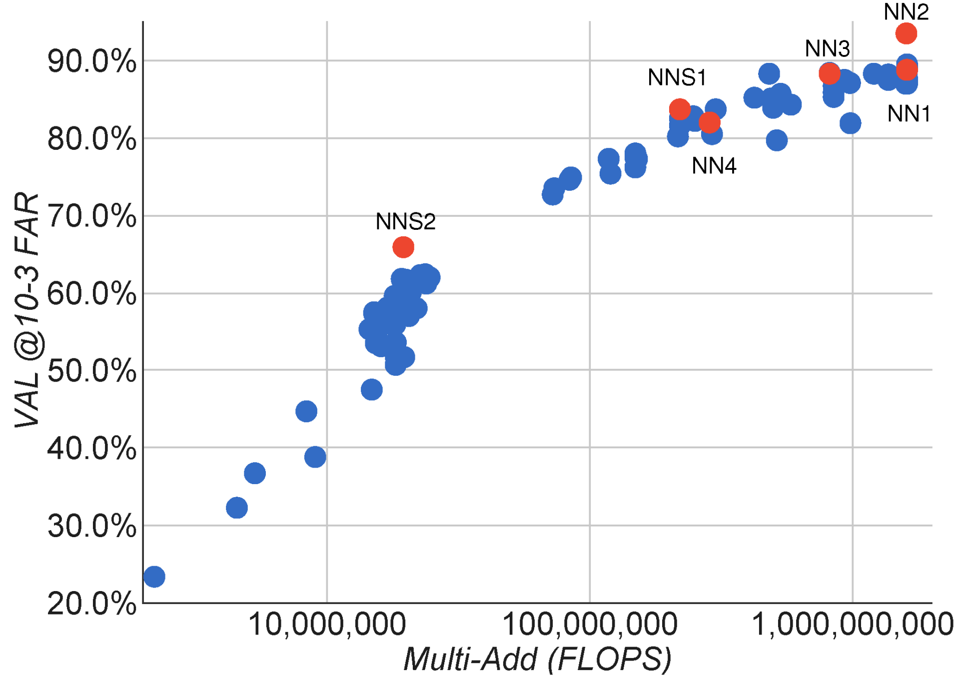
Method	Net. Loss	Outside data	# models	Aligned	Verif. metric	Layers	Accu.
DeepFace [97]	ident.	4M	4	3D	wt. chi-sq.	8	97.35±0.25
Canon. view CNN [115]	ident.	203K	60	2D	Jt. Bayes	7	96.45±0.25
DeepID [92]	ident.	203K	60	2D	Jt. Bayes	7	97.45±0.26
DeepID2 [88]	ident. + verif.	203K	25	2D	Jt. Bayes	7	99.15±0.13
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

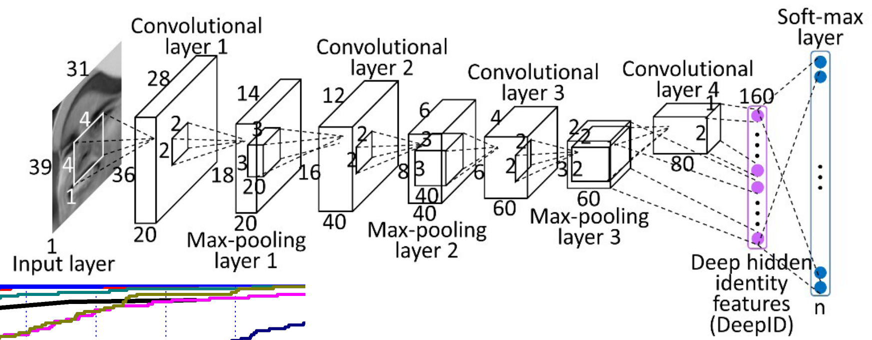


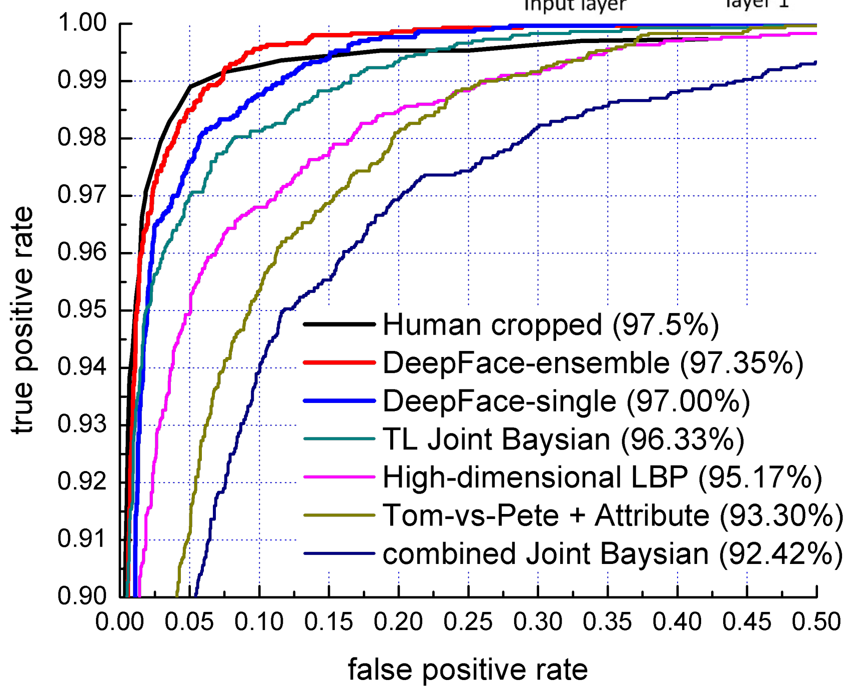


F. Schroff, D. Kalenichenko, J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering", CVPR 2015.

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

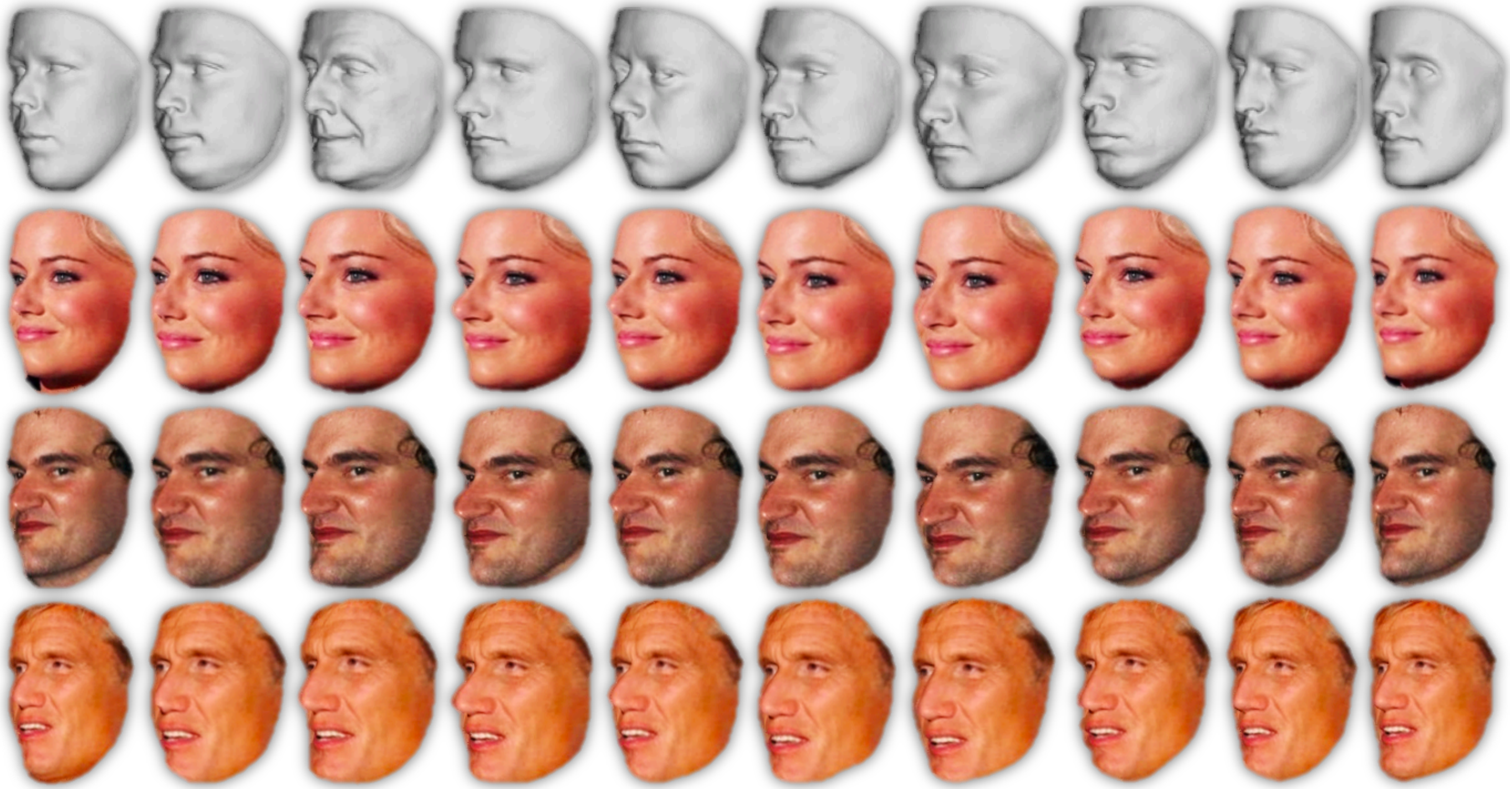




"The performance of these systems is ironically matched by our present ignorance of why they work as well as they do."

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

Efficient CNN learning



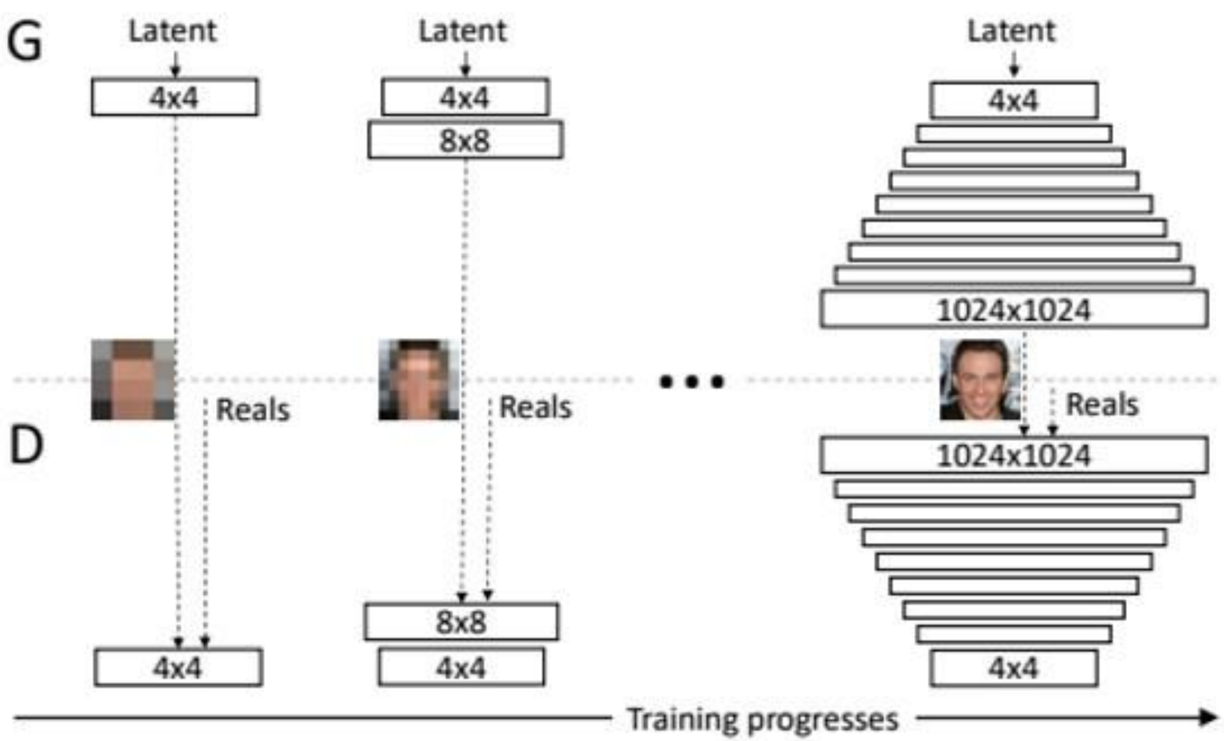
Augmenting faces by using different generic 3D models for rendering.

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.

Vision Lab

Generating Faces (1)



Generative Adversarial Networks: The key idea is to grow both the generator and discriminator progressively.
Starting from a low resolution, new layers are added to model increasingly fine details as training progresses. This both speeds the training up and greatly stabilizes it, allowing to produce images of unprecedented quality, e.g., CelebA images at 1024^2.

Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen; "Progressive Growing of GANs for Improved Quality, Stability, and Variation" arXiv preprint arXiv:1710.10196v2 [cs.NE], 3 Nov 2017.

Generating Faces (2)

Generative Adversarial Networks: The key idea is to grow both the generator and discriminator progressively.
Starting from a low resolution, new layers are added to model increasingly fine details as training progresses. This both speeds the training up and greatly stabilizes it, allowing to produce images of unprecedented quality, e.g., CelebA images at 1024^2.

Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen; "Progressive Growing of GANs for Improved Quality, Stability, and Variation" arXiv preprint arXiv:1710.10196v2 [cs.NE], 3 Nov 2017.

Face recognition

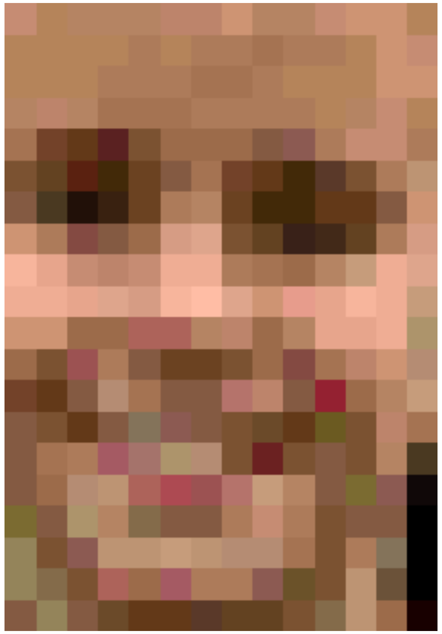
I. (PAST) What happened in 20+ years of research in face recognition?

II. (PRESENT) What can we learn?

III. (FUTURE) What is still to be done?

Human Face Perception

How many pixels to detect a face?

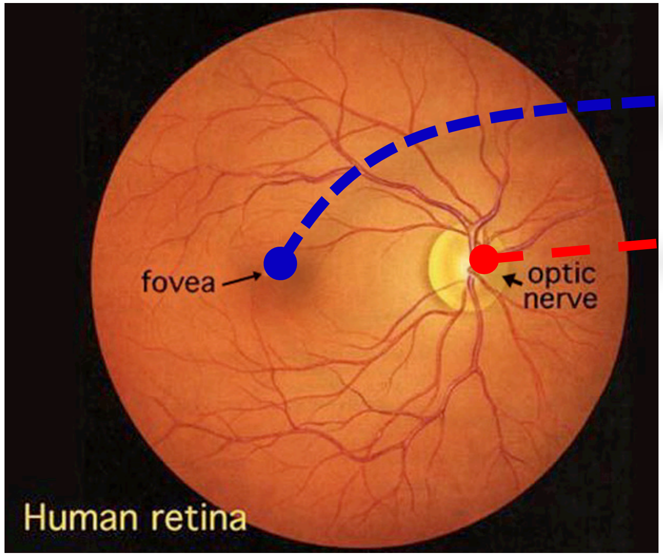


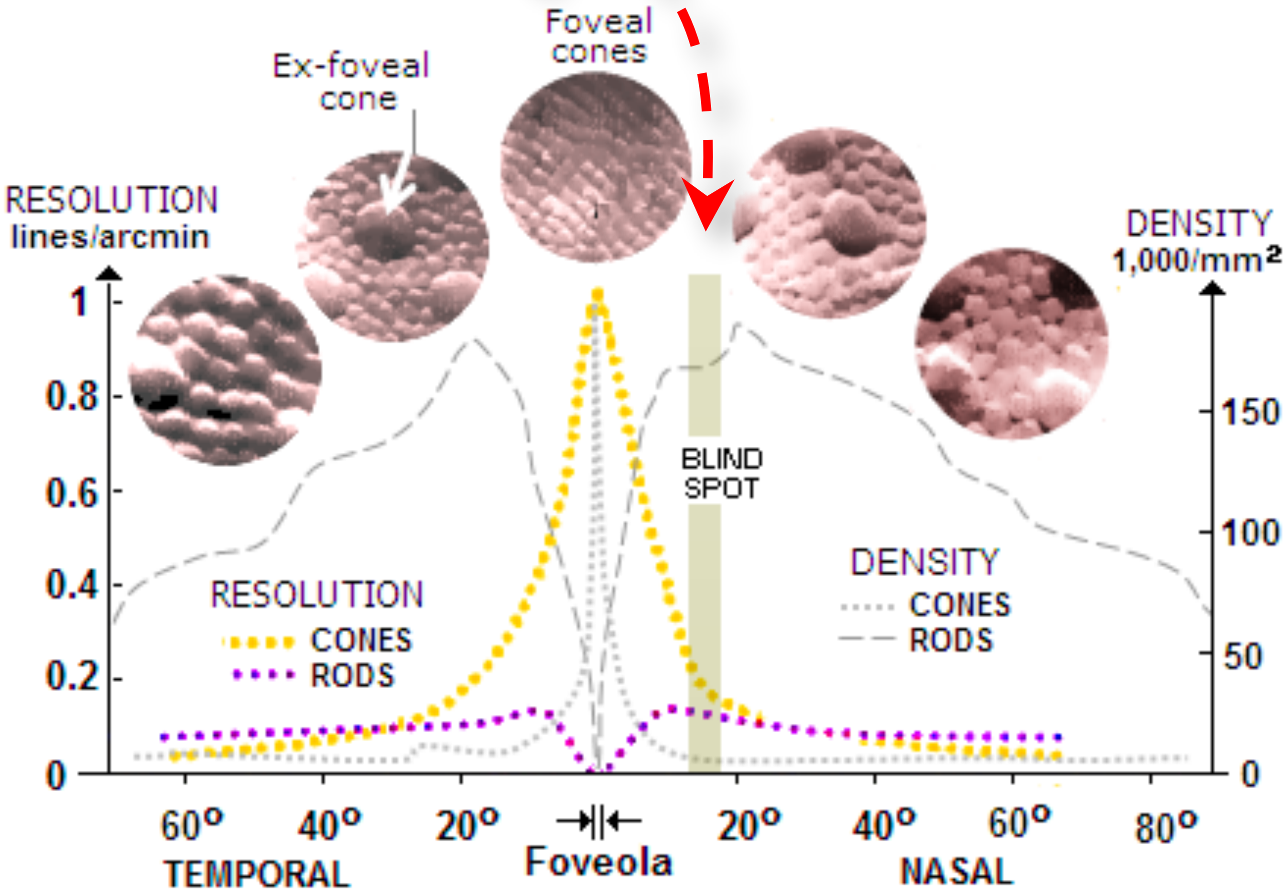
... Not many ... *(20x14)*

It's more a question of spatial distribution and …

proper frequency tuning

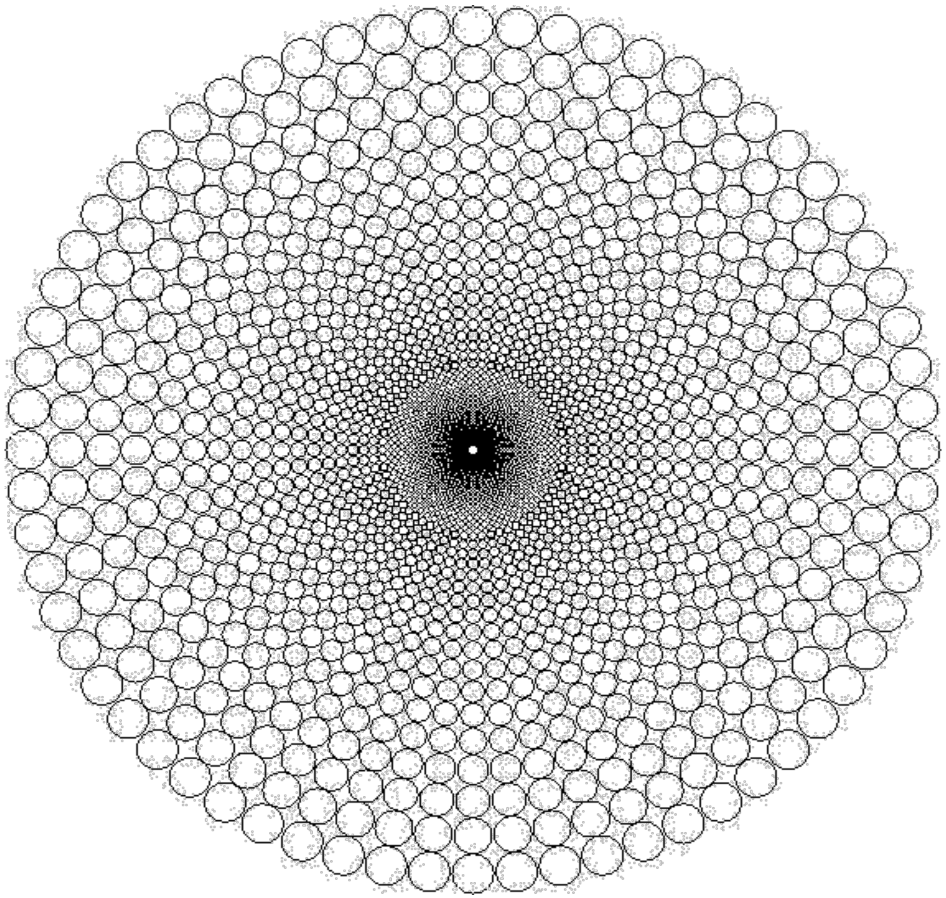
The retina layout (1)





