Face Recognition in the Wild: Challenges and Perspectives

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

1

Credits

- From the laboratory staff:

  Linda Brodo
  Marinella Cadoni
  Filippo Casu
  Massimo Gessa
  Enrico Grosso
  Souad Khellat Khiel
  Andrea Lagorio
  Ludovica Lorusso
  Gianluca Masala
  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)

2

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

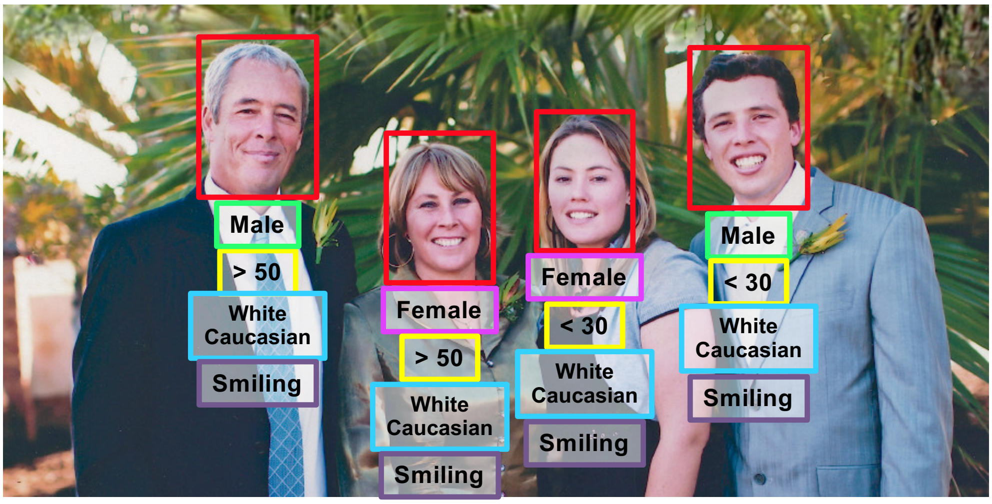
Multi-Foresee

IDENTITY

Computer Vision Enabled Multimedia Forensics and People Identification

Biometric Face Analysis

Vision Lab

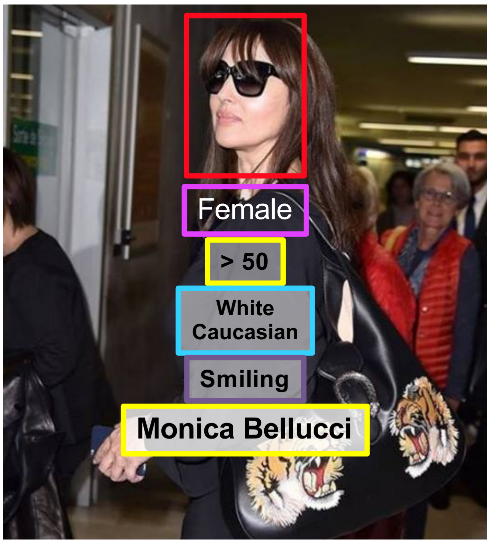


Information from a **single** face image

Biometric Face Analysis

Vision Lab

Monica Bellucci



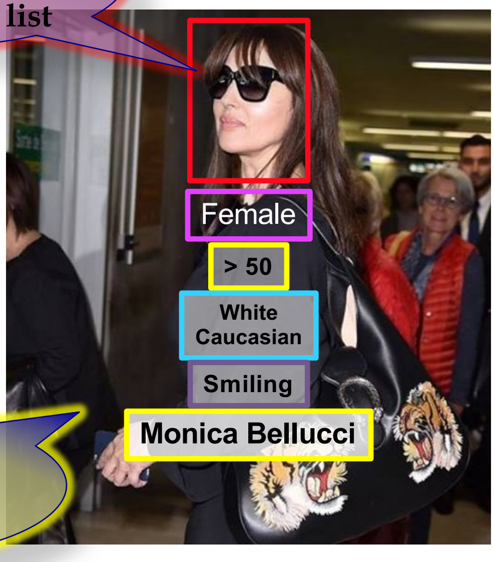
Information from **two** face images

Biometric Face Analysis

Vision Lab

Not found in Shenzhen's terrorist black list

Monica Bellucci



Not found among Shenzhen University staff

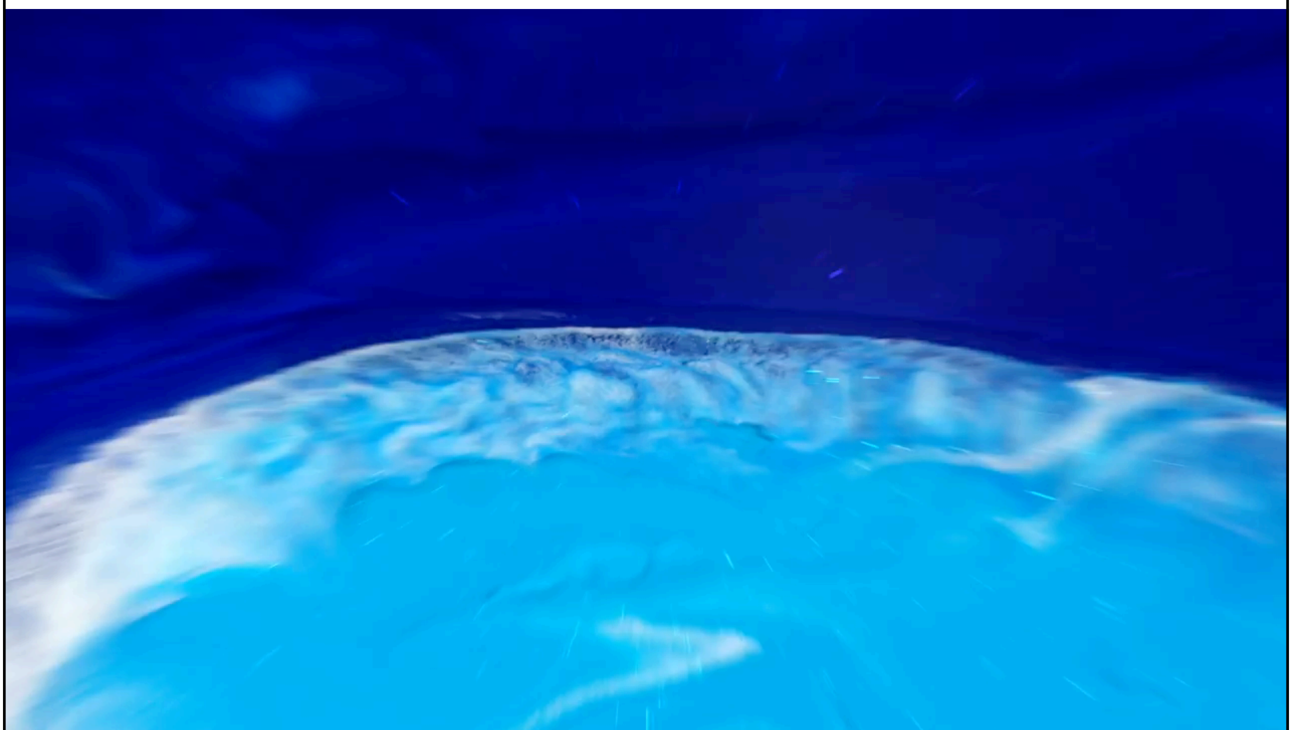
Information from **many** face images

CSI Fiction

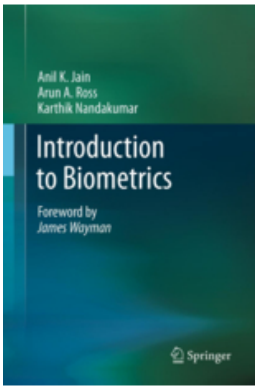
Vision Lab

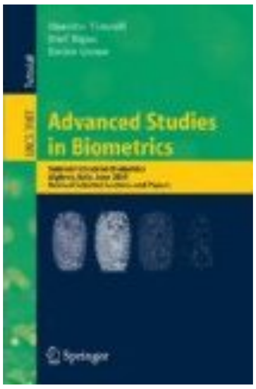
Media Fiction