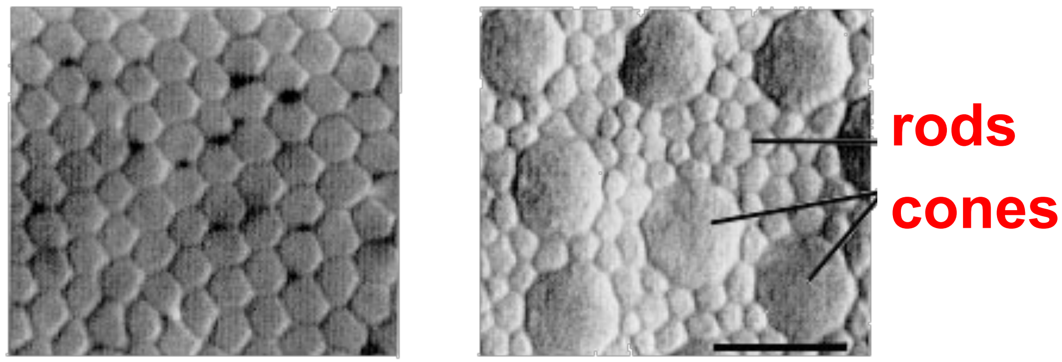
The retina layout (2)

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



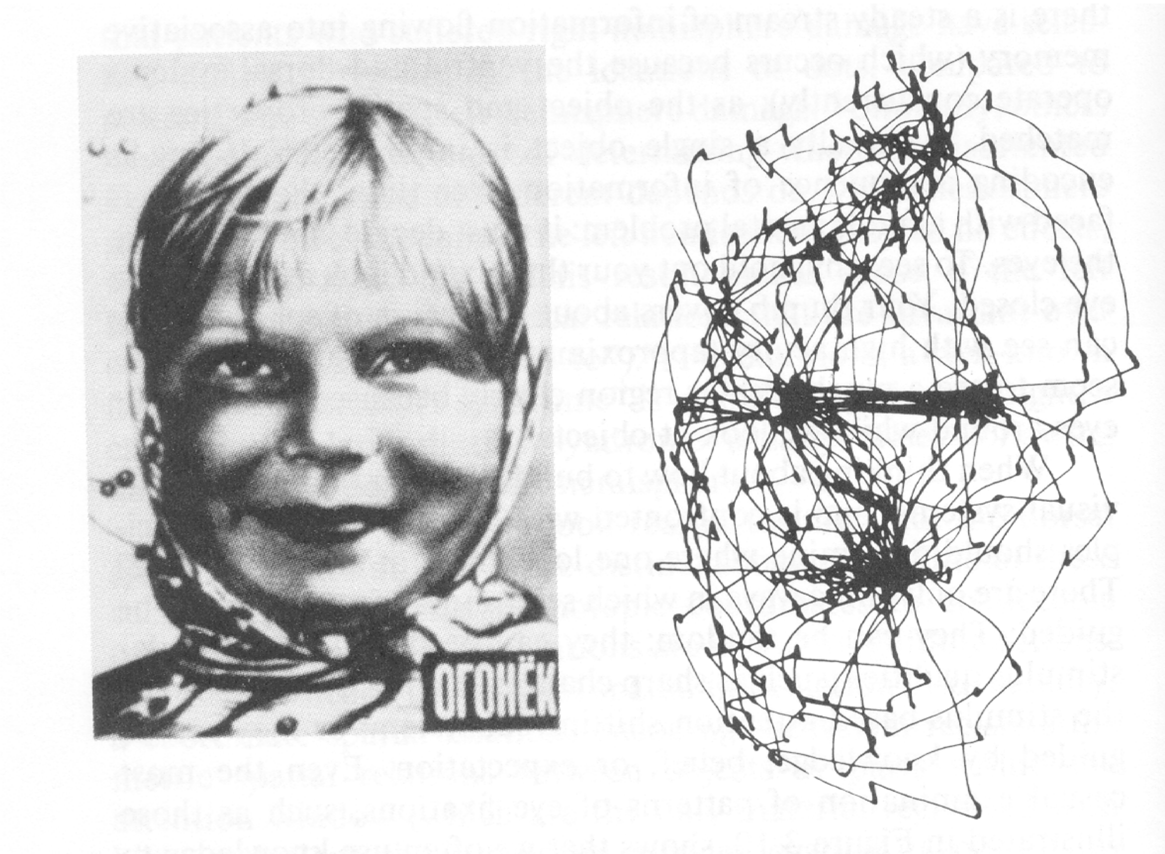
$$\begin{cases} x = \rho \sin\theta \\ y = \rho \cos\theta \end{cases} \qquad \begin{cases} \xi = \log_a\left(\rho / \rho_0 \right) \\ \eta = q\theta \end{cases}$$

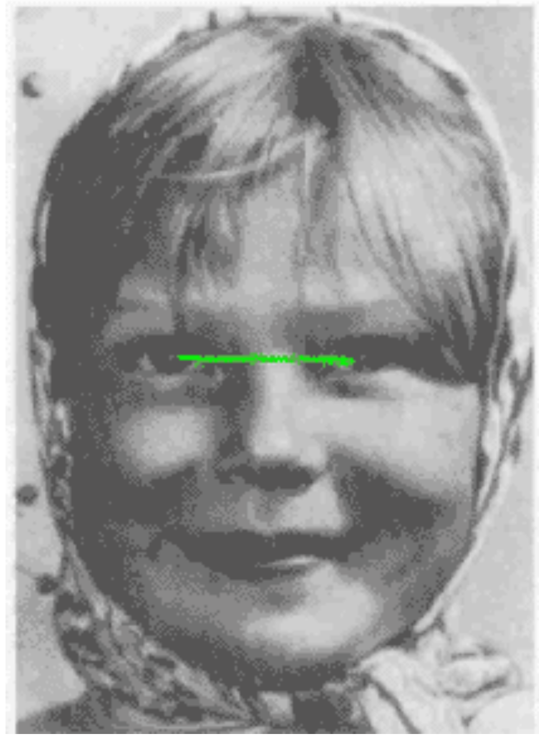
Fovea **Periphery**



Cones density and size in the fovea and near periphery

Context analysis: Visual attention

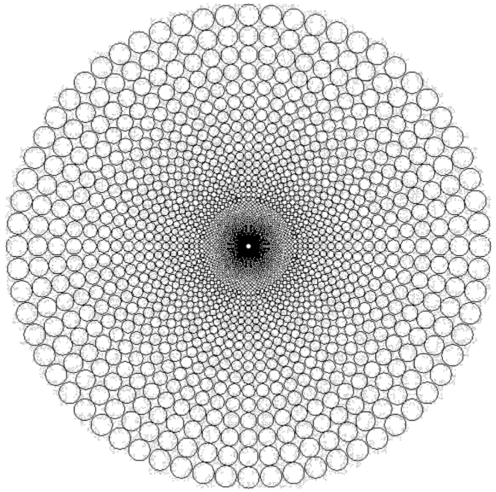




Eye movements while watching a girl’s face

A.L. Yarbus, “Eye Movements and Vision”, Plenum Press, 1967

Space variant imaging



Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space variant imaging

Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space variant imaging

Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space variant imaging

Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach” *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space variant imaging

Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach” *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

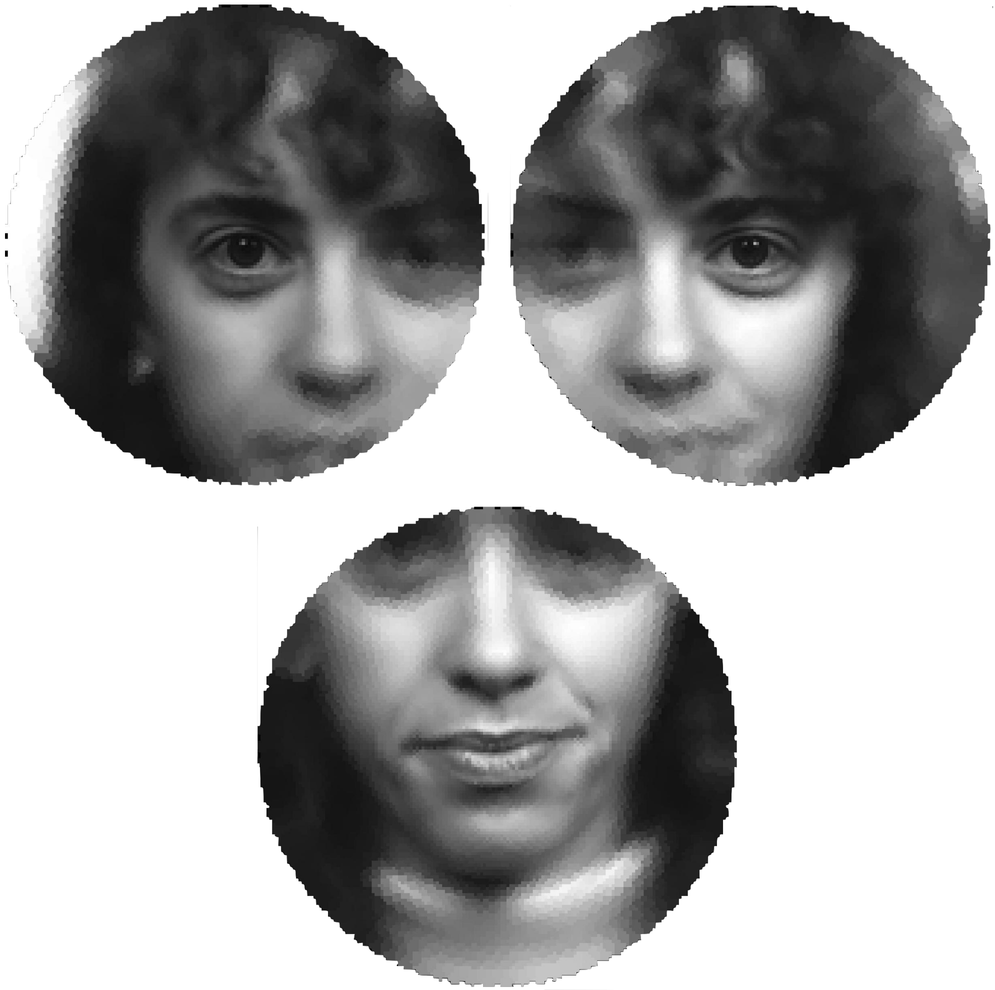
Space variant imaging

Tistarelli, M. and Grosso, E. (1997) "Active face r ... *ognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision ... *omputing*, Vol. 18, no. 4, pp 299-314, 2000

Space variant imaging

Tistarelli, M. and Grosso, E. (1997) "Active face r *ognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision *mputing*, Vol. 18, no. 4, pp 299-314, 2000

Space variant imaging



Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space variant imaging

Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space variant imaging

Vision Lab

Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

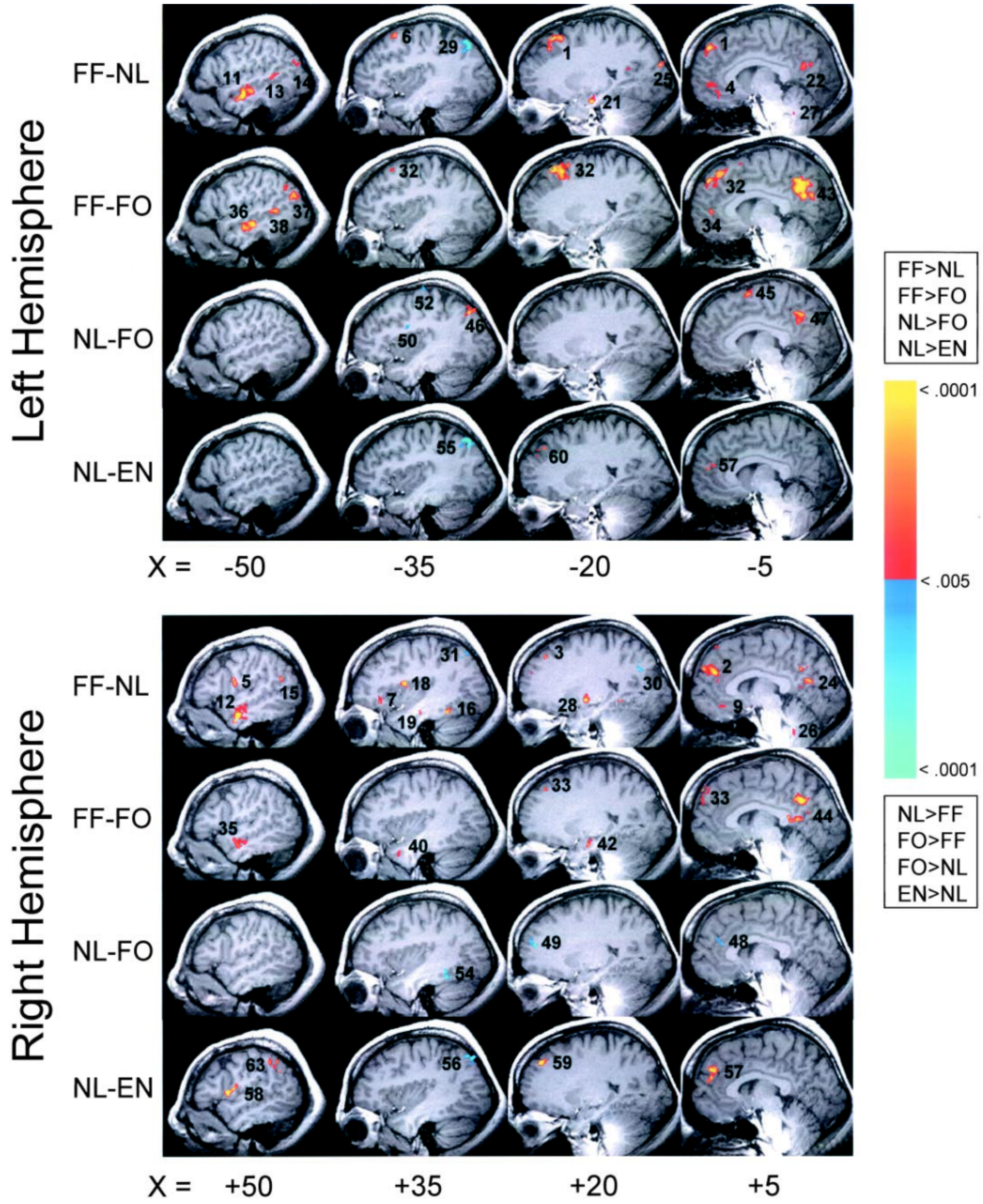
Space variant imaging

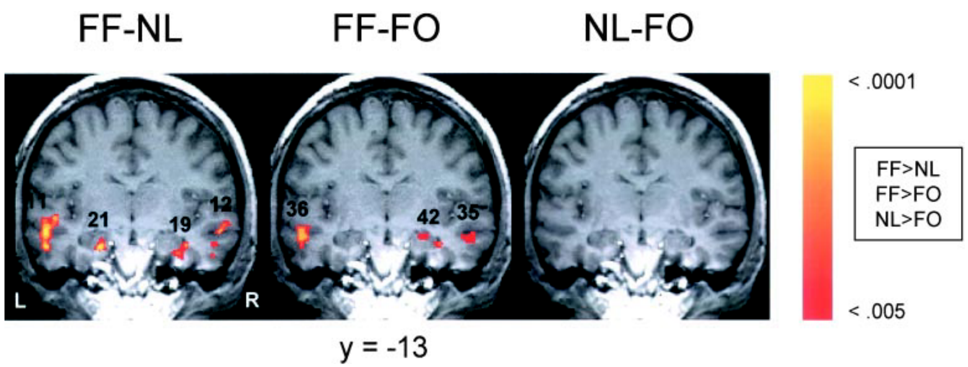
Tistarelli, M. and Grosso, E. (1997) "Active face recognition with an hybrid approach" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997
Tistarelli, M. and Grosso, E. (2000) "Active vision-based face authentication" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

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

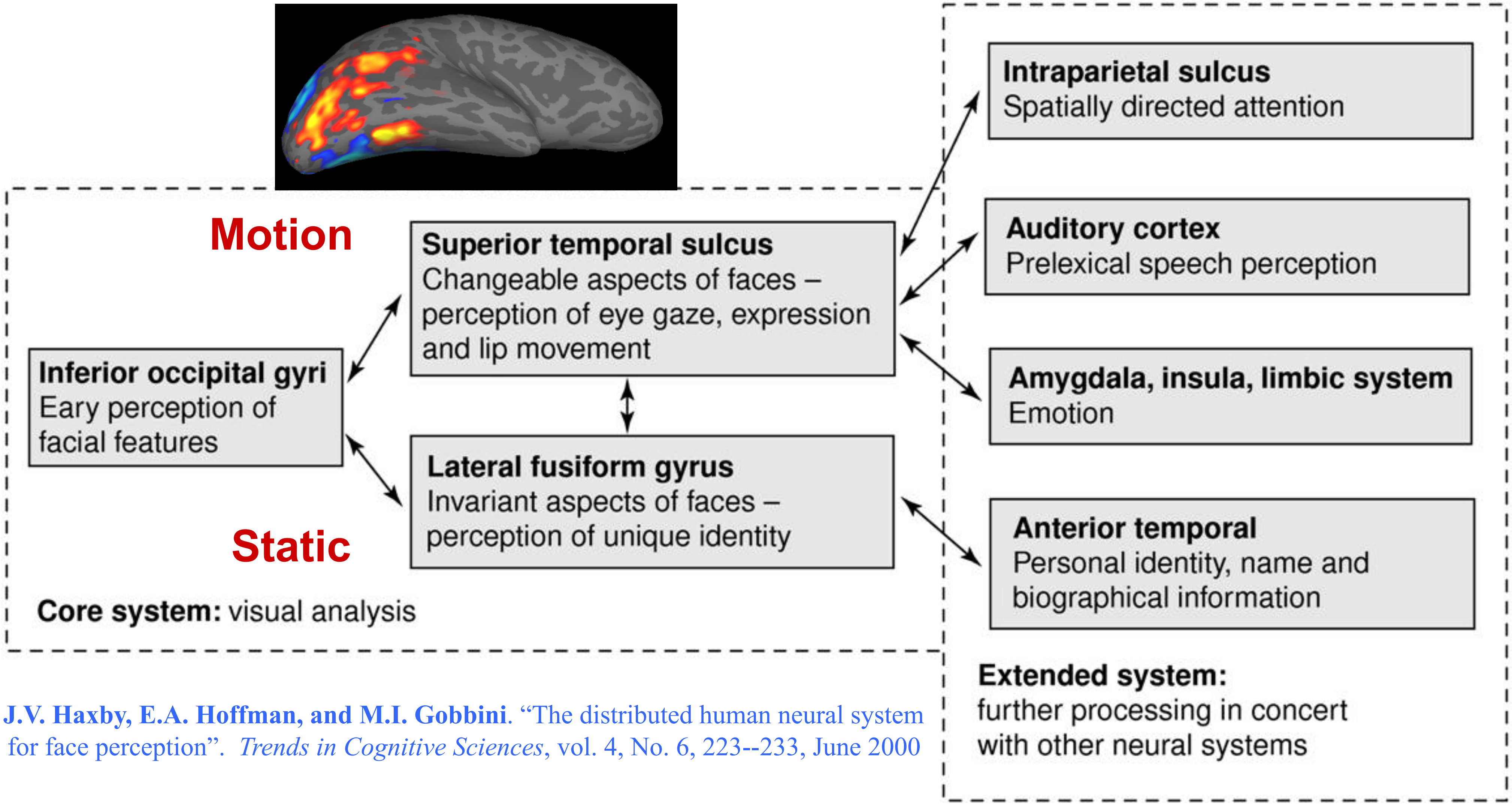
C. L. Leveroni et al. **"Neural Systems Underlying the Recognition of Familiar and Newly Learned Faces",** The Journal of Neuroscience, January 15, 2000, *20*(2):878–886

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

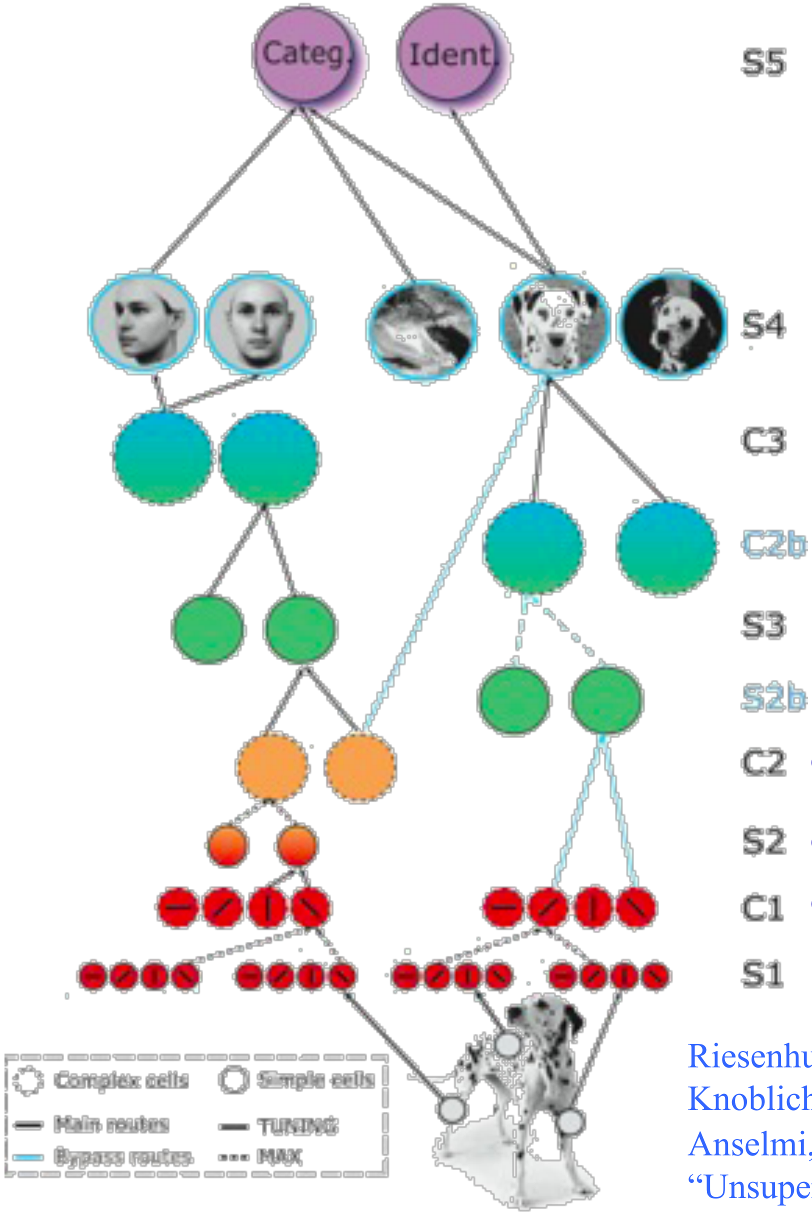
Region is defined as center of mass. The first column refers to location numbers demarcated in Figures 2 and 3 (italicized numbers indicate locations not shown in figures). Coordinates represent distance in millimeters from anterior commissure: x right (+)/left (−); y anterior (+)/posterior(−); z superior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

Neural architecture of face perception



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

Brain models

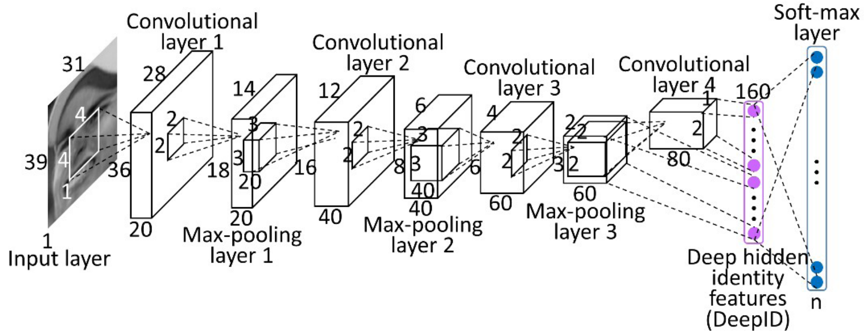


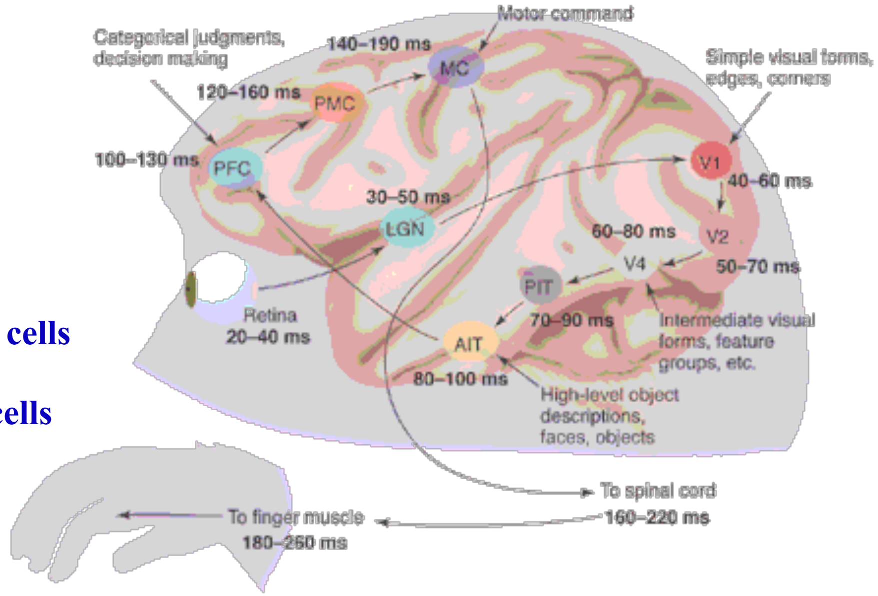
C2 Complex composite cells

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.

Face representation in the HVS

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			
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	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(−).

BRAIN Total 1400 ml
100 billion neurons
71.5 Mneurons/ml

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

Face representation in the HVS

aces (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
it 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

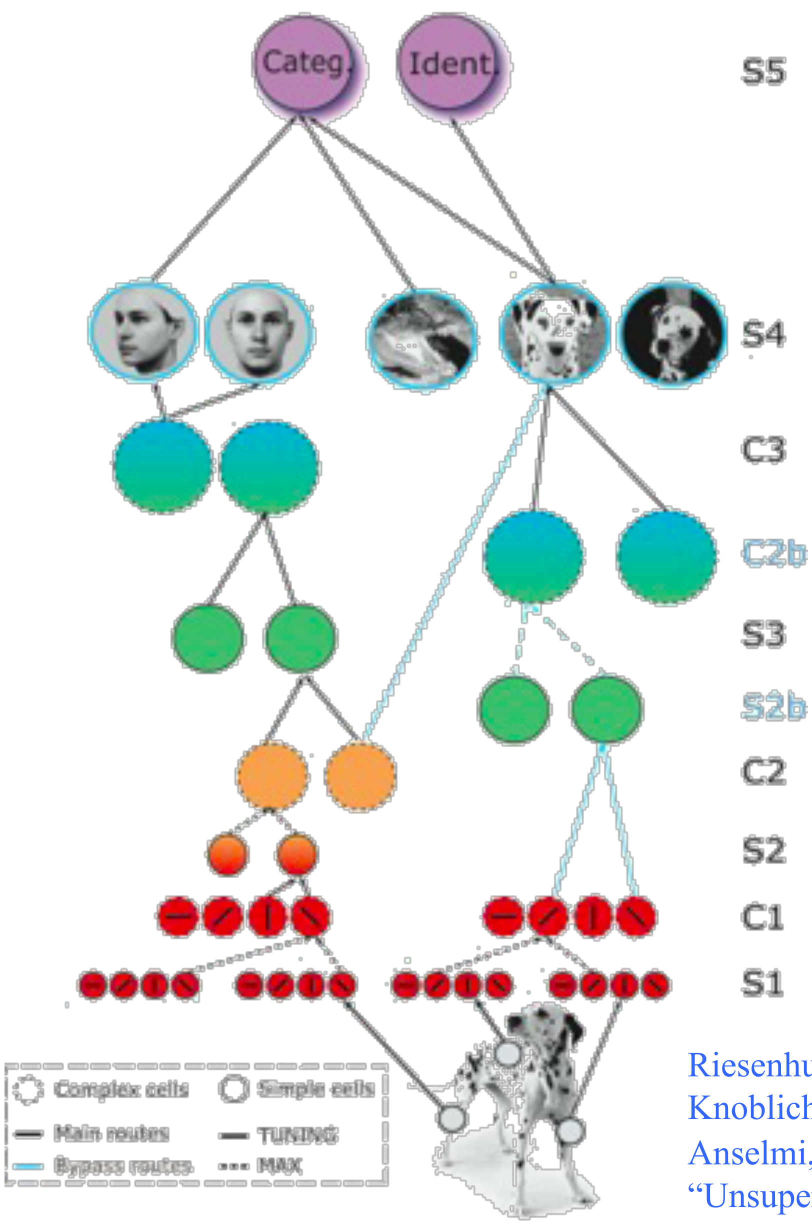
Summing up the volumes of all active areas, the total volume is 21,2 ml
or … 1.5 Bneurons

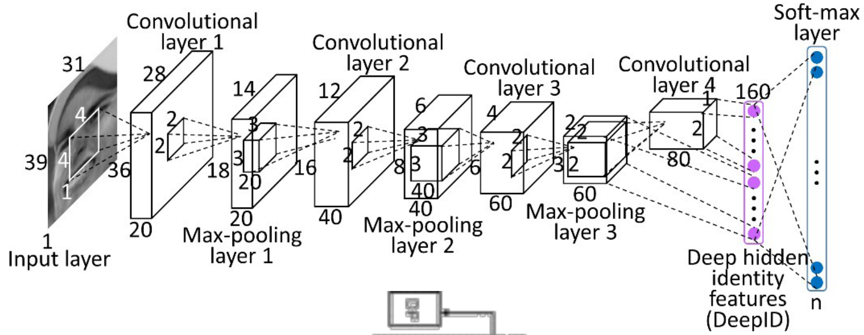
… with 12K Synapses/neuron!

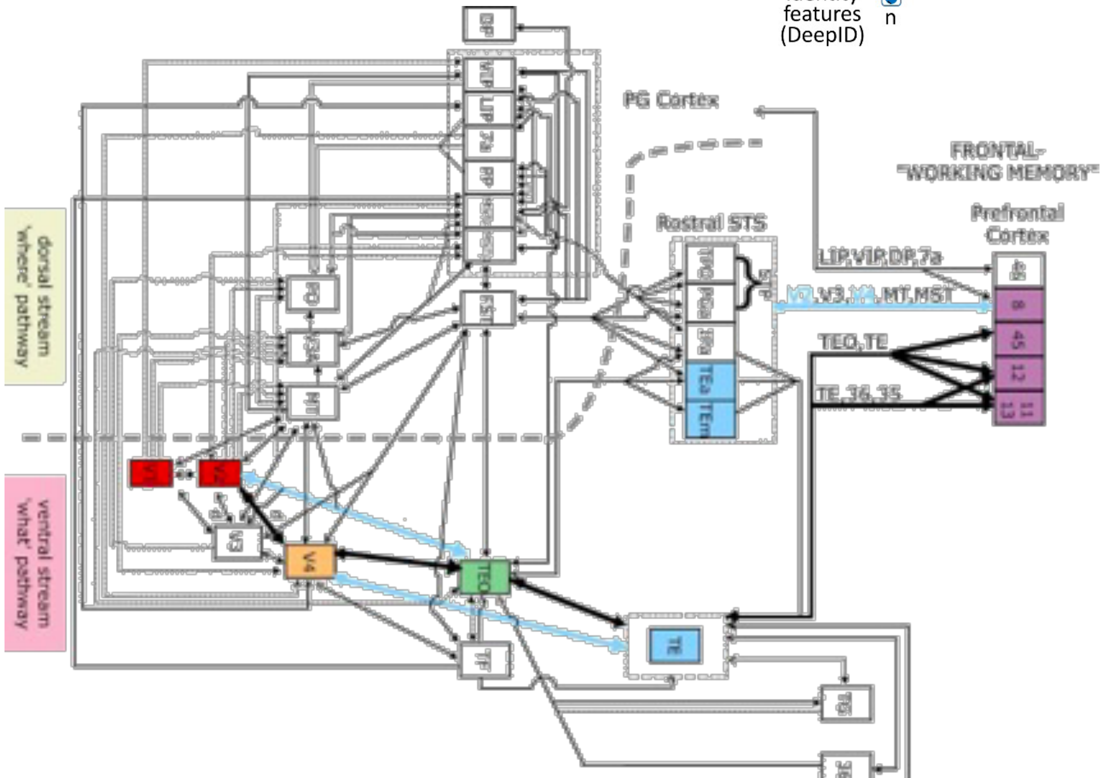
= 18 trillion synapses!
= 2.3 trillion Bytes?

If we can learn, say 10,000 faces this corresponds to
220 MB/face
(or a 7 sec. video stream of 1Kx1K images)

Brain models

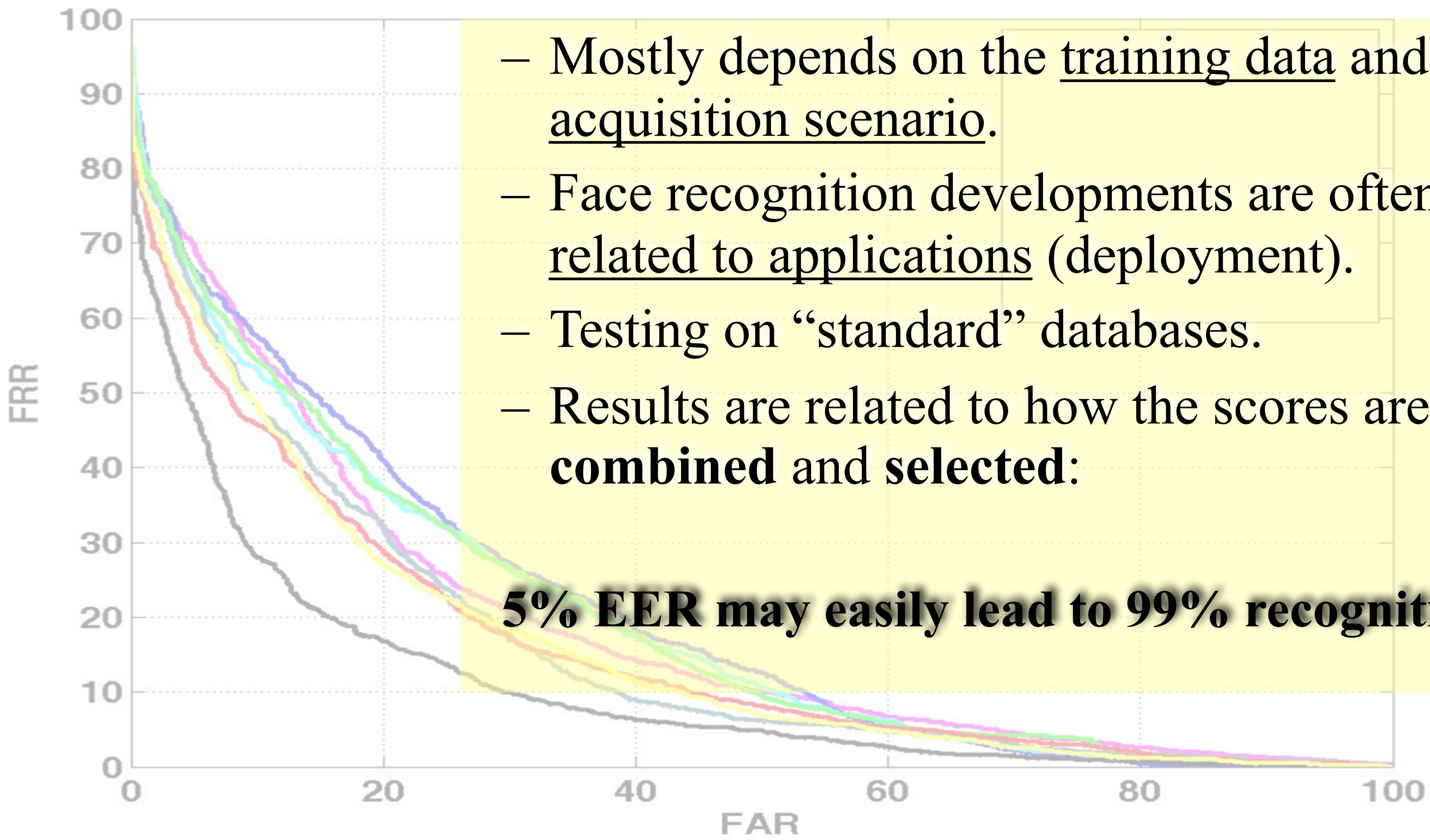






Riesenhuber & Poggio 1999, 2000; Serre Kouh Cadieu
Knoblich Kreiman & Poggio 2005; Serre Oliva Poggio 2007
Anselmi, F., Leibo, J. Z., Rosasco, L., Mutch, J., Tacchetti, A., and Poggio, T., "Unsupervised learning of invariant representations", Theoretical Computer Science, 2015.