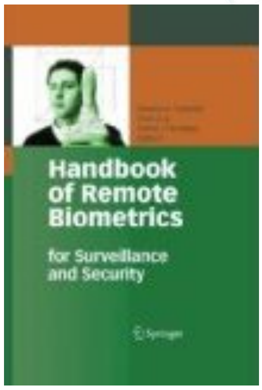
Vision Lab

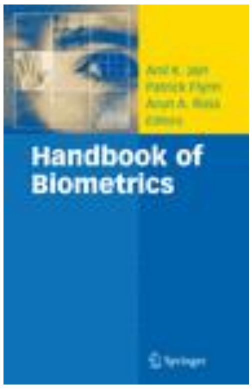


Textbooks

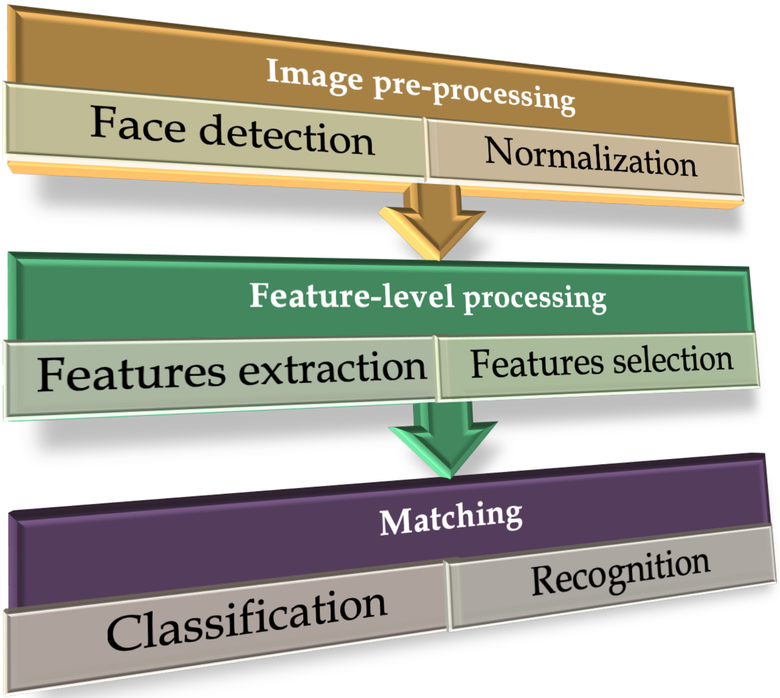








Elements of Face Biometrics



Match scores

Vision Lab

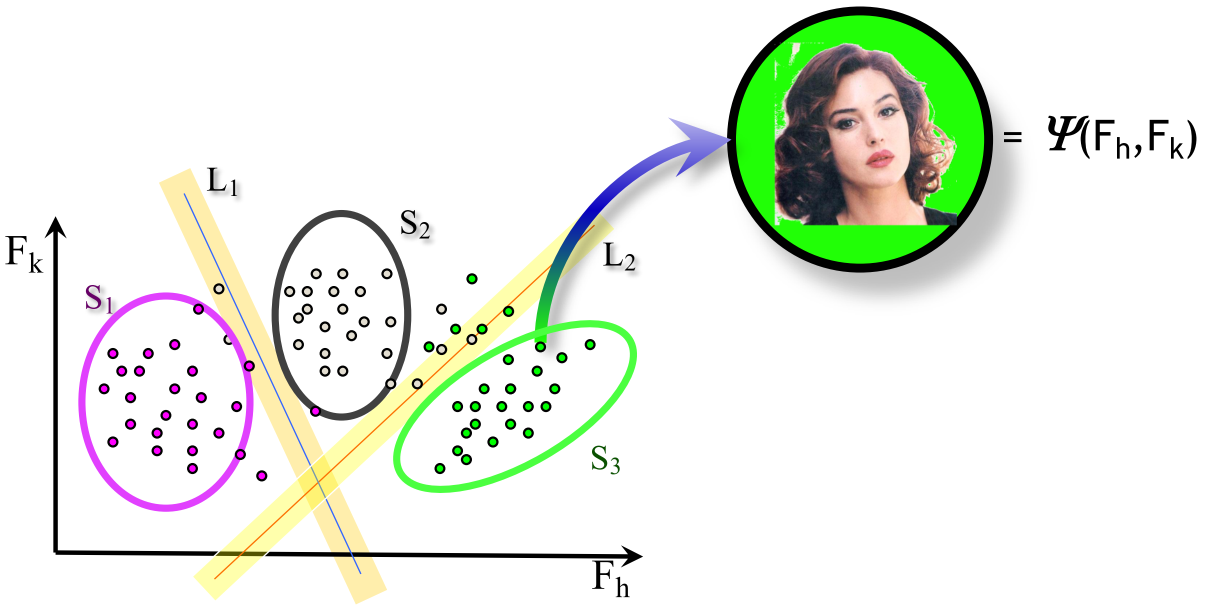
The output of a biometric matcher is the **match score**, (typically a single number), that quantifies the **similarity** between a reference sample (***probe image***) and the newly acquired sample (***gallery image***)

The **higher** the *similarity* score, the **more certain** is the system that the two biometric measurements come from the same person

Face Identification

Vision Lab

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



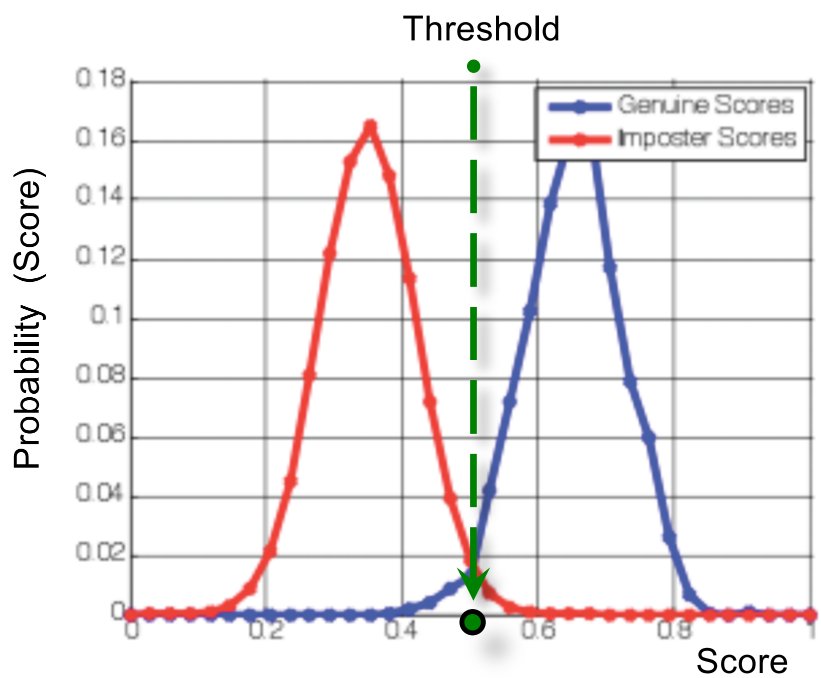
Genuine and Impostor scores

- **Genuine score**: Match score obtained when two biometric samples from the **same** individual are compared
- **Impostor score**: Match score obtained when two biometric samples originating from **different** individuals are compared

  Therefore, **a genuine user score should be greater than an impostor score**

- A **threshold** (or **classifier**) is used to determine if a score is related to a genuine user or an impostor