Face recognition performances



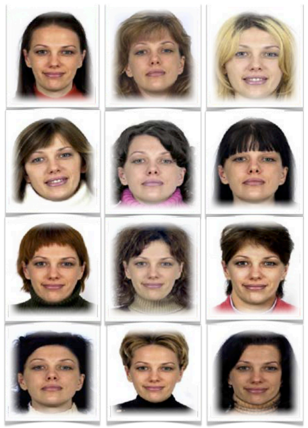
- Mostly depends on the training data and the acquisition scenario.
- Face recognition developments are often related to applications (deployment).
- Testing on “standard” databases.
- Results are related to how the scores are **combined** and **selected**:

5% EER may easily lead to 99% recognition

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

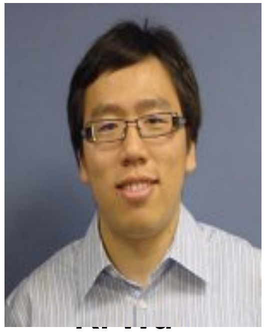

G. Medioni, Co-PI


R. Nevatia, Fusion


P. Debevec, Illumination


W. AbdAlmageed
Indexing, LSML


J. Choi
Face Recognition

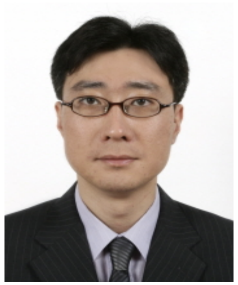

FD, Systems


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

Four Big Problems

Vision Lab

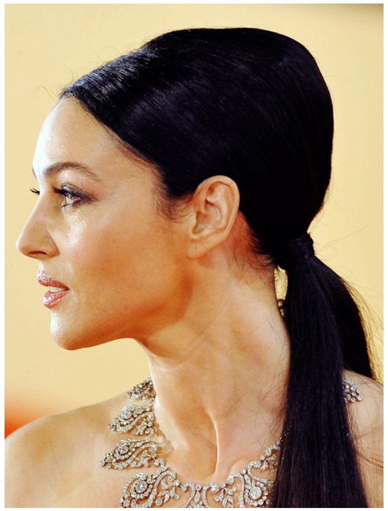
A – Aging

P – Pose

I – Illumination

E - Expression



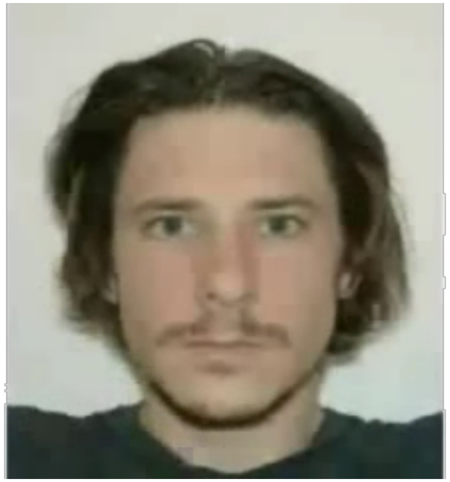
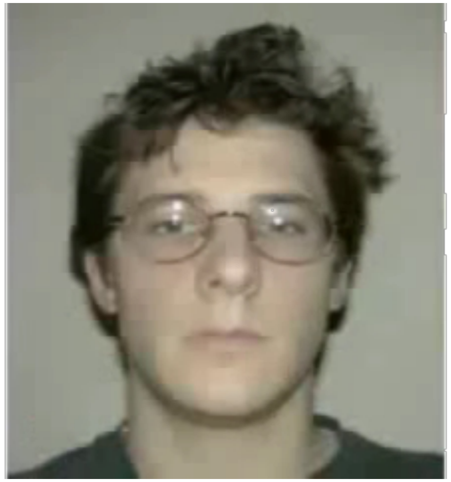
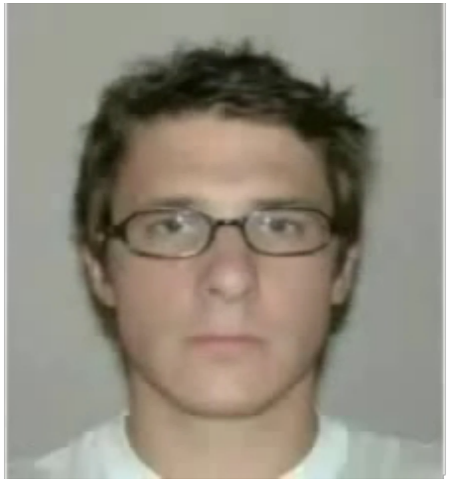



DEALING WITH AGE PROGRESSION

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)

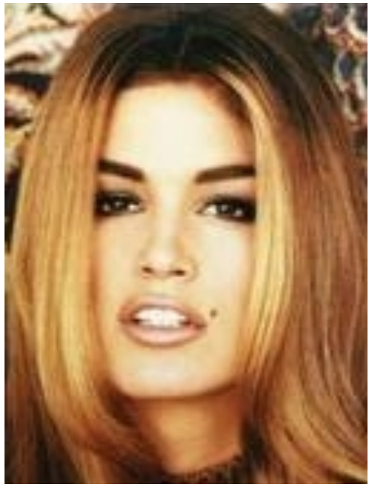
Aging effects

Time duration: 2 years

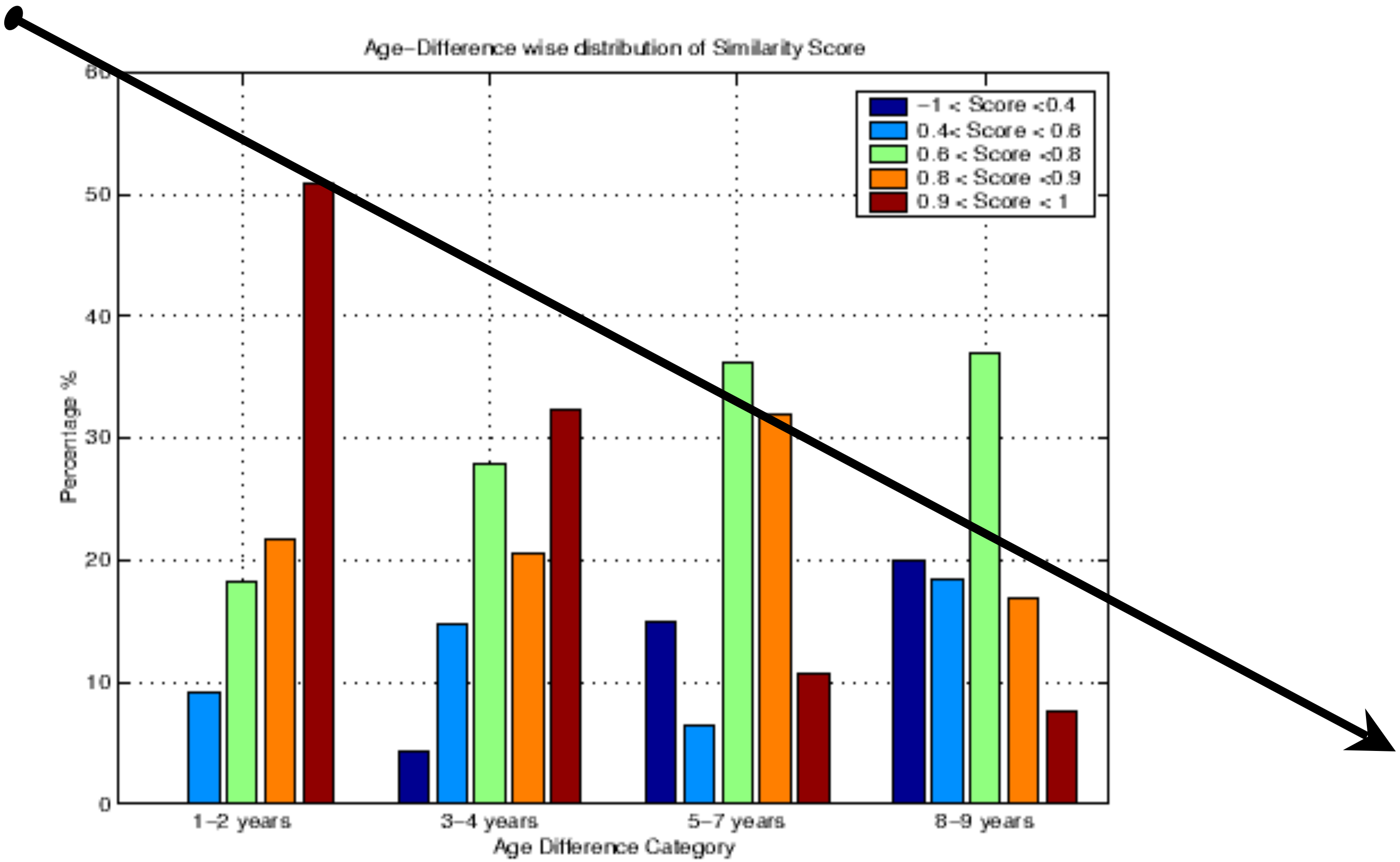




Time duration: several years



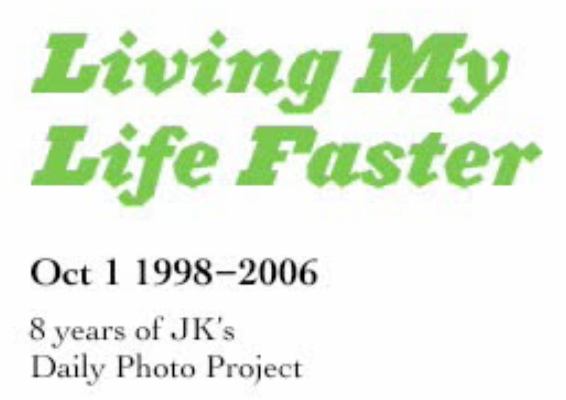



Effects of age progression

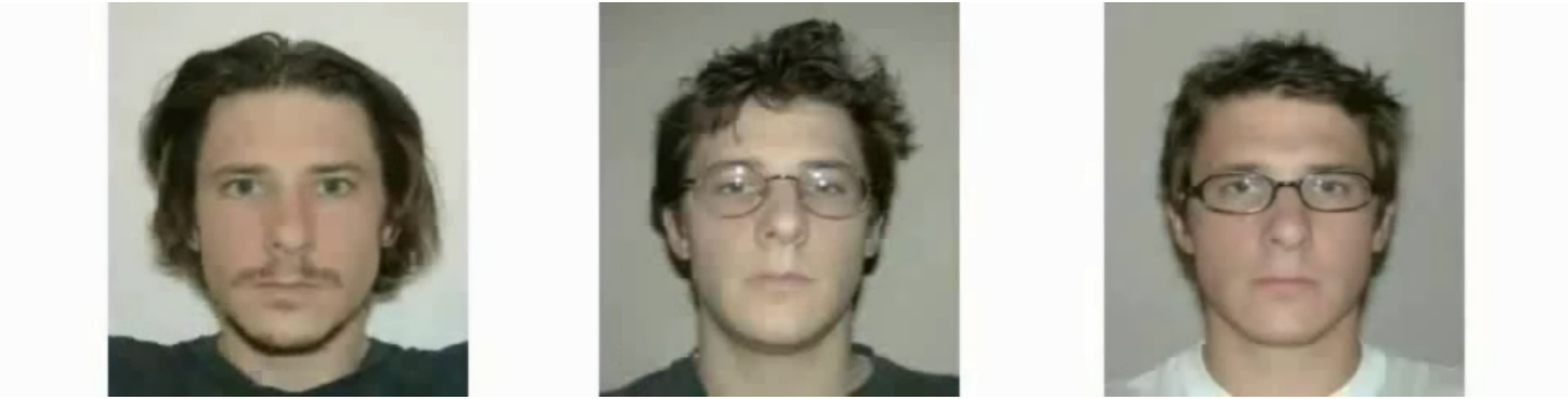


Similarity scores decreased as age difference increases

Aging … over time



Aging … over time

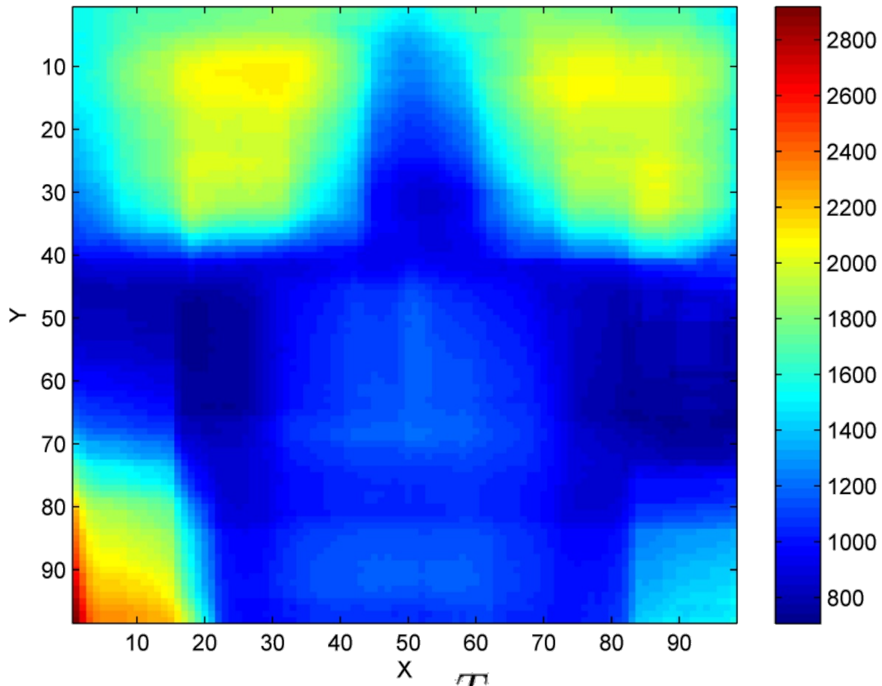


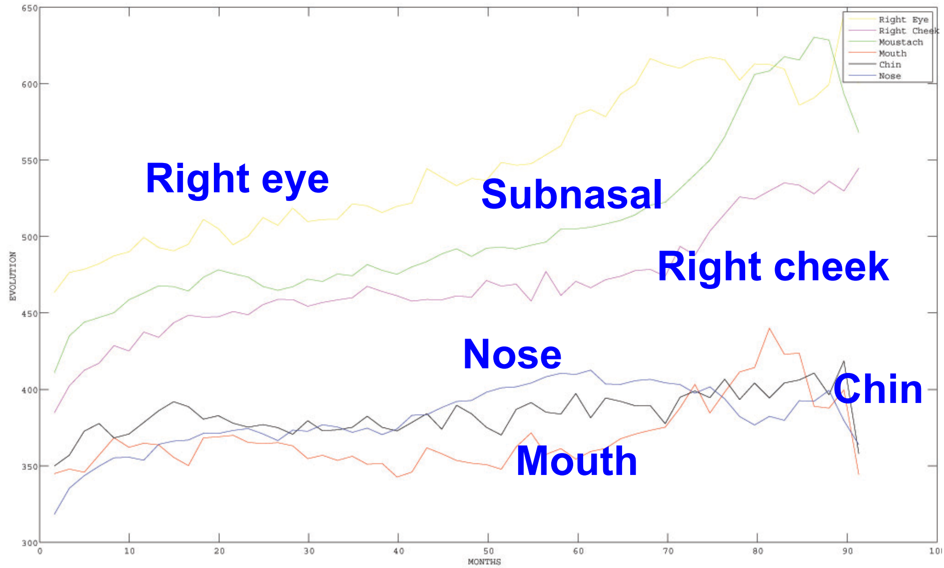
$$d_{I_1,I_2}(x,y) = \frac{1}{2}\left(\frac{1}{|P_{I_1}|}\sum_{p\in P_{I_1}}\omega(p) + \frac{1}{|P_{I_2}|}\sum_{q\in P_{I_2}}\omega(q)\right)$$

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 4 years





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

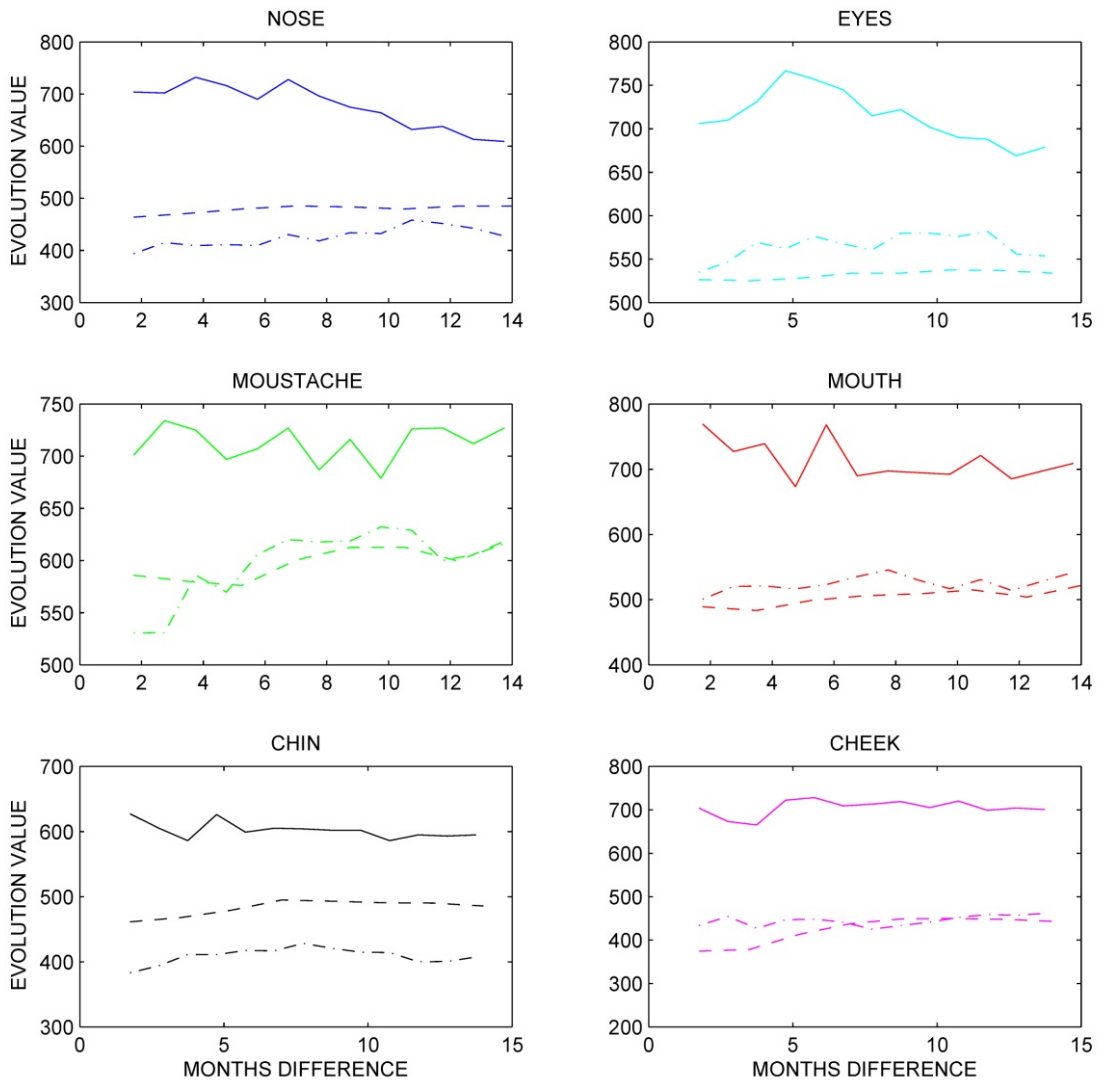
$$E(L,t) = \frac{1}{w_L h_L}\sum_{(x,y)\in L} E(x,y,t)$$

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)

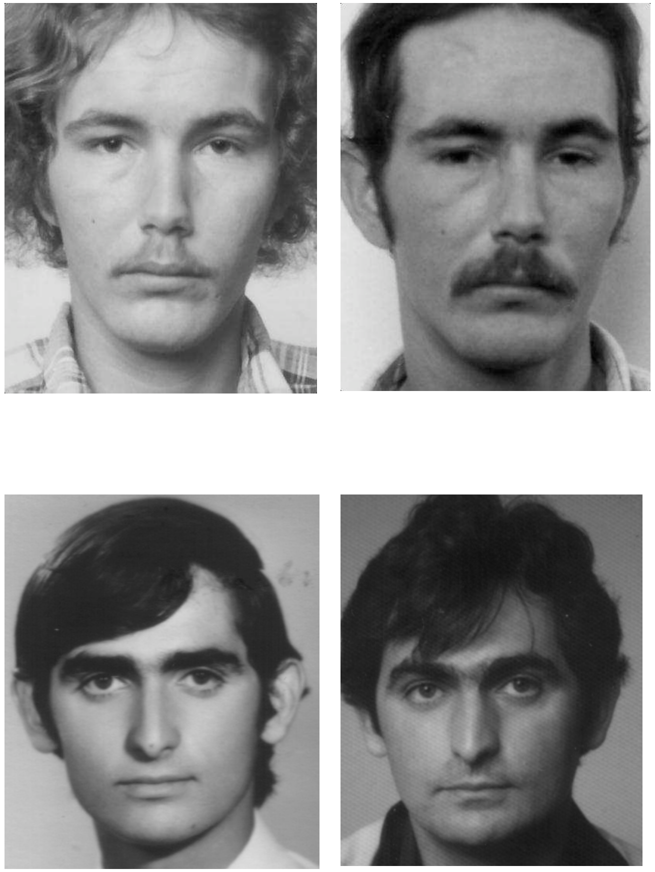
Photometric effects

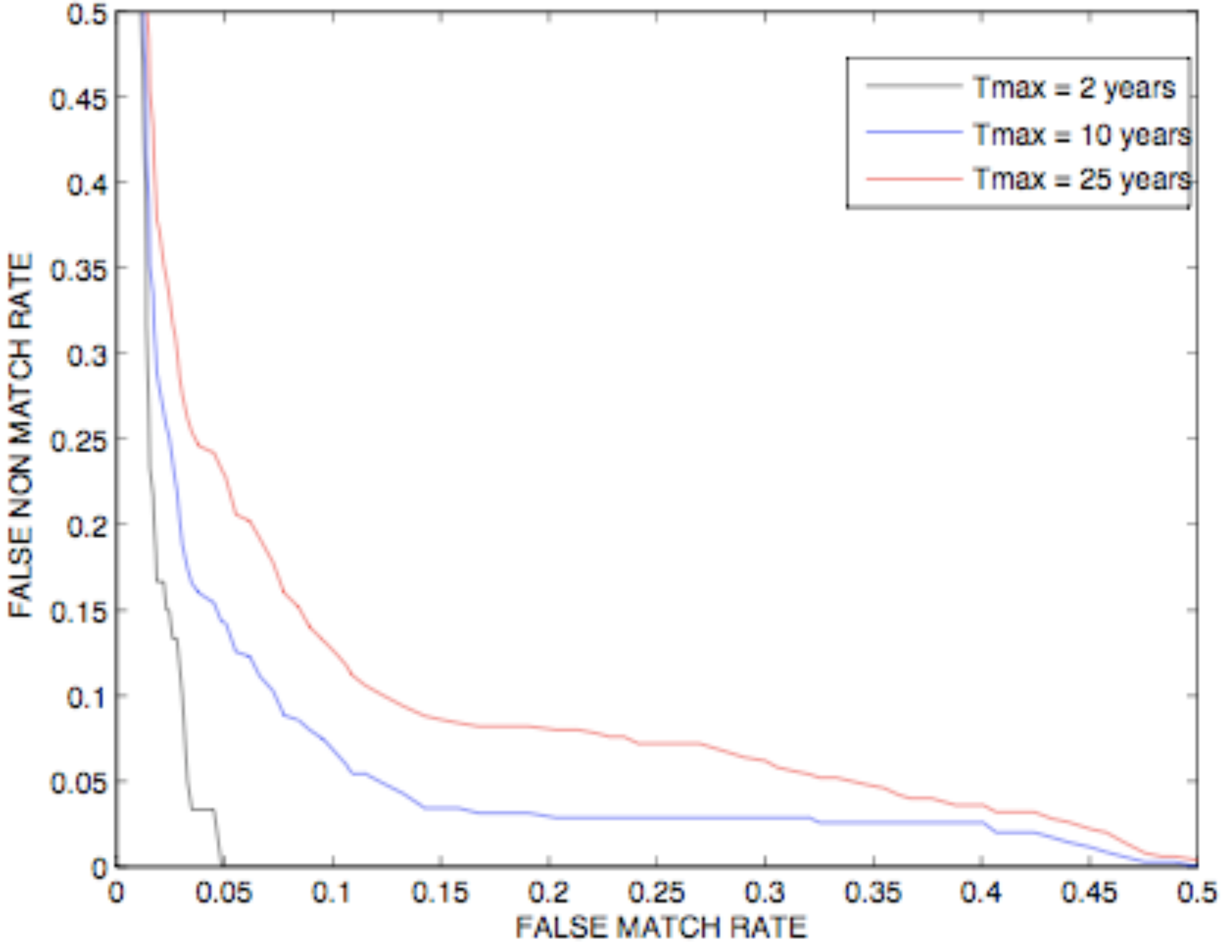
Vision Lab

Comparative time evolution of features for different subjects



Face matching across age



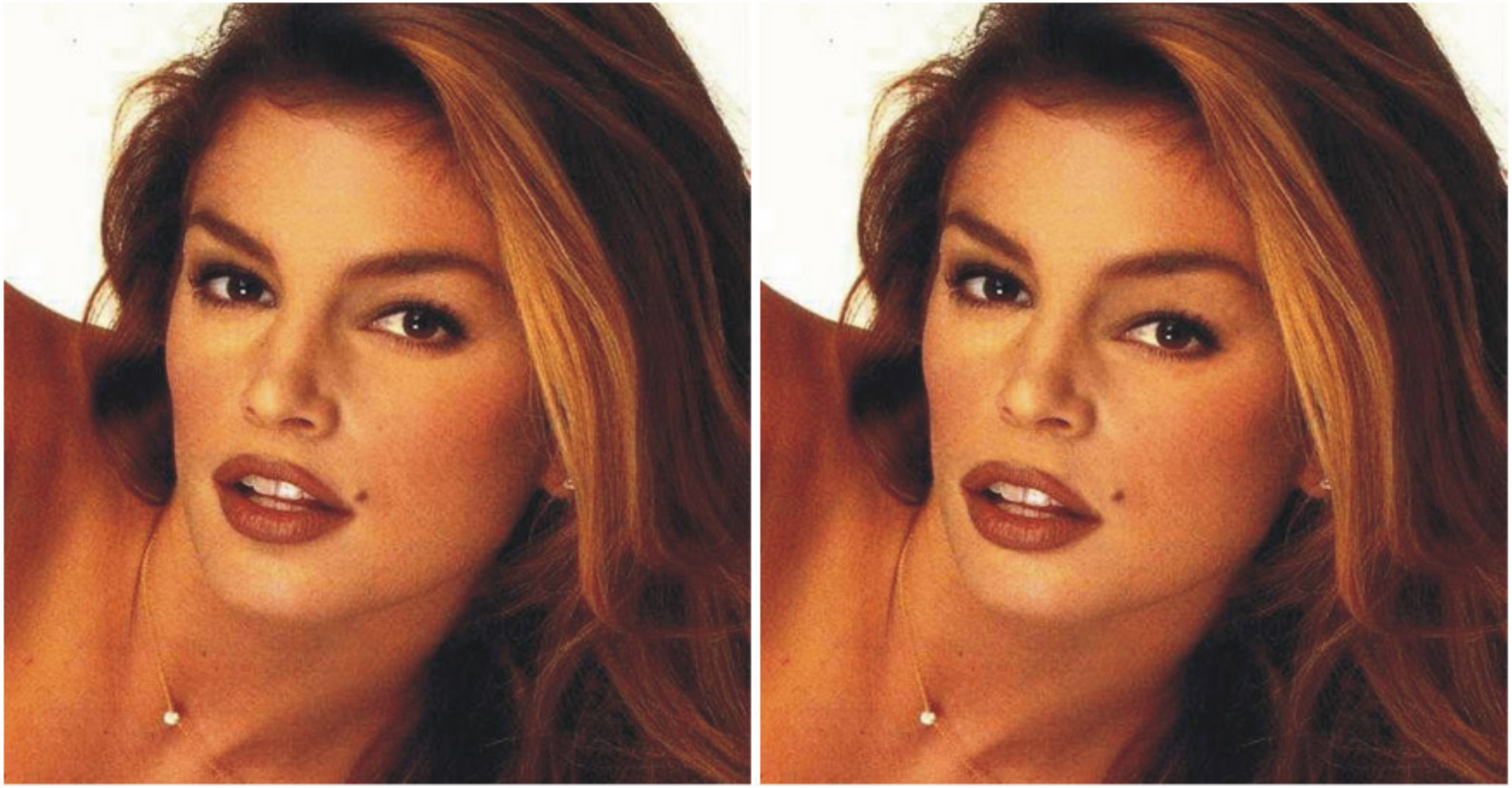


M. Ortega, L. Brodo, M. Bicego, M. Tistarelli, "Measuring changes in face appearance through aging,", Proc. of IEEE Computer Vision and Pattern Recognition Workshop, pp. 107-113, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (2009)

DEALING WITH FACIAL POSE

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

Face alignment



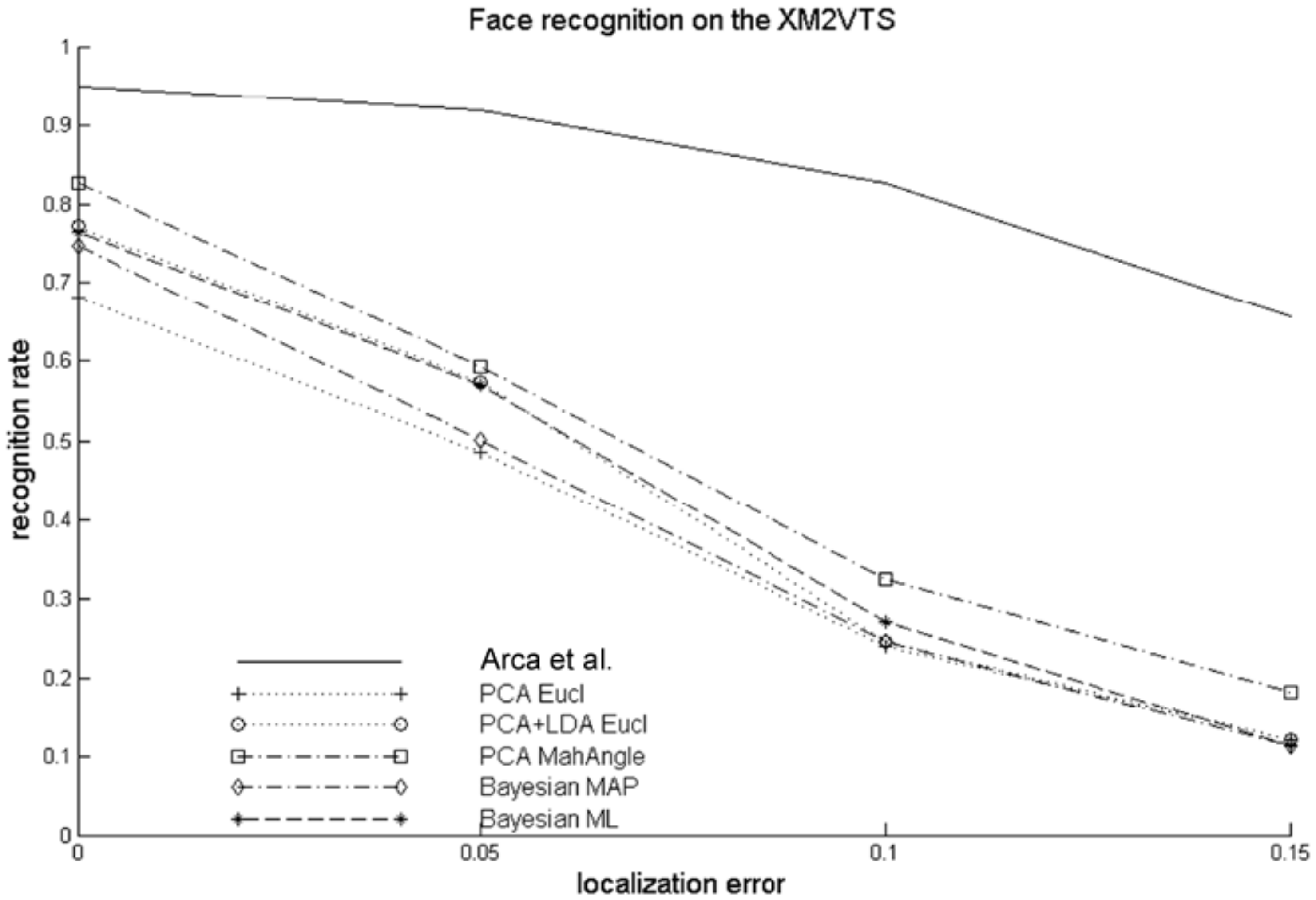
… But the data is still there

Face alignment

In cognitive psychology it is called

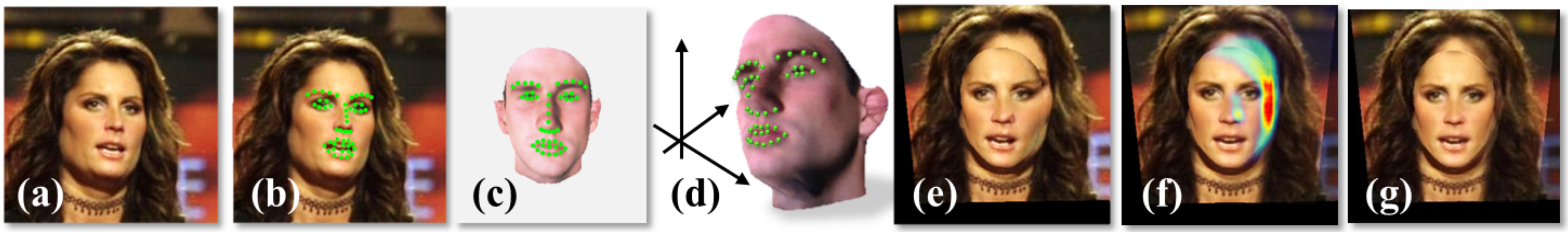
perceptual organization

Face alignment



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.

2D Frontalization



(a) Query photo; (b) landmarks detection; (c) textured 3D computer graphics model with landmarks;
(e) The estimated projection matrix is used to back-project the query intensities to the reference coordinate system;
(f) Estimated visibility overlaid on the frontalized result. Warmer colors reflect less visible pixels.
(g) Facial appearance produced by borrowing texture from simmetric face areas.

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

DEALING WITH FACIAL ILLUMINATION

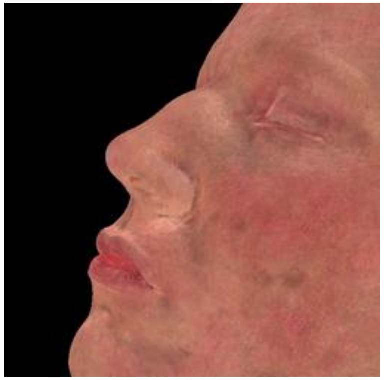
R. Gross and V. Brajovic, "An Image Preprocessing Algorithm for Illumination Invariant Face Recognition", International Conference on Audio- and Video-Based Biometric Person Authentication, 2003.

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.

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)

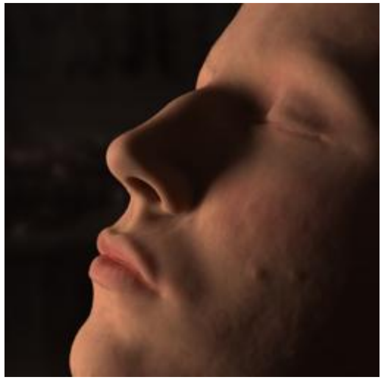
Modelling the face skin



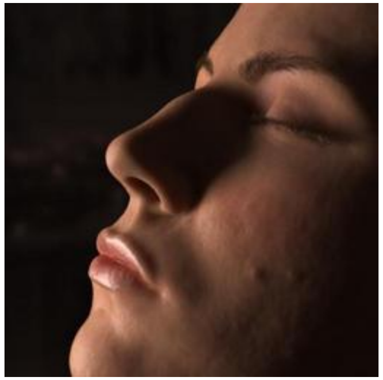
Skin chromaticity map

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

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

$$F(L) = \iint_{\varpi} \rho(x,y)(L(x,y) - I(x,y))^2 dxdy + \lambda \iint_{\varpi} (L_x^2 + L_y^2) dxdy \quad (1)$$

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.

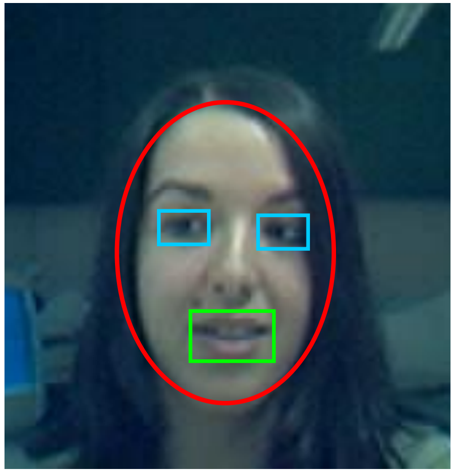
D. Jobson, Z. Rahmann and G. Woodell, "A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observations of Scenes", IEEE Transanctions on Image Processing, volume 6, Issue 7, 1997.

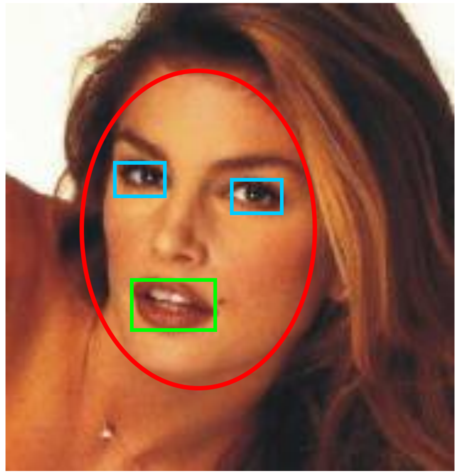
SUBJECT-SPECIFIC FACE REPRESENTATION

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.