Match score distributions



Inter-class *similarity*

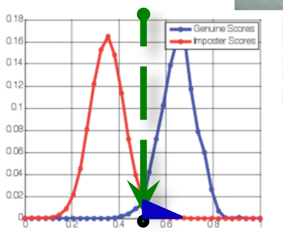
Vision Lab

Two different people with very similar appearance

FALSE MATCH

www.marykateandashley.com

Twins



/english/in_depth/americas/2000/us_elections

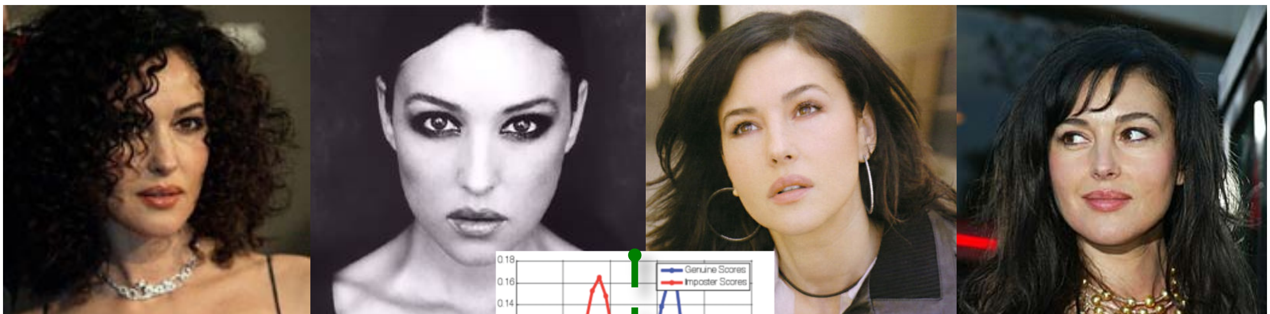
Father and son

Intra-class *variability*

Vision Lab

The same person with very different biometric samples

FALSE NON MATCH



Monica Bellucci

Biometric uniqueness

Vision Lab

Biometric traits develop:

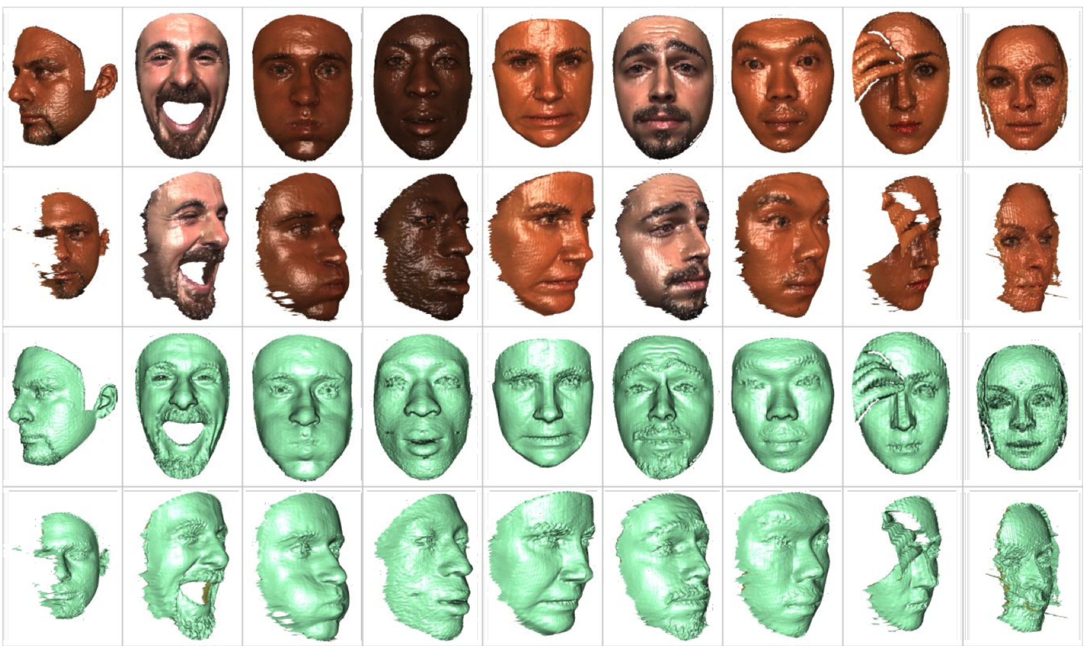
1. through genetics: **Genotypic**
2. through random variations in the early phases of an embryo's development: **Phenotypic**
3. through training: **Behavioral**

Biometric Trait	genotypic	phenotypic	behavioral
Fingerprint (minutiae)	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 back 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

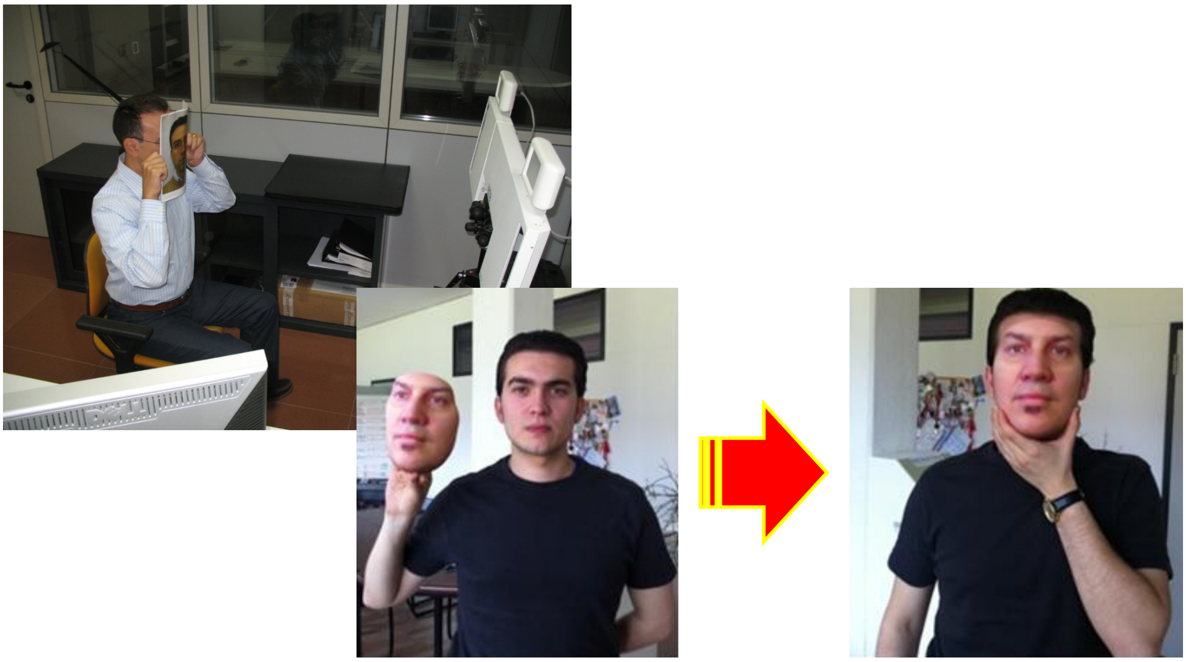
Face shape and texture

Vision Lab



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.