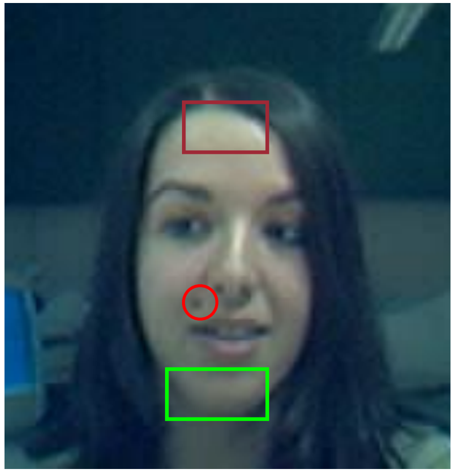
Subject-specific representation

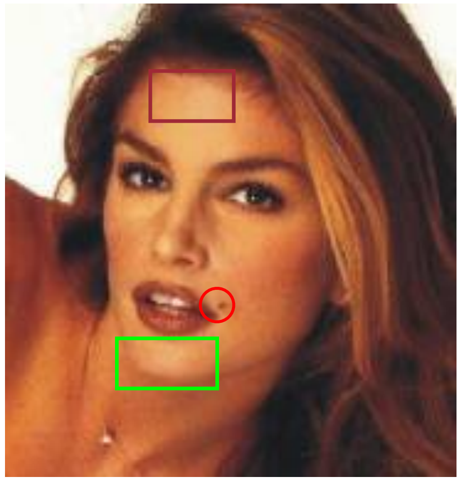
For **localization** and **tracking** we are interested on what every face has **<u>in common</u>** (to tell a face from “non-faces”)



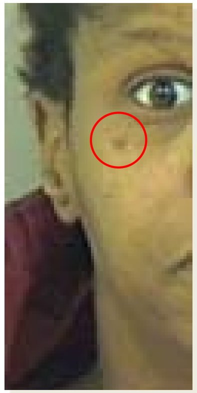


For **identification** we are not interested on what faces have in common but rather **<u>what differentiates</u>** one face from another.

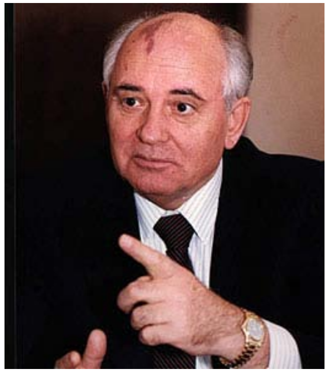




Facial Marks

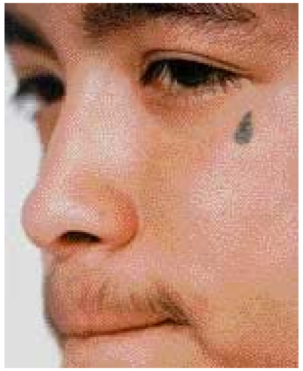


Mole in partial view

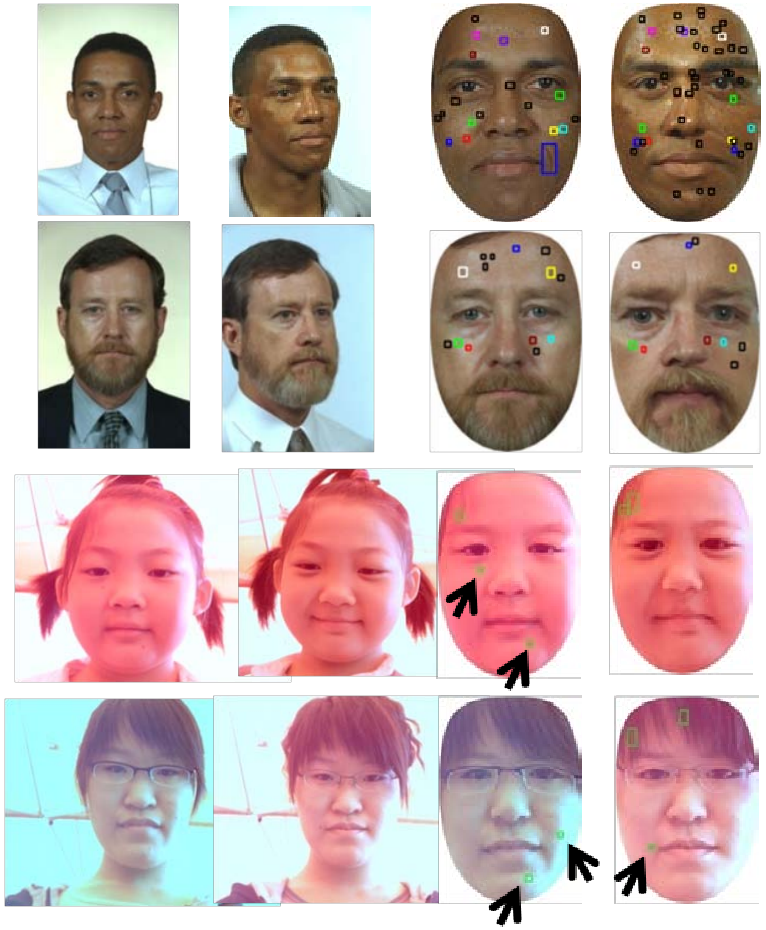


Birth mark

Mole in side view



Face tattoo



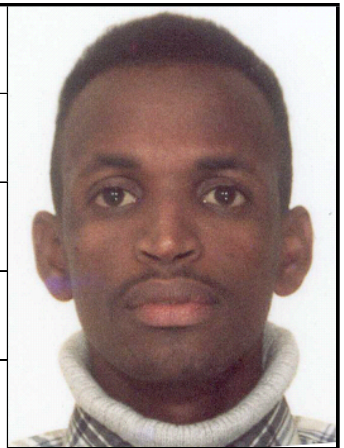
(a) (b) (c) (d)

Figure 10. Example image pairs where facial marks helped to improve the matching performance. (a) (b) Probe and gallery images for the first two rows and two images of identical twins for the third and fourth rows. (c) (d) Facial mark detection results from (a) and (b).

A.K. Jain, B. Klare, U. Park. (2011) "Face Recognition: Some Challenges in Forensics", 9th IEEE Int'l Conference on Automatic Face and Gesture Recognition, Santa Barbara, CA, March, 2011.

Forensic face evaluation

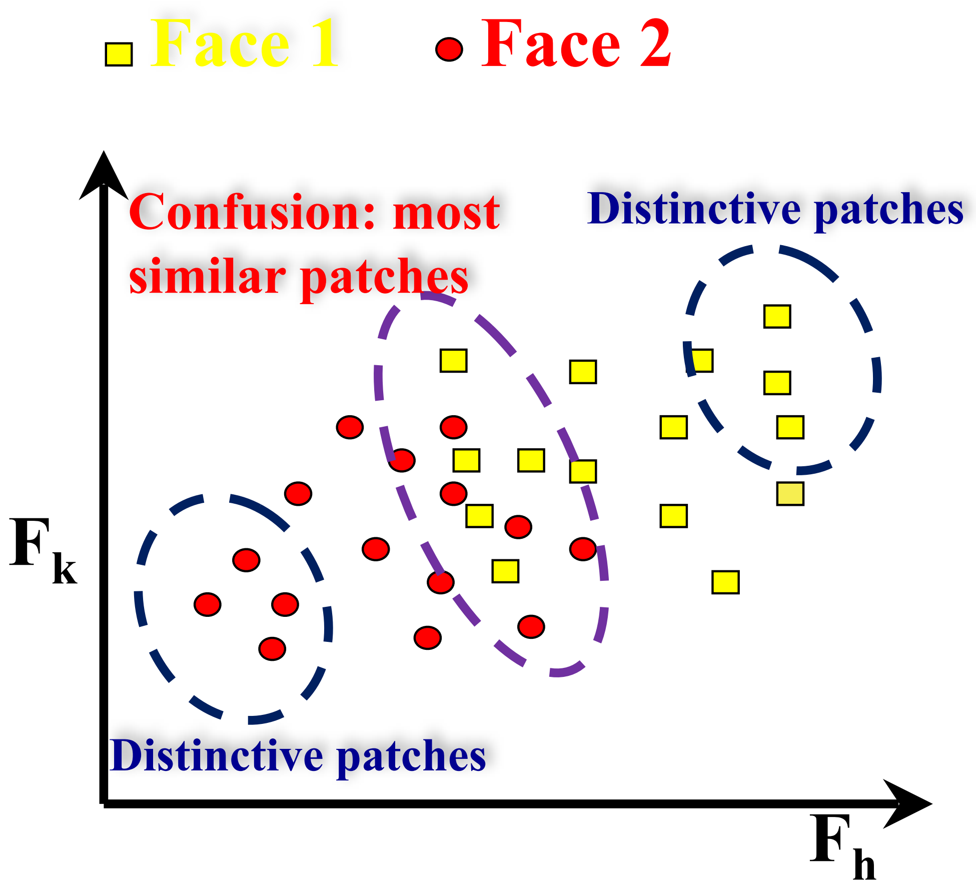
Feature	Details	Similarities and differences					
		S	NO	D	SI	Explanation for differences: Assumption of same source	Explanation for differences: Assumption of different source
Face	Shape	X					
	Proportions	X					
	Hairline Hairgrowth			X		Age, hairdo	
Forehead	Shape	X					
	Bumps	X					
	Horizontal creases		X				
	Eyebrows	X					

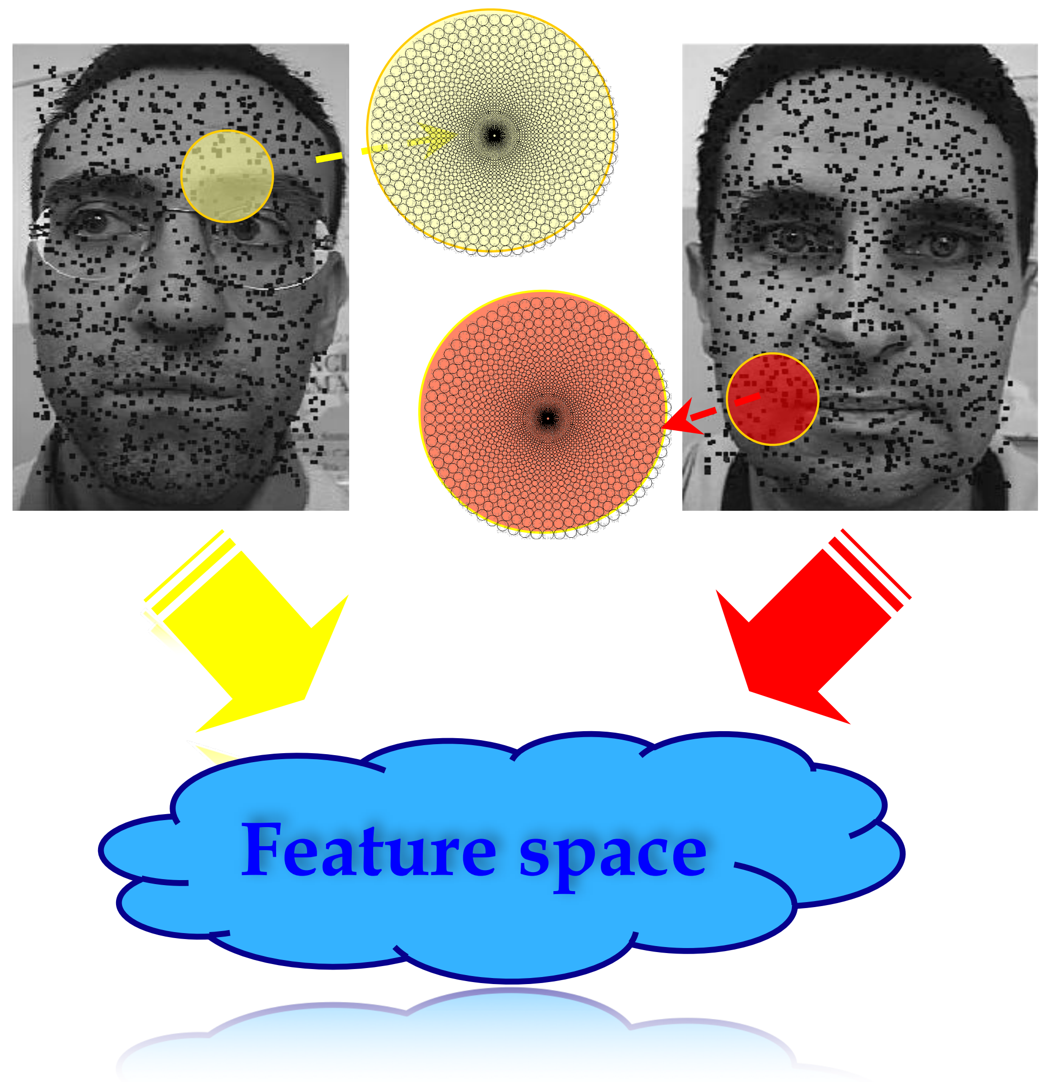


52 Didier Meuwly | The use of biometric information in forensic practice | 12.06.2014 | UNIVERSITY OF TWENTE. EEMCS | Netherlands Forensic Institute *Ministry of Security and Justice*

Courtesy of Didier Meuwly NFI

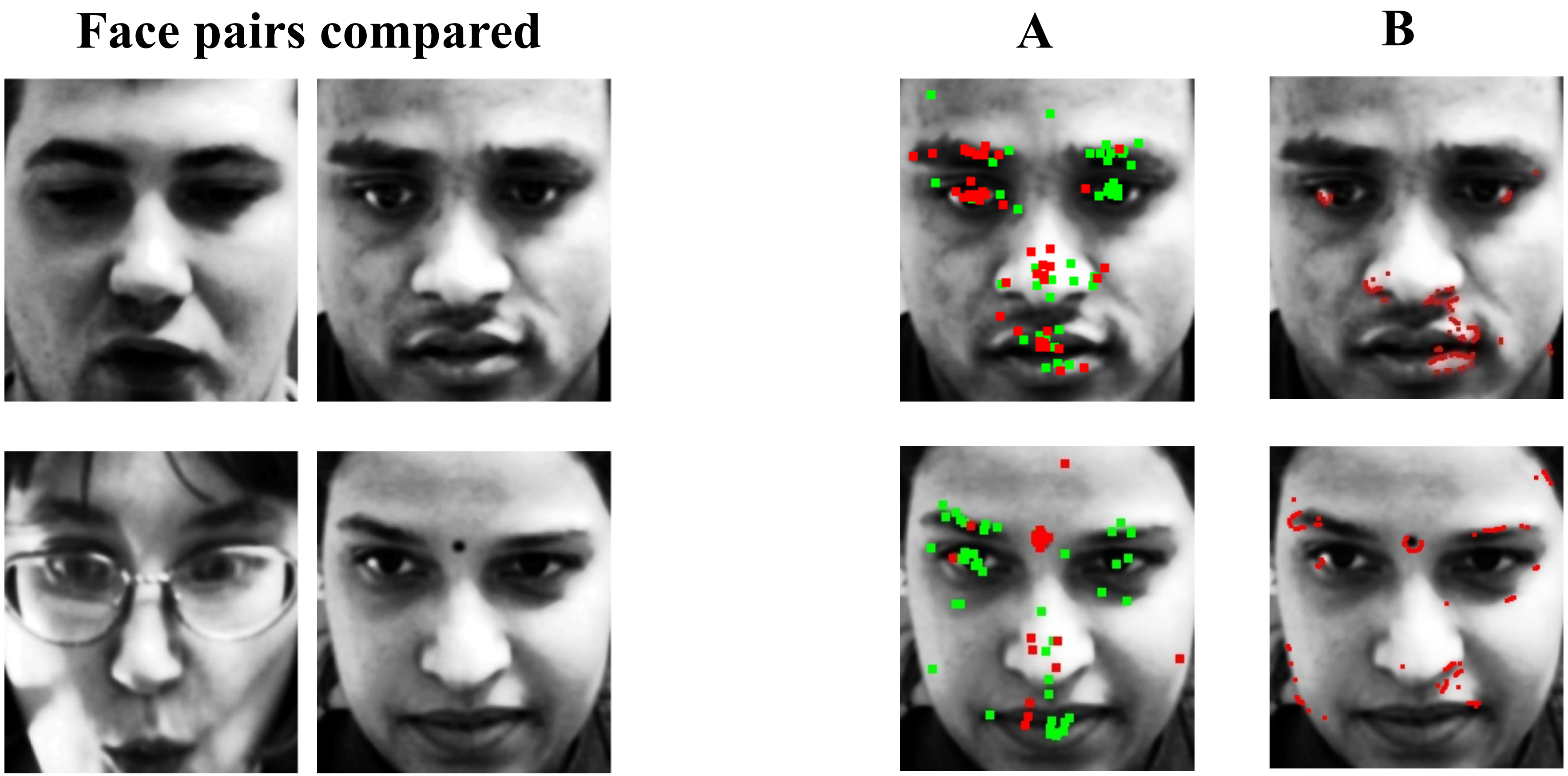
Face Characterization





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

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.

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

Vision Lab

- 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 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.

Common cosmetic procedures

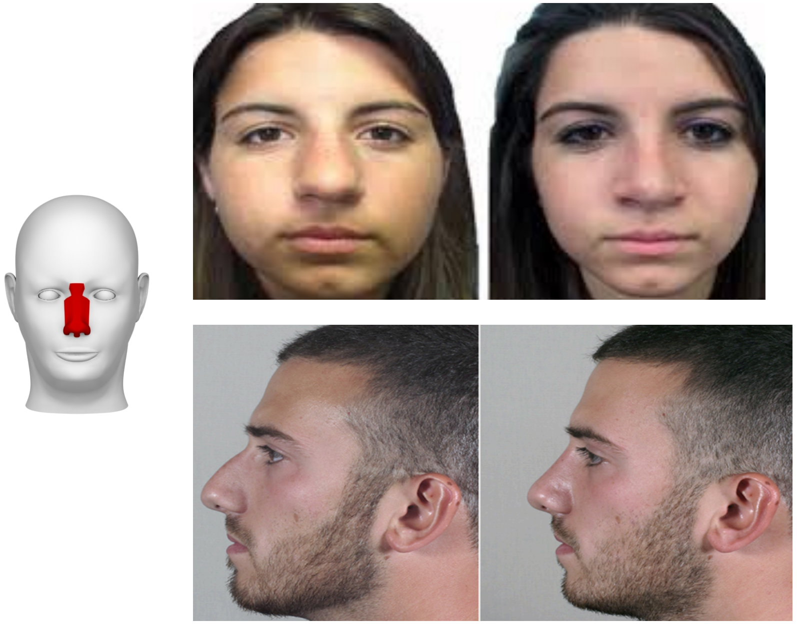
Botulinum toxin

Chemical peel

Dermal fillers

Dermoabrasion

Common cosmetic procedures



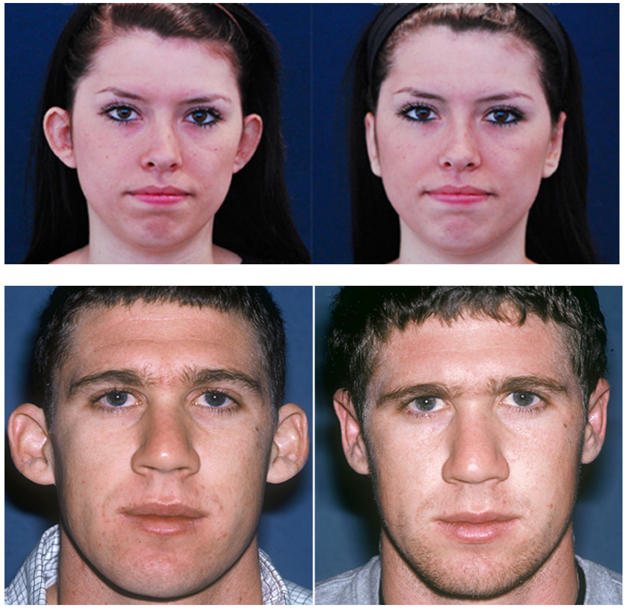
Rhinoplasty

Rhytidectomy (face lift)