Visual challenges



Visual challenges

Vision Lab

Visual challenges

Vision Lab

Visual challenges

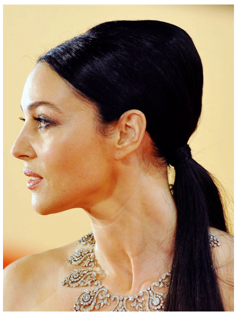
Vision Lab

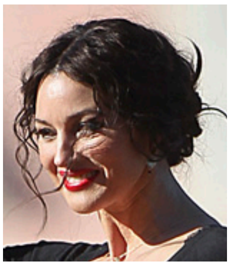
A – Aging

P – Pose

I – Illumination

E - Expression



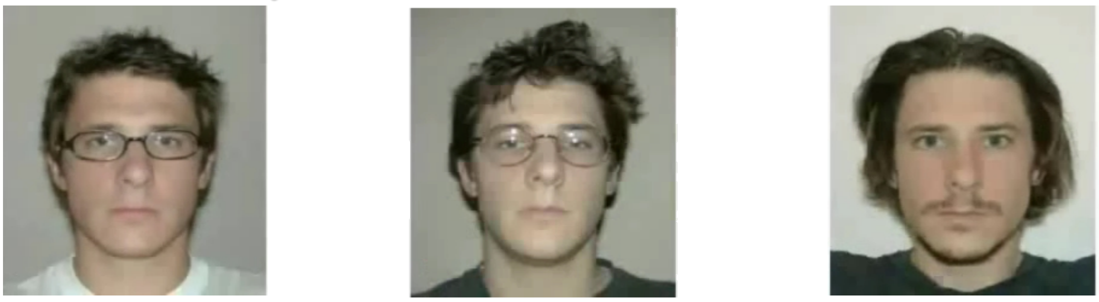


DEALING WITH AGE PROGRESSION

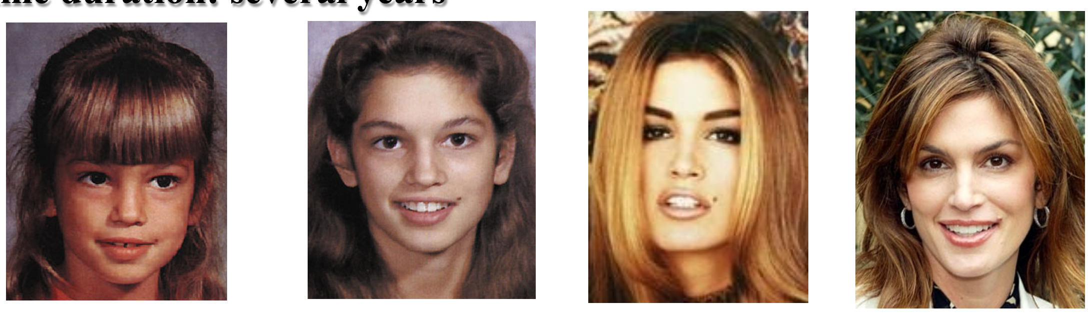
M. Ortega, L. Brodo, M. Bicego, M. Tistarelli, "Measuring changes in face appearance through aging,", Proc. of IEEE Computer Vision and Pattern Recognition Workshop, pp. 107-113, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (2009)

Aging effects

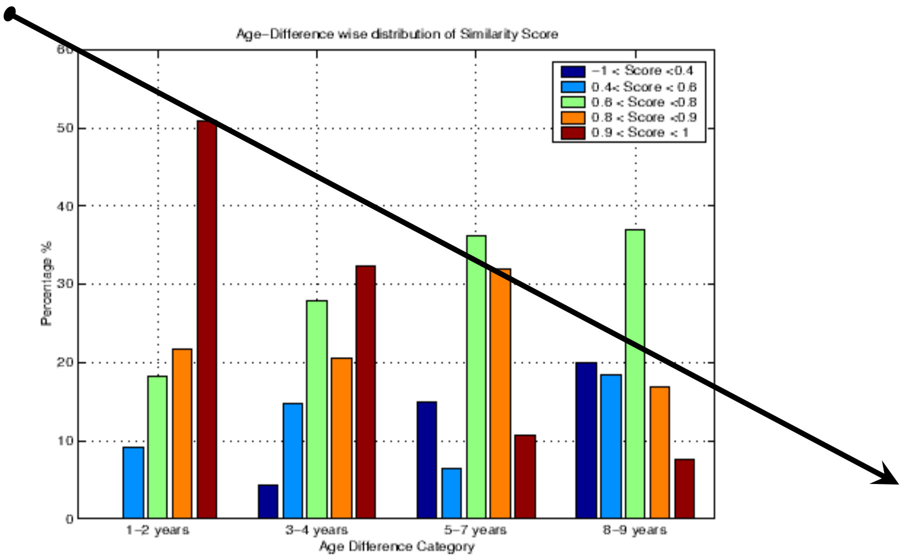
Time duration: 2 years



Time duration: several years

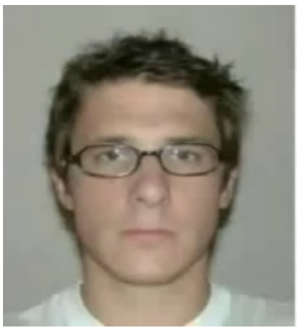


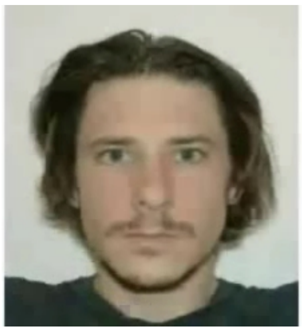
Effects of age progression



Similarity scores decrease as age difference increases

Aging … over time





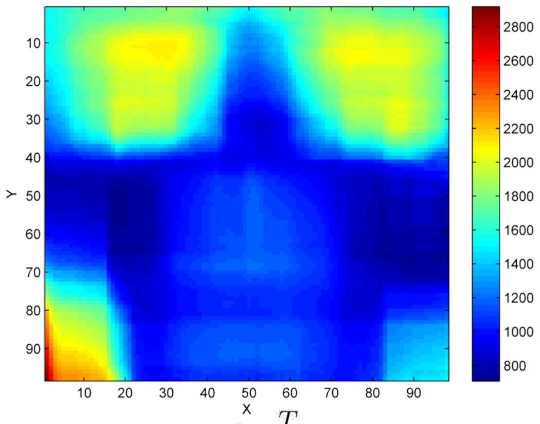
$$d_{I_1,I_2}(x,y) = \frac{1}{2}\left(\frac{1}{|P_{I_1}|}\sum_{p\in P_{I_1}}\omega(p) + \frac{1}{|P_{I_2}|}\sum_{q\in P_{I_2}}\omega(q)\right)$$

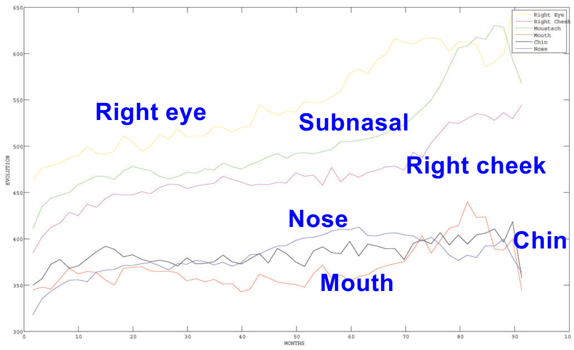
M. Ortega, L. Brodo, M. Bicego, M. Tistarelli, "Measuring changes in face appearance through aging,", Proc. of IEEE Computer Vision and Pattern Recognition Workshop, pp. 107-113, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (2009)

Photometric effects

Vision Lab

Time evolution of facial features over 4 years





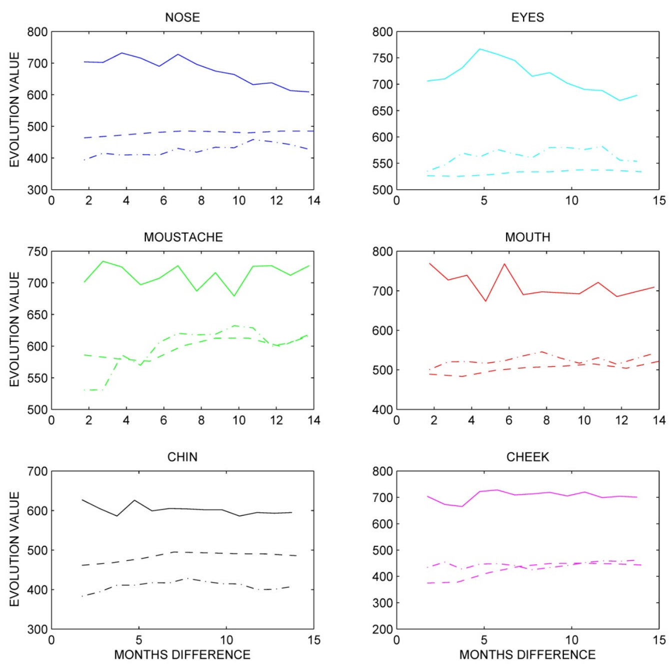
$$E(x,y,t) = \frac{1}{T}\sum_{i=1}^{T} d_{I_i, I_{i+t}}(x,y) \qquad E(L,t) = \frac{1}{w_L h_L}\sum_{(x,y)\in L} E(x,y,t)$$

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)

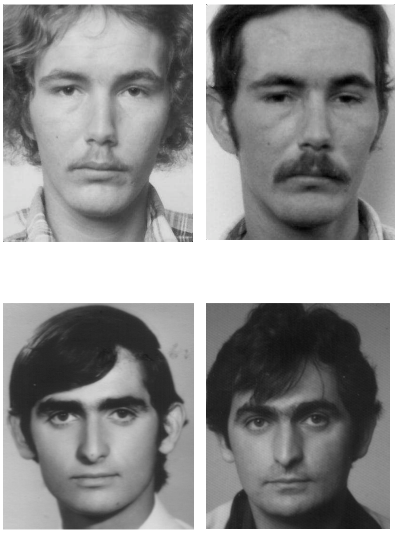
Photometric effects

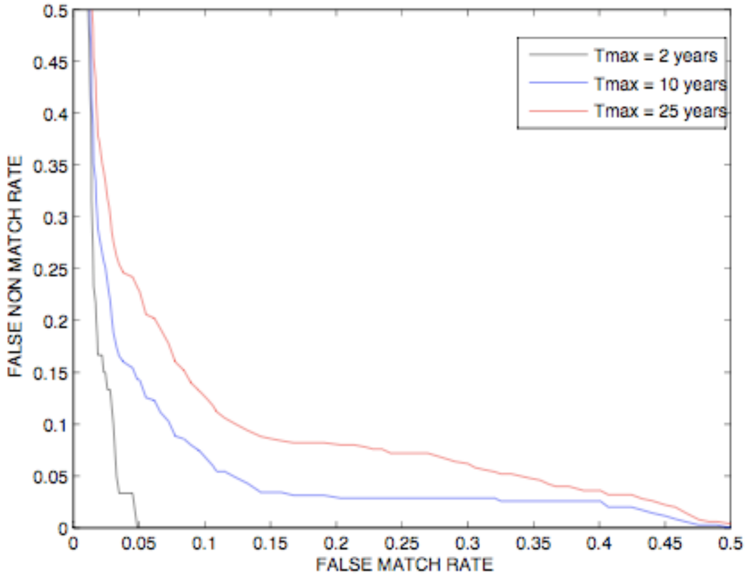
Vision Lab

Comparative time evolution of features for different subjects



Face matching across age





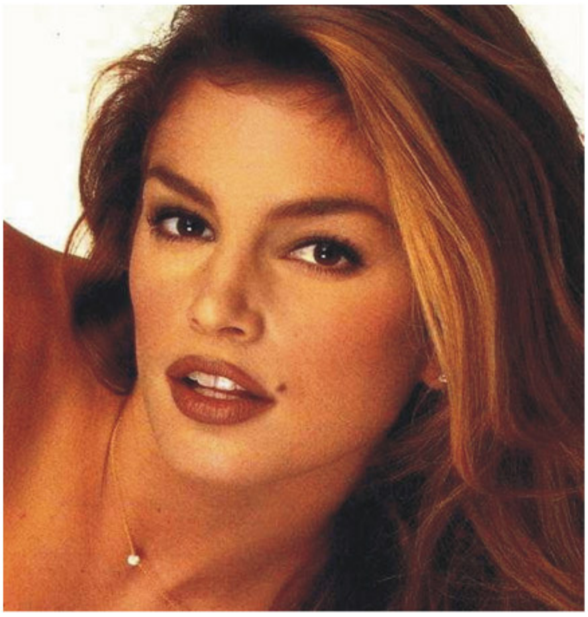
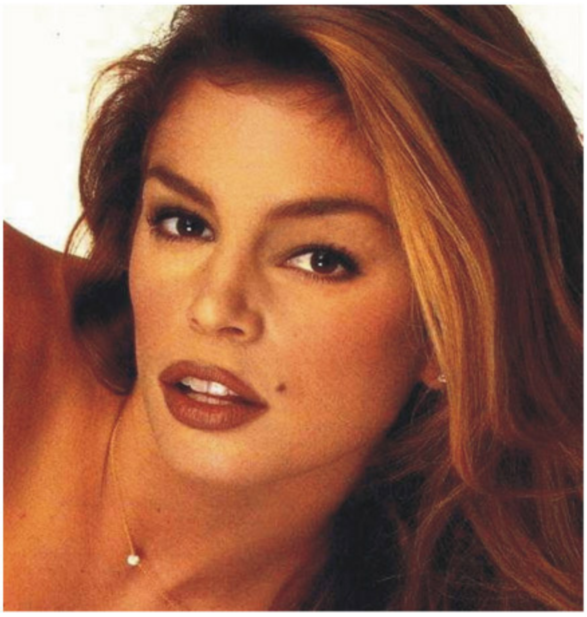
M. Ortega, L. Brodo, M. Bicego, M. Tistarelli, "Measuring changes in face appearance through aging,", Proc. of IEEE Computer Vision and Pattern Recognition Workshop, pp. 107-113, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (2009)

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

Vision Lab

… But the data is still there

Face alignment

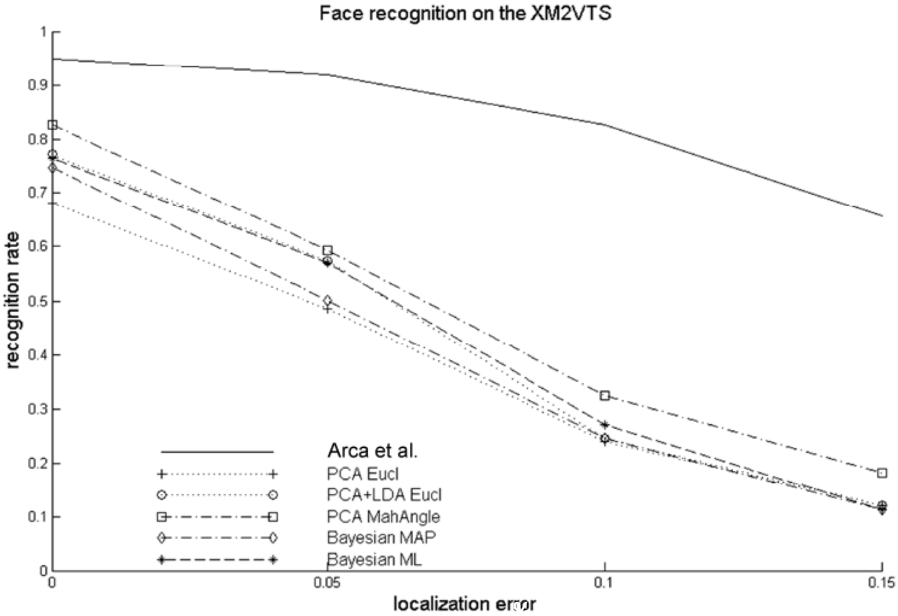
Vision Lab

In cognitive psychology it is called

perceptual organization

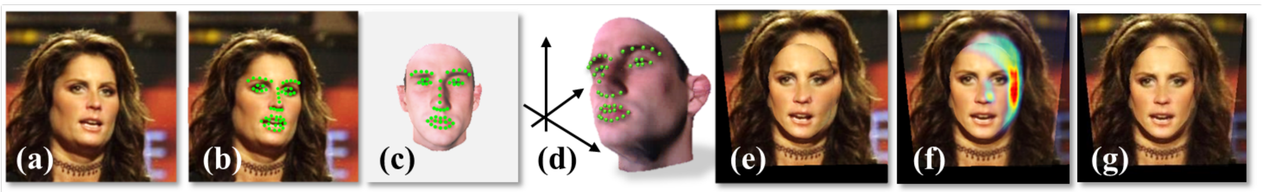


Face alignment



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.

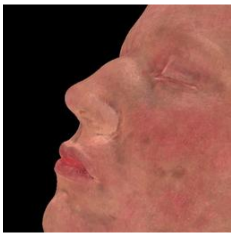
Tal Hassner, Shai Harel, Eran Paz, Roee Enbar; "Effective Face Frontalization in Unconstrained Images" The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4295-4304

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.