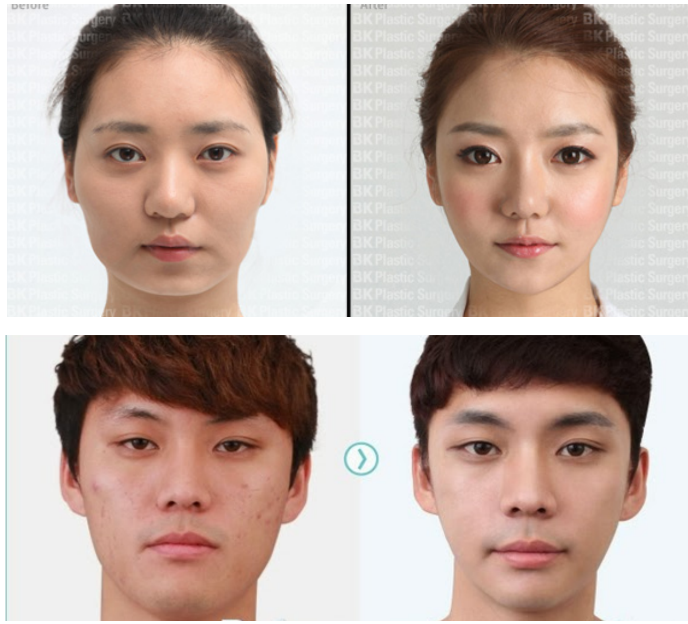
Blepharoplasty

Brow/Forehead lift

Common cosmetic procedures



Otoplasty



Cheek bones reshaping

Mentoplasty

Vision Lab

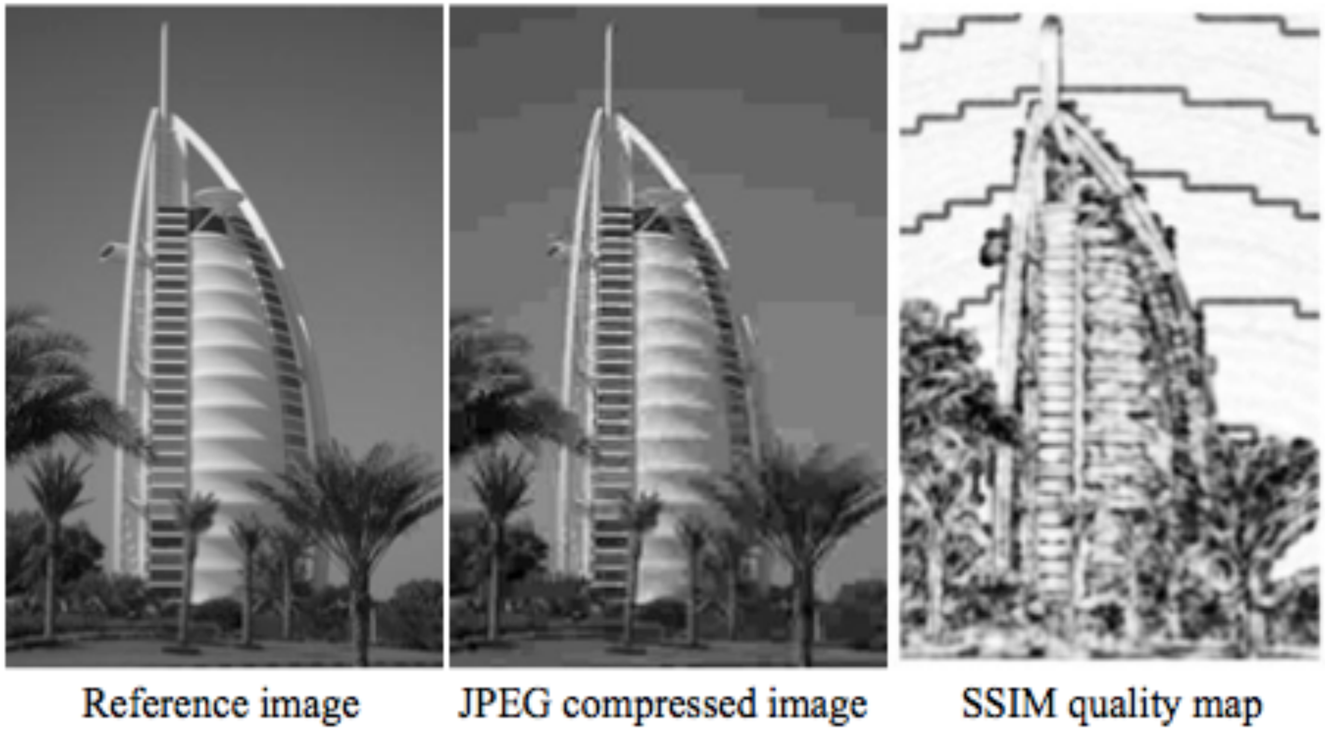
Common cosmetic procedures

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 region	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 region	Low	Medium	High	0.15%
Cheekbones reshaping	Zygomatic region	Low	Medium	High	0.1%
Ear surgery (*Otoplasty*)	Ears	Low	Limted	Low	0.2%

Face recognition algorithms

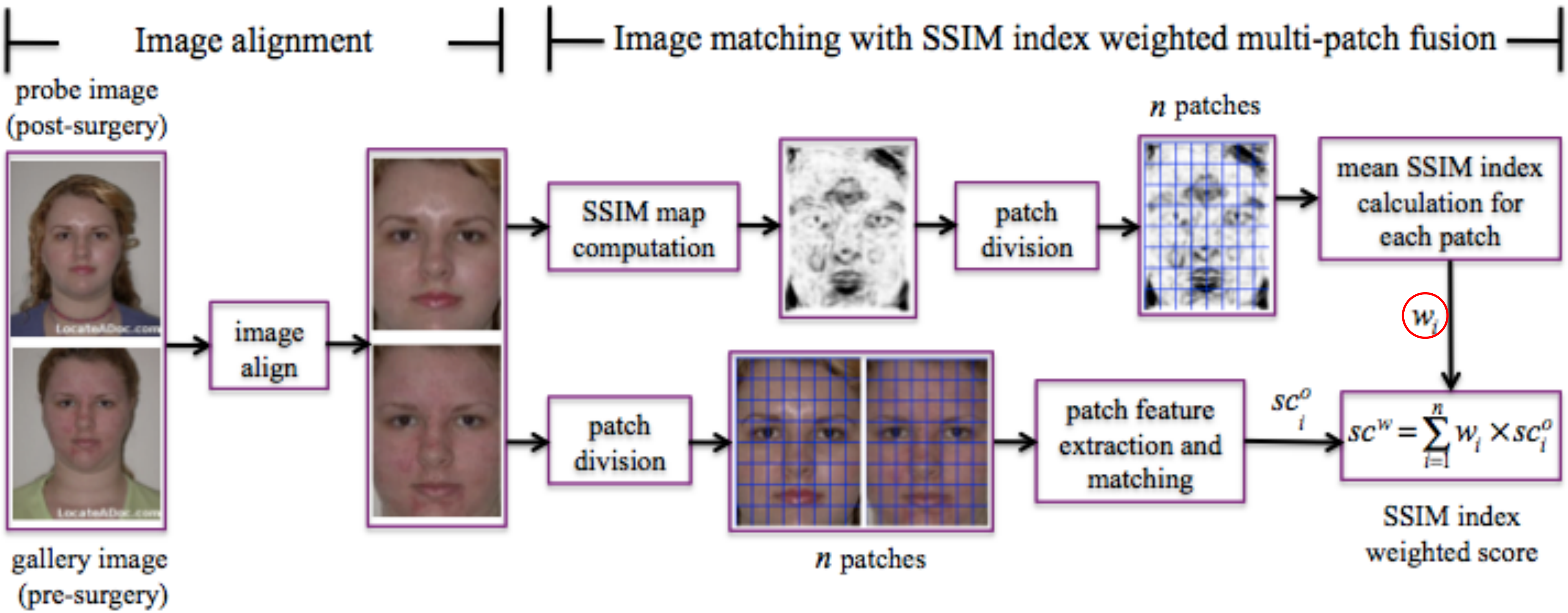
#	Reference	Dataset	Key Features					Algorithm
			GLOBAL	LOCAL	TEX	3D	RR%	
1	Aggarwal et al. [26]	Plastic Surgery Face Database	N	Y	N	N	*77.9*	Part-wise and Sparse representation
2	Bhatt et al. [40]	Plastic Surgery Face Database	Y	Y	Y	N	*78.6*	Uniform Circular Local Binary Pattern (UCLBP) + Speeded Up 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

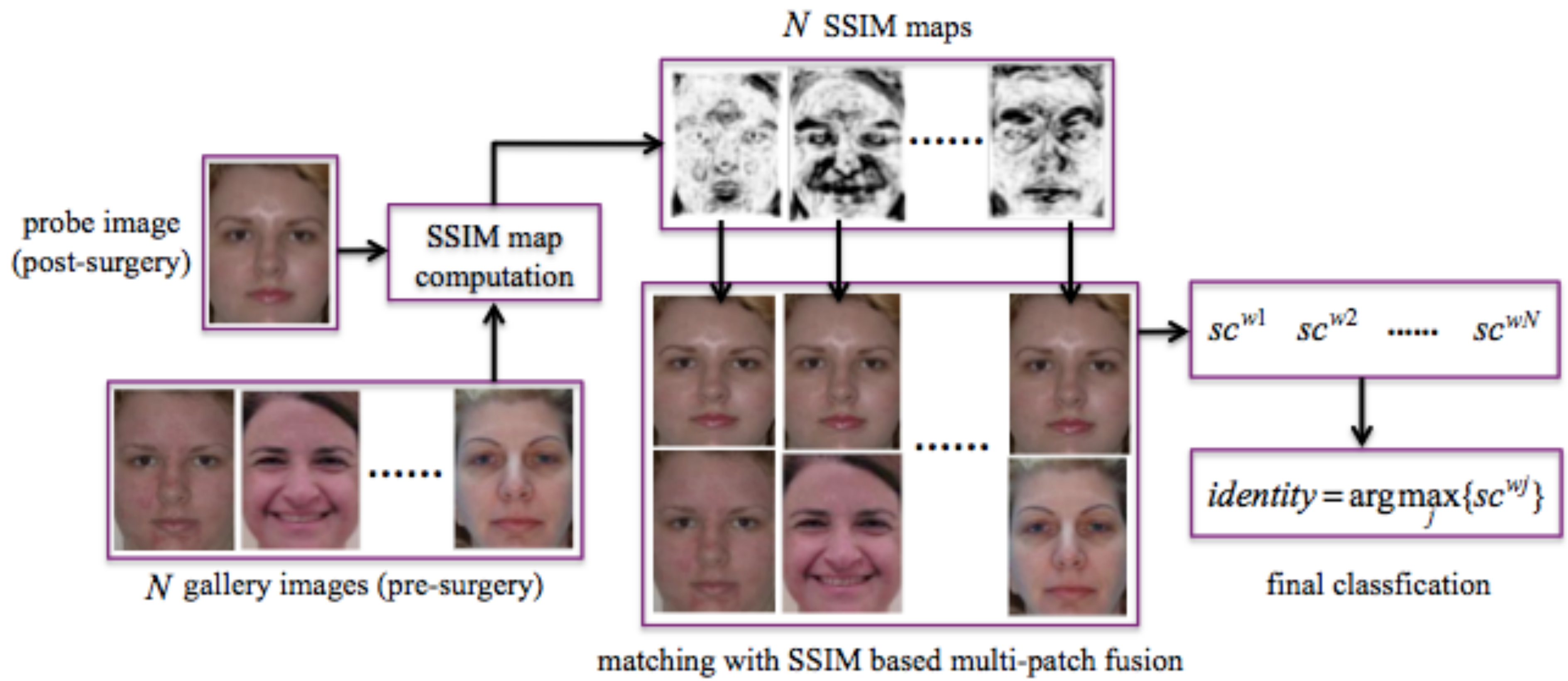


$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Face recognition using SSIM



Face recognition using SSIM



SSIM examples

Blepharoplasty

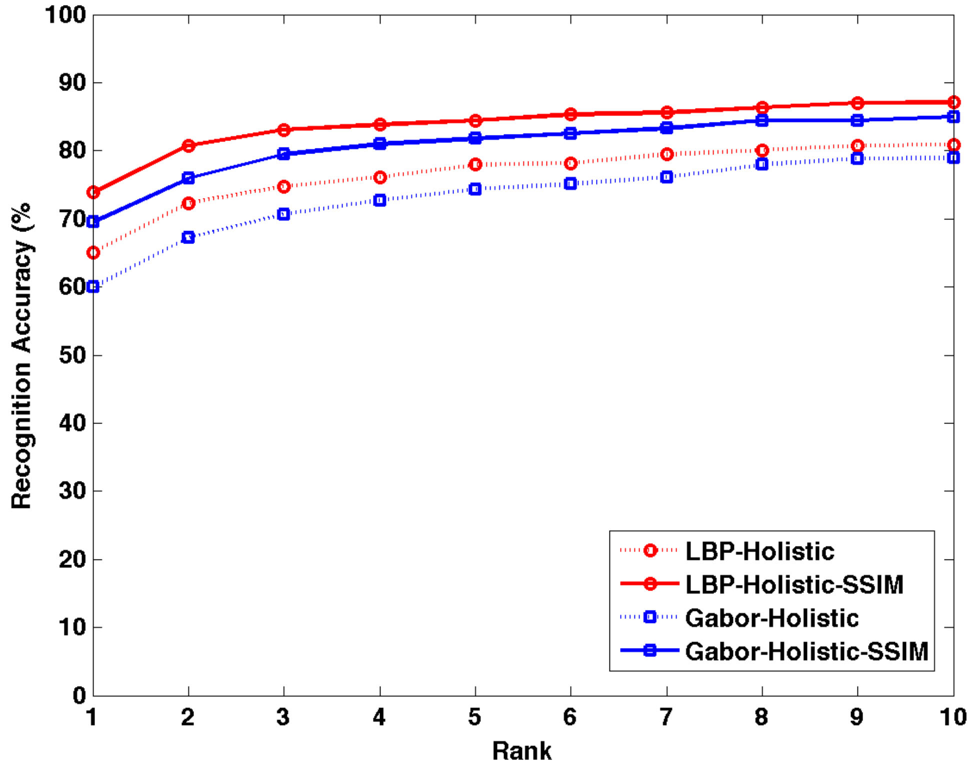
Laser skin resurfacing

Rhinoplasty

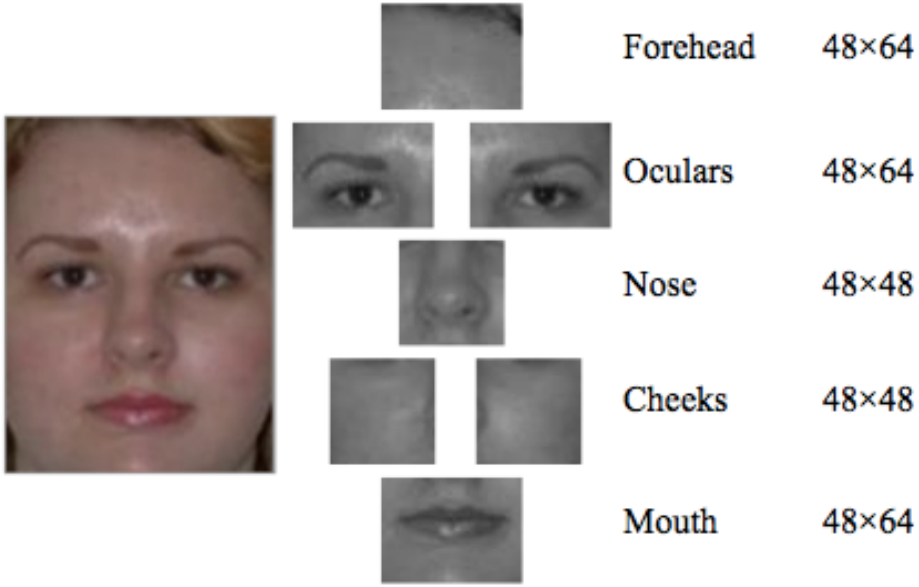
Lip augmentation

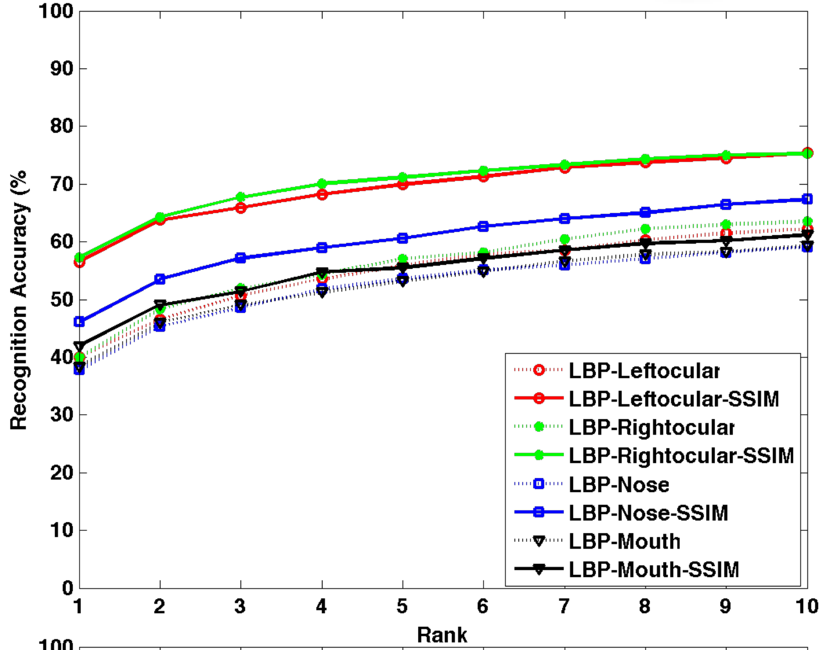
Experimental results

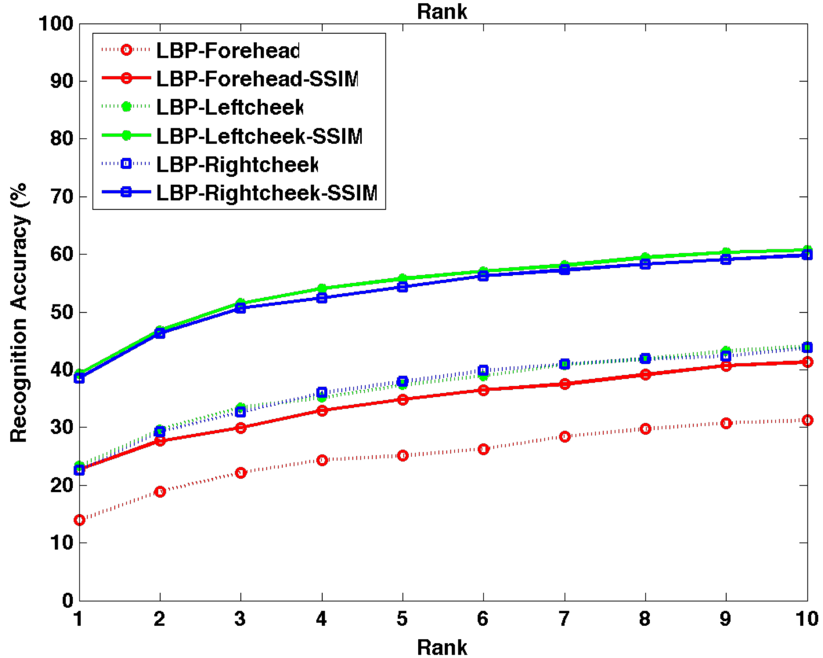
Plastic surgery database containing 576 images of 784 subjects taken from the web.



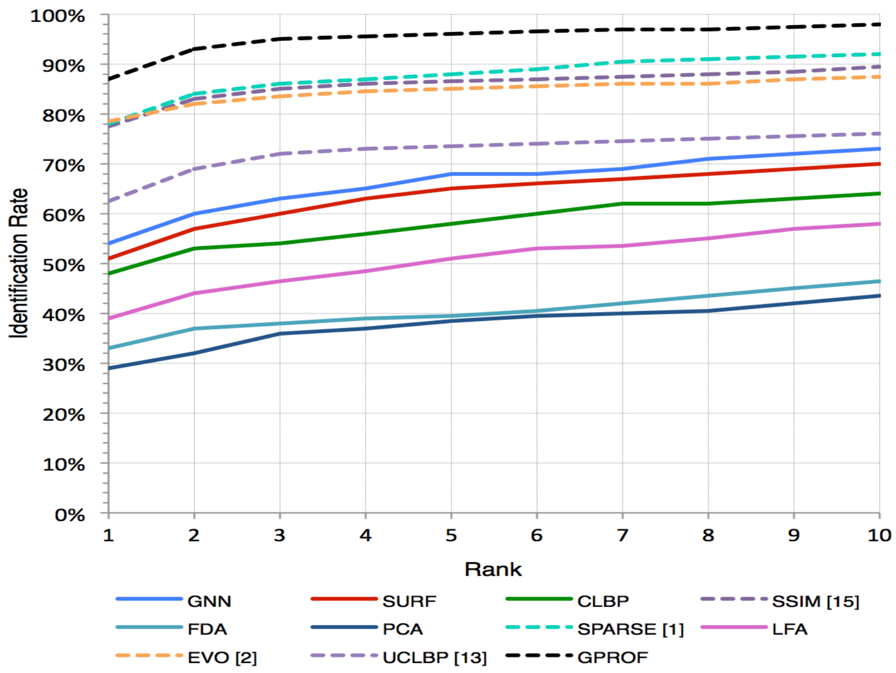
Component-based matching

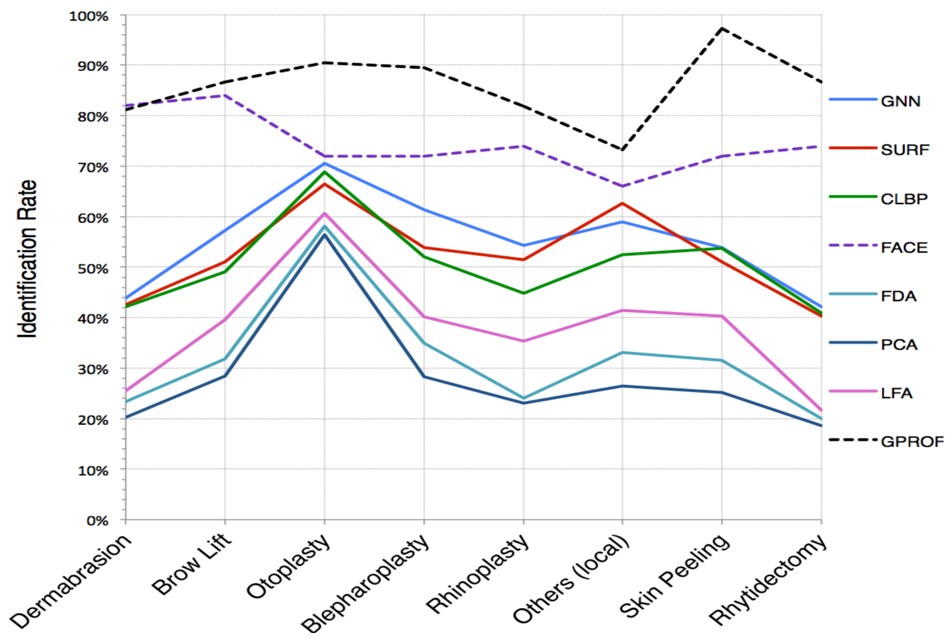






Comparative performance

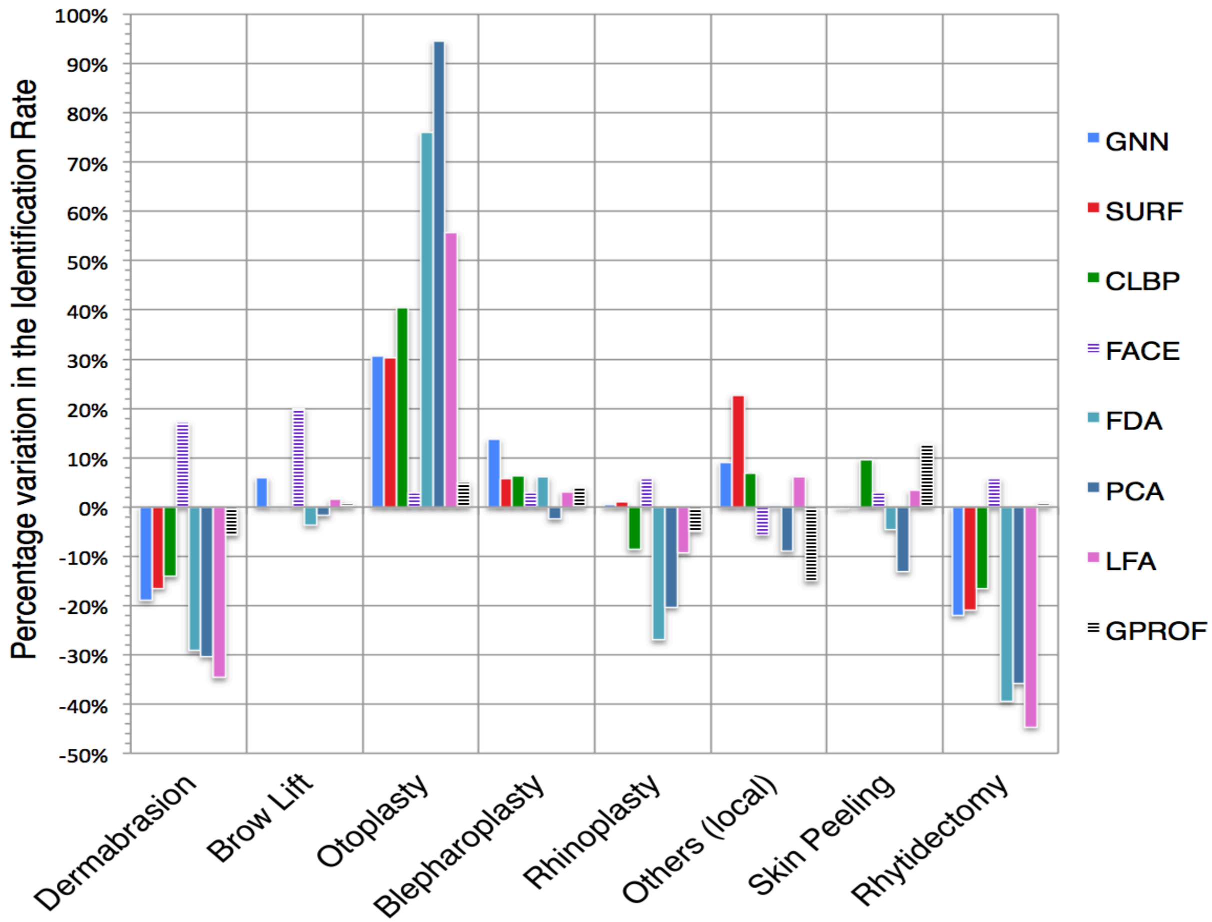




Comparison of the Cumulative Match Characteristic curves computed from eleven different face recognition algorithms applied to the same plastic surgery database. Dashed lines refer to region-based approaches, while solid lines refer to holistic approaches.

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.

Comparative performance

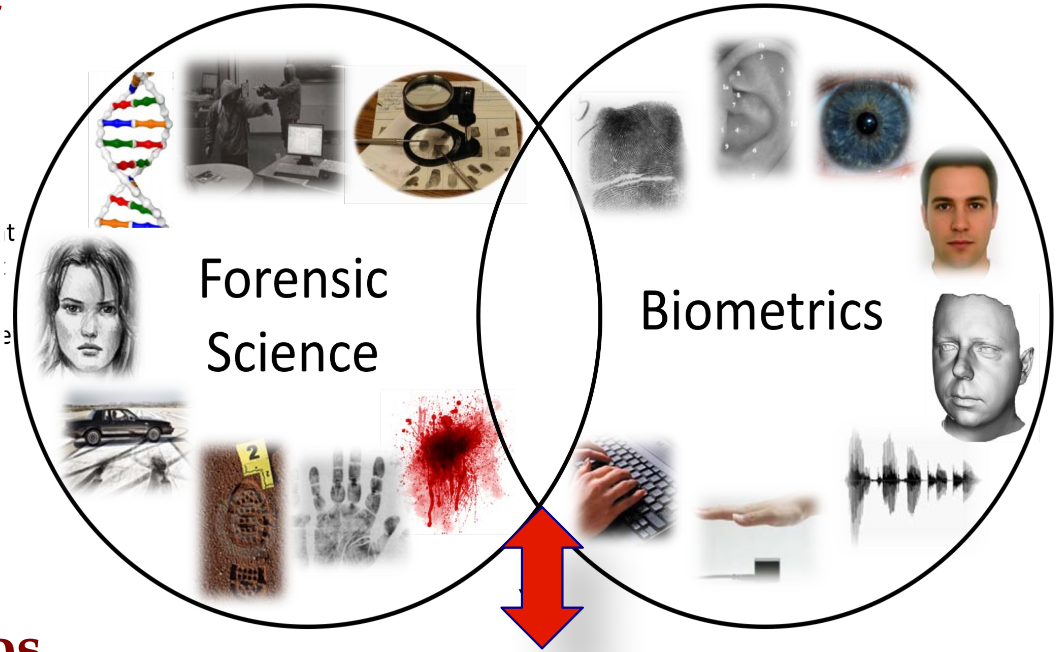


Comparison of overall vs. procedure-wise performance of eight different algorithms. The identification rate is normalized to 1.

APPLICATIONS TO 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

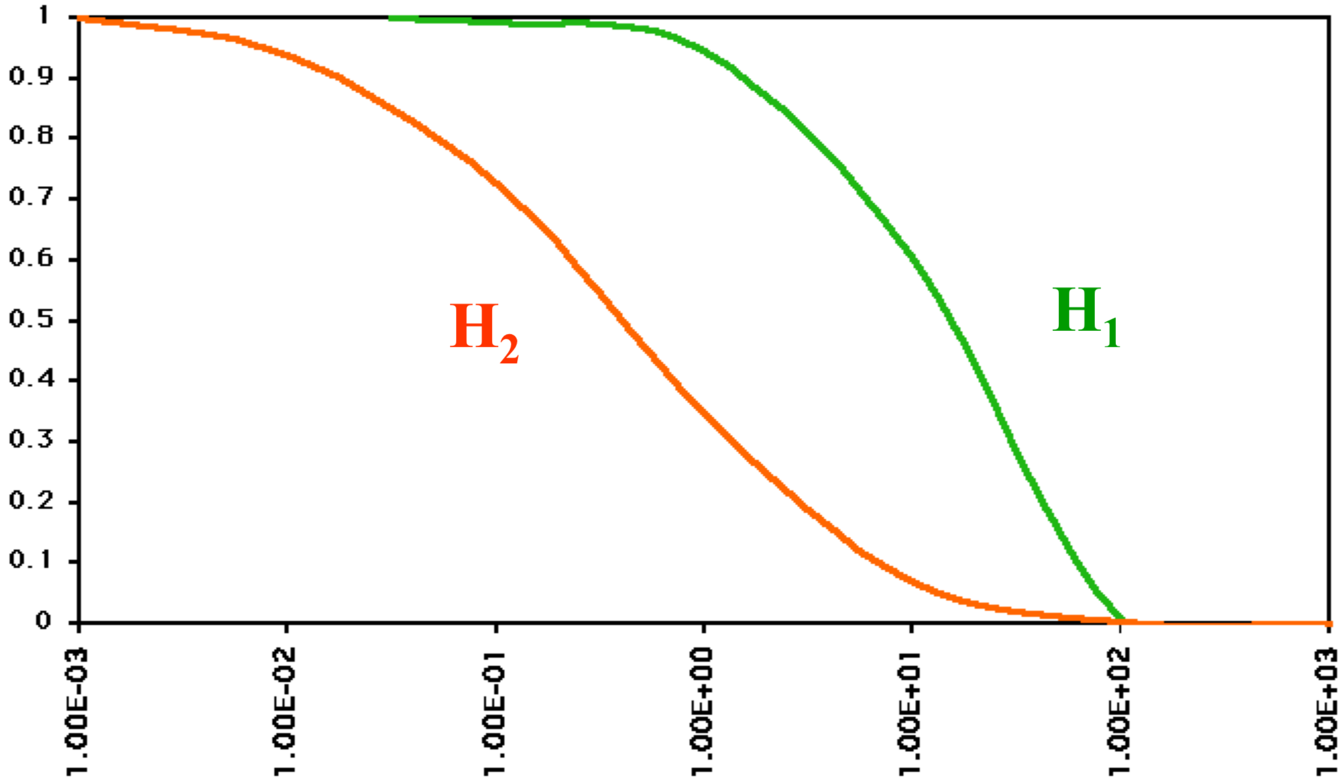
Biometrics: Identification of living persons by their traits in “real-time”

Forensics: Use of “trace evidence” from the crime scene to identify objects or persons

Biometrics and Forensic Science

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

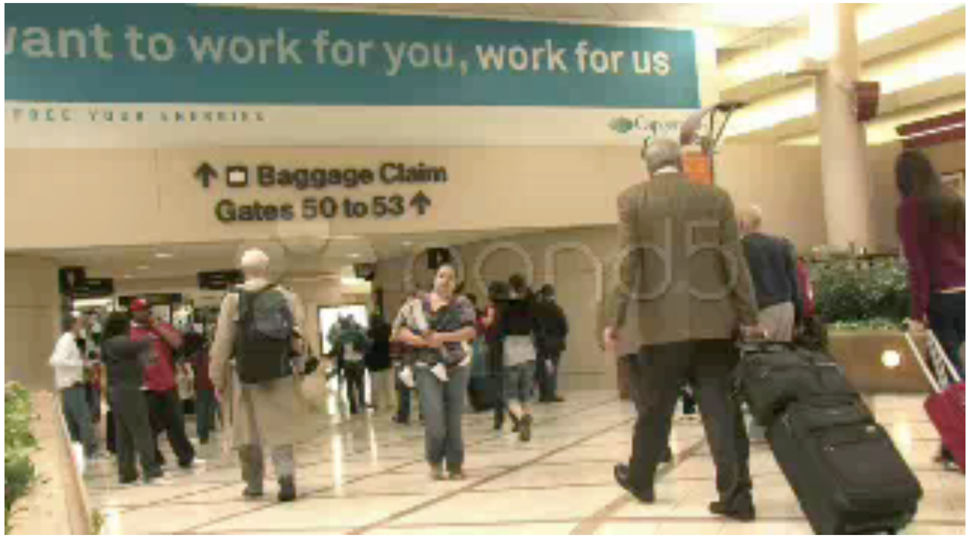
Vertical axis: Indication of the probability that the result of the experiment exceeds a given value of likelihood ratio



Horizontal axis: Graduation with increasing values of likelihood ratios

Working environments

Working environments



Videos & Sketches

First Composite Sketch Created in 1987

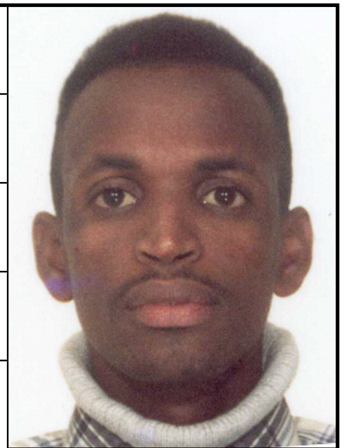
Later Sketch which became widely circulated

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

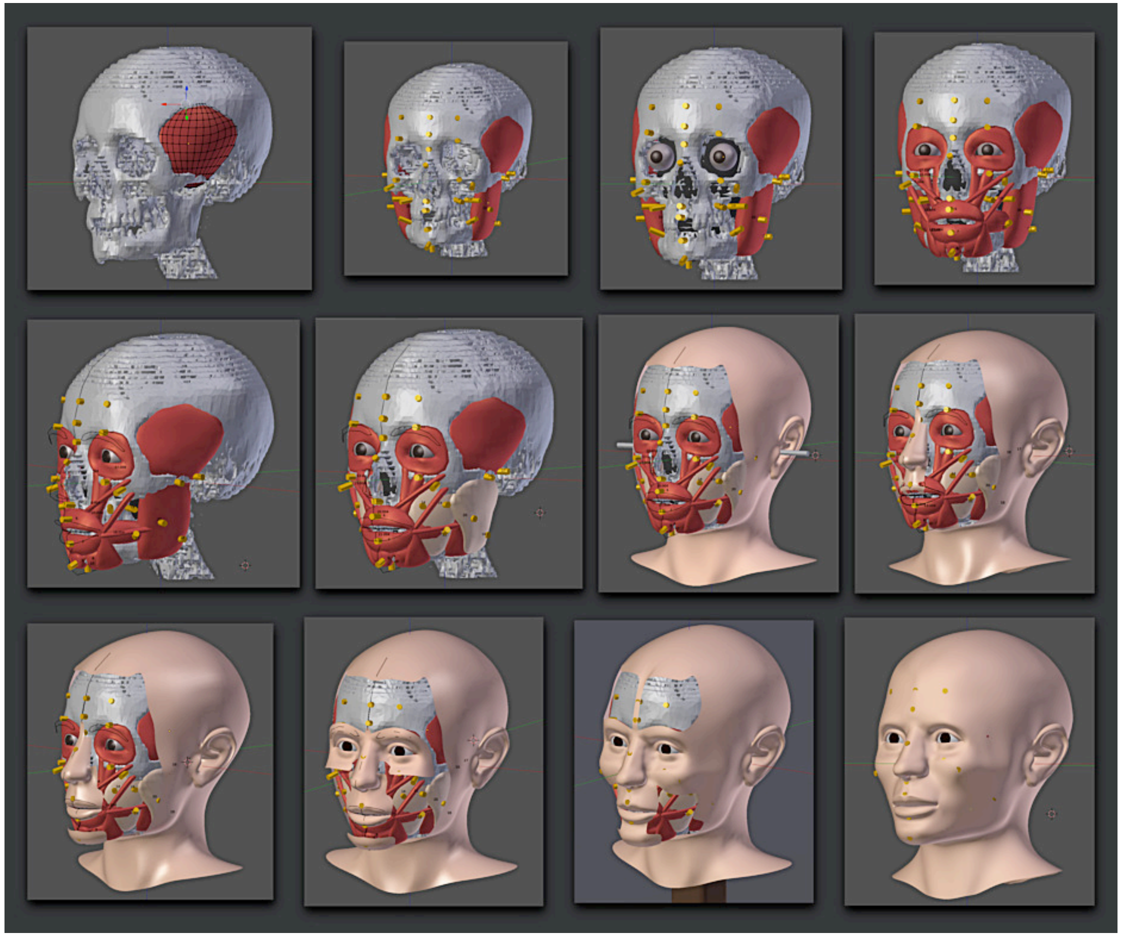
Feature	Details	Similarities and differences					
		S	NO	D	SI	Explanation for differences: Assumption of same source	Explanation for differences: Assumption of different source
Face	Shape	X					
	Proportions	X					
	Hairline Hairgrowth			X		Age, hairdo	
Forehead	Shape	X					
	Bumps	X					
	Horizontal 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 recognition

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

II. (PRESENT) What can we learn?

III. (FUTURE) What is still to be done?

Future of face recognition

… Who knows?

Ask Apple…

or maybe Google

…if you can

Future of face recognition

… Go to ICB, BTAS or next Biometrics School….

Future of face recognition

Vision Lab

- **Age - Pose - Illumination - Expression**
  - APIE-invariant representations: more dimensions & familiarity
- **Mobile Applications in the visible domain**
  - Shading can be as good as shape: iPhone X & embedded
- **Exploit more qualitative information**
  - From recognition to characterisation
- **Spoofing and camouflage**
  - Exploit additional information not only related to the face
- **3D shape and texture**
  - Not just fusion, more cooperative strategies
- **Exploit fine details**
  - Learn from forensic… don’t indulge too much with CNNs…

Vision Lab

THANK YOU
FOR YOUR ATTENTION

…AND PATIENCE