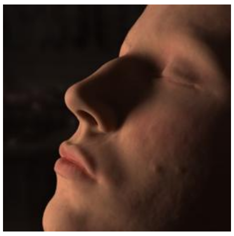
Modelling the face skin



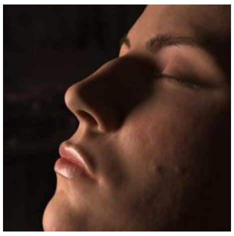
Skin chromaticity map

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/)

Illumination compensation

Sample techniques:

- **Histogram-based adaptive equalization** (applied on image patches)
- **Image Re-lighting**
- **Illumination-invariant representations** (such as the *Hue* component in color space)

Image re-lighting

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

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

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.

An ill-posed problem

Vision Lab

An inverse problem is *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

Vision Lab

Two adverse conditions:

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

Solution: Regularization

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), 501–4.

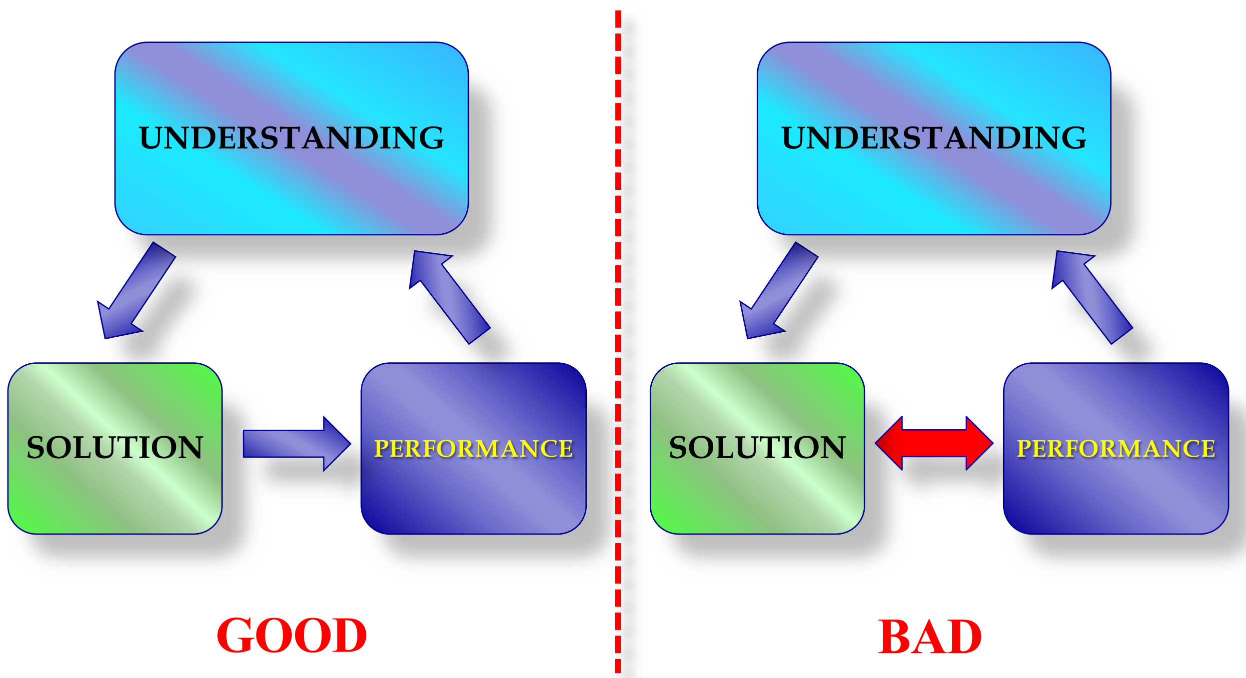
A.N. Tikhonov, "**On the regularization of ill-posed problems**". Dokl. Akad. Nauk SSSR 153(1) (1963), 49–52 (in Russian).

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

FACE RECOGNITION RESEARCH AND TECHNOLOGIES

41

Good research or bad research?



42

Common mistakes

Vision Lab

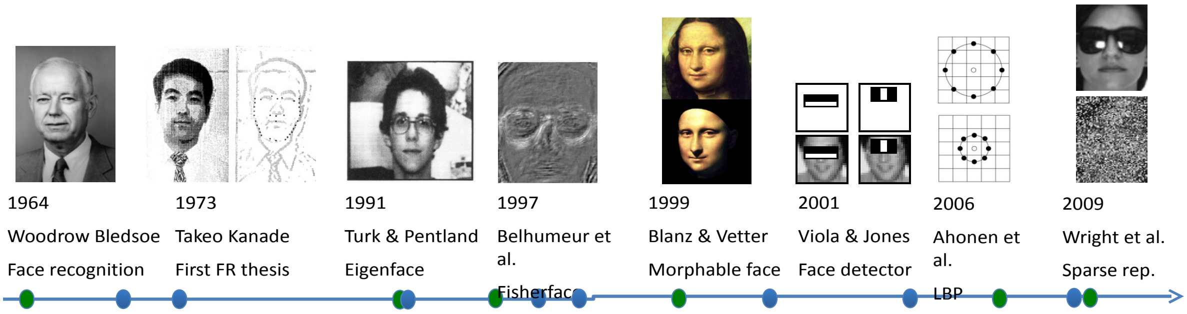
1. Start **programming** before **thinking**;
2. Building a system **blindly** combining a number of already available algorithms;
3. Performing **blind tests** with available tools;
4. Performing **blind tests** on available datasets;
5. Twickling the **parameters** until you obtain the **desired performance**;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

Addressing the problem

Vision Lab

1. Analyze the **problem**, the available **data** and the **constraints**;
2. Make a **bibliographical search** (don't try to re-invent the wheel);
3. Define a **model** describing the **physics** of the **event**;
4. Find a **mathematical framework** which may bring to 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 algorithms at the state of the art.

Face recognition milestones

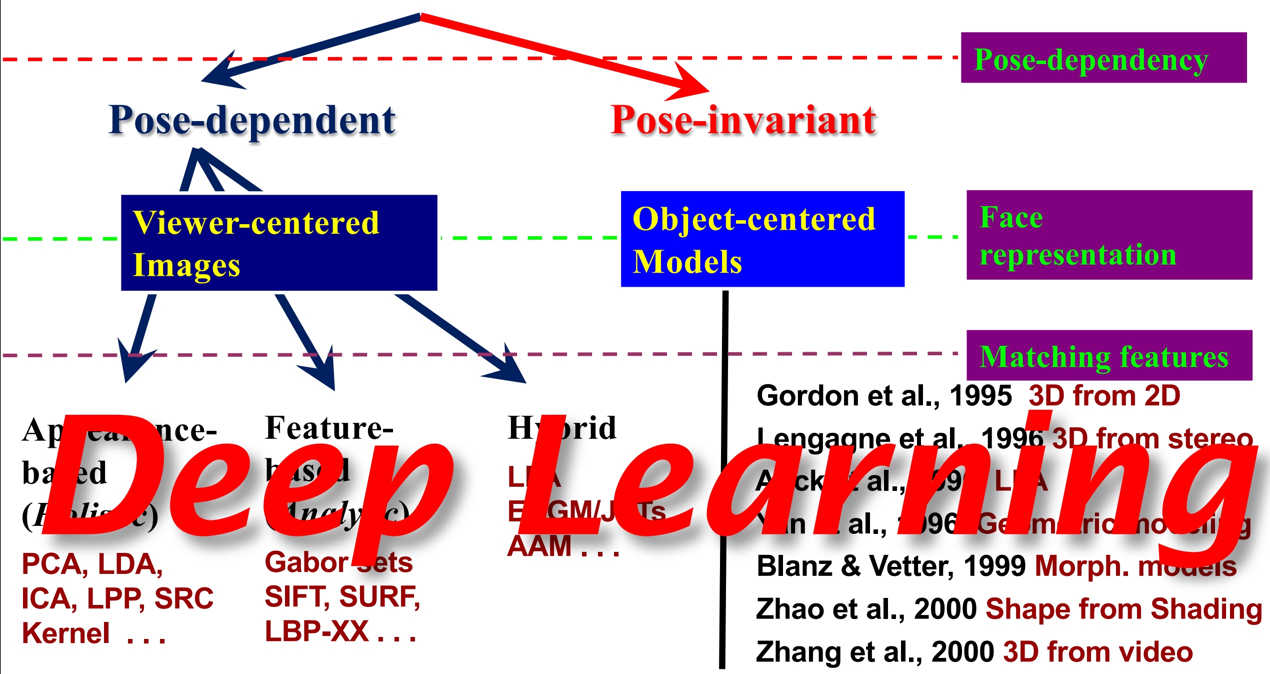


1964	1973	1991	1997	1999	2001	2006	2009
Woodrow Bledsoe	Takeo Kanade	Turk & Pentland	Belhumeur et al.	Blanz & Vetter	Viola & Jones	Ahonen et al.	Wright et al.
Face recognition	First FR thesis	Eigenface	Fisherface	Morphable face	Face detector	LBP	Sparse rep.

1915	1991	1990s	2000	2010-2013	April 2013
35mm still camera	Kodak	Surveillance camera	Sharp	Wearable camera	Samsung Galaxy S4
	Digital camera	480p @ 30fps	First camera phone	480P @ 30fps	1080p @ 30fps
	1024p		320p	Google Glass	
				720p @30fps	

Bledsoe, W. W. 1964. The Model Method in Facial Recognition, TR PRI 15, Panoramic Research, Inc., California.
Takeo Kanade, Picture Processing System by Computer Complex and Recognition of Human Faces, Kyoto Univ., 1973.
M. Turk and A. Pentland, Eigenfaces for recognition. Journal of Cognitive Neuroscience 3 (1): 71–86, 1991.
Belhumeur, P.N. et al., Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, PAMI, 19-7, 1997.
V. Blanz and T. Vetter, A morphable model for the synthesis of 3D faces, SIGGRAPH 1999.
http://photodoto.com/camera-history-timeline/
Viola, Jones: Robust Real-time Object Detection, IJCV 2001.
Ahonen, et al. Face Description with Local Binary Patterns: Application to Face Recognition, PAMI, 2006.
J. Wright et al. Robust Face Recognition via Sparse Representation, PAMI, 31-2, 2009.
http://static7.businessinsider.com/image/4d013ea7cadcbb7033010000/looxcie-video-camera.jpg

Taxonomy of face recognition



Holistic face recognition

- The basic idea 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***:

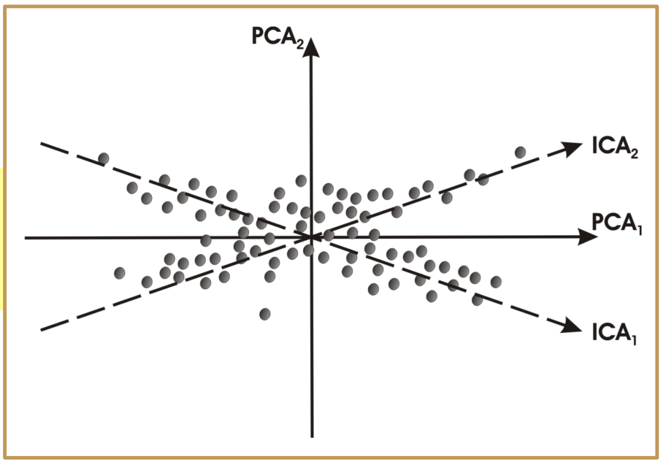
$$A = U \cdot \Sigma \cdot W$$

- 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

ICA

Orthonormality
vs
Maximal variance

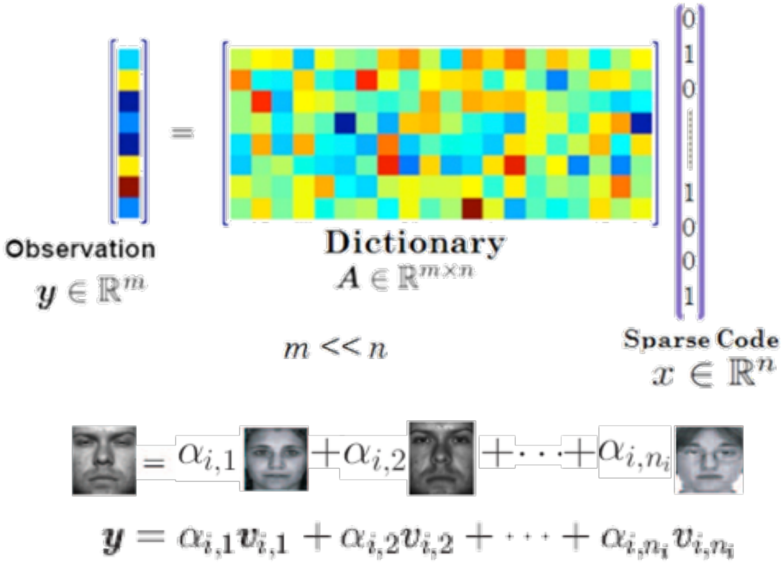


PCA

Sparse Representations

John Wright et al. PAMI 2009

1. Represent the test face as a **sparse linear combination** of the training faces.
2. Estimate the class of the test image from the **sparse coefficients**.



Dictionary matrix A

Holistic face recognition

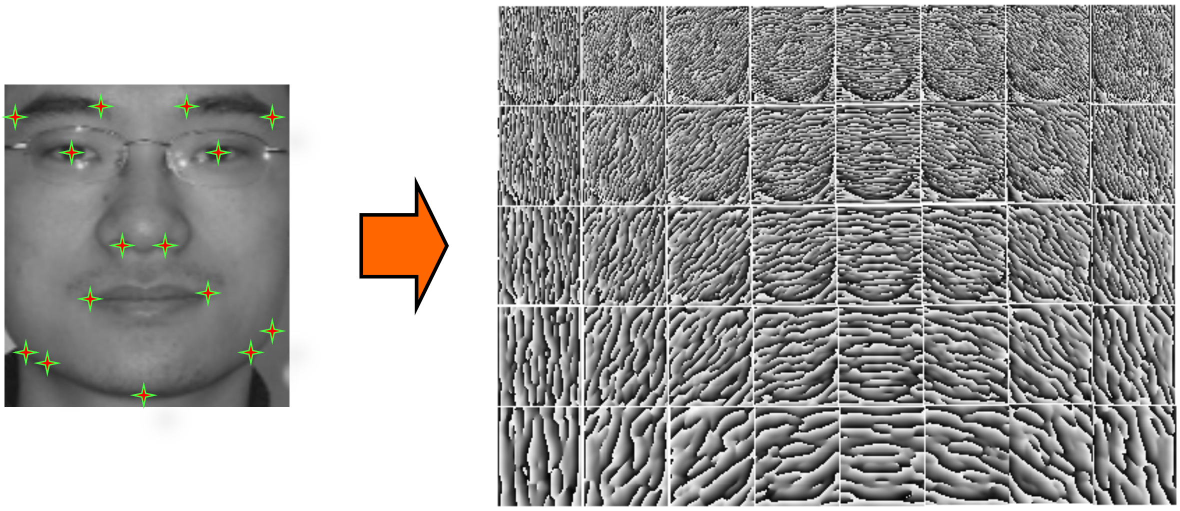
Sparse Representation *vs* Principal Component Analysis

Similar formulation, different objectives

Facial features

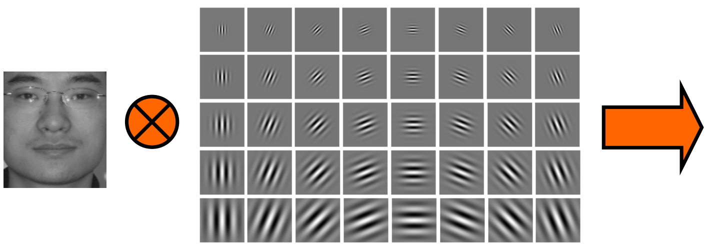
Physical **Landmarks**

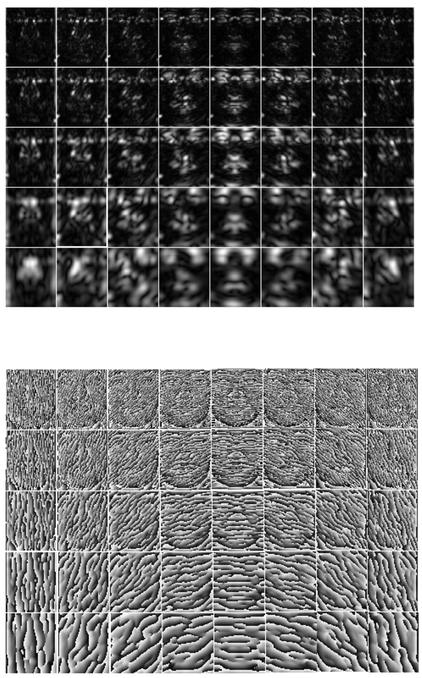
Gray level **oriented patterns/ photometric properties**



Gabor wavelets

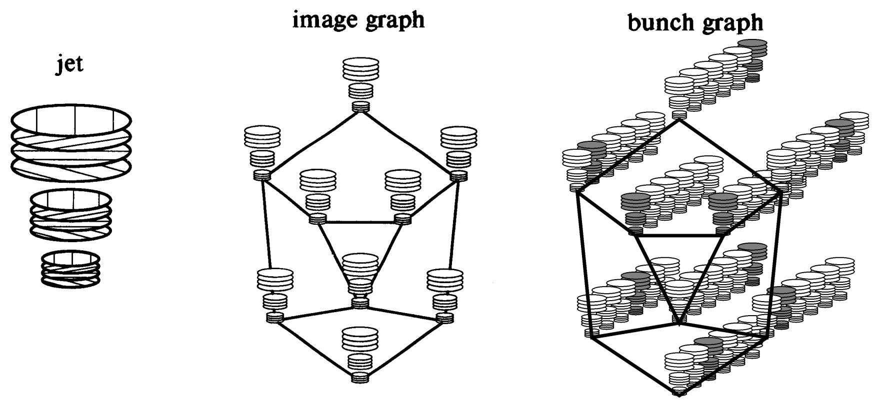
- Produce a complex vector description of the local structure of the gray-level patterns

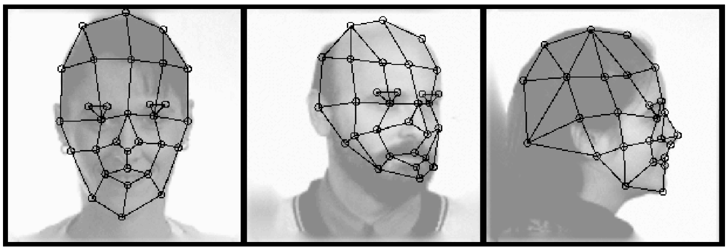




- Convolution with a bank of frequency-tuned complex filters

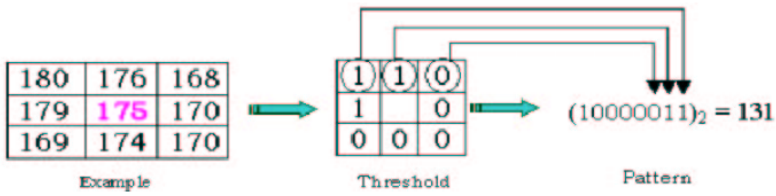
Gabor Jets

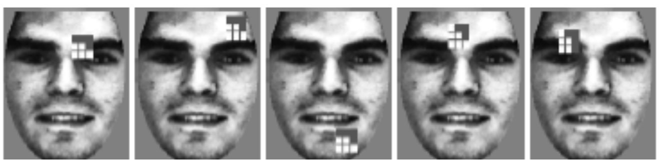




L. Wiskott, J-M Fellous, N. Krüger, and C.von der Malsburg “Face Recognition by Elastic Bunch Graph Matching“. IEEE Transactions on PAMI 19(7):775-779, July 1997.

Local Binary Patterns





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

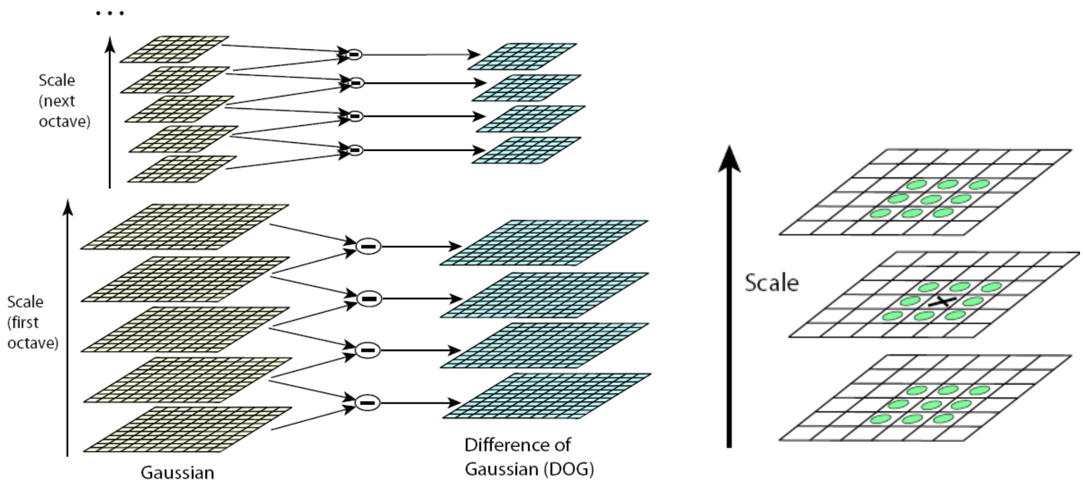
The **histogram of the labels** is used as a **texture descriptor**.

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

Scale Invariant Features

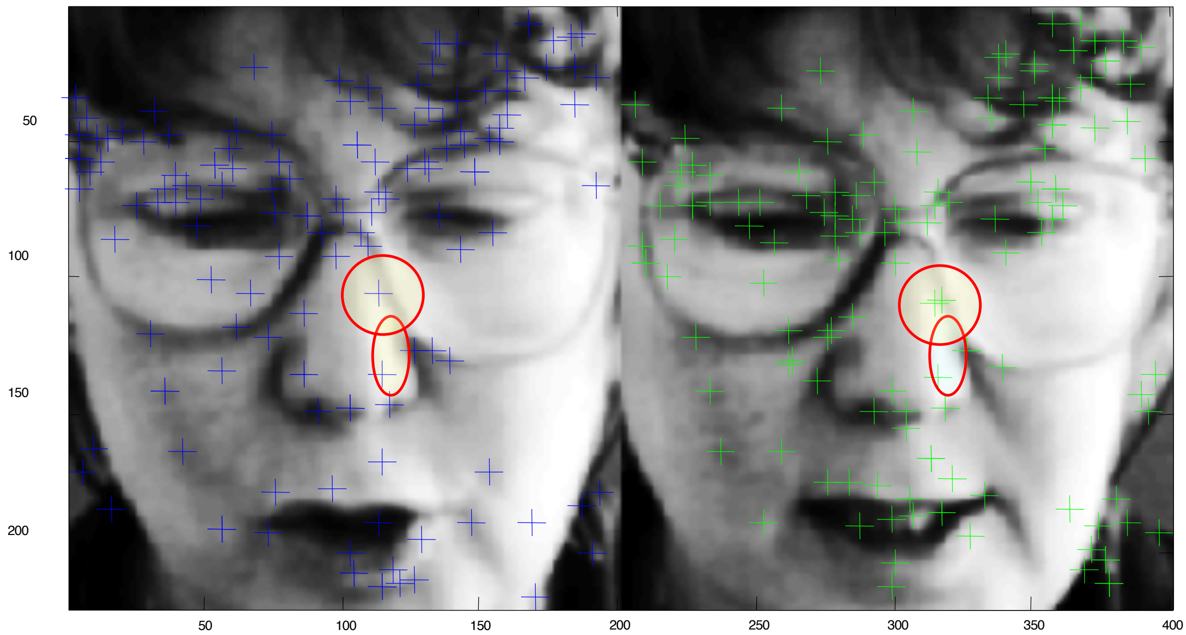
$$\mathbf{D}(x, y, \sigma, k) = (\mathbf{G}(x, y, k\sigma) - \mathbf{G}(x, y, \sigma)) * \mathbf{I}(x, y)$$

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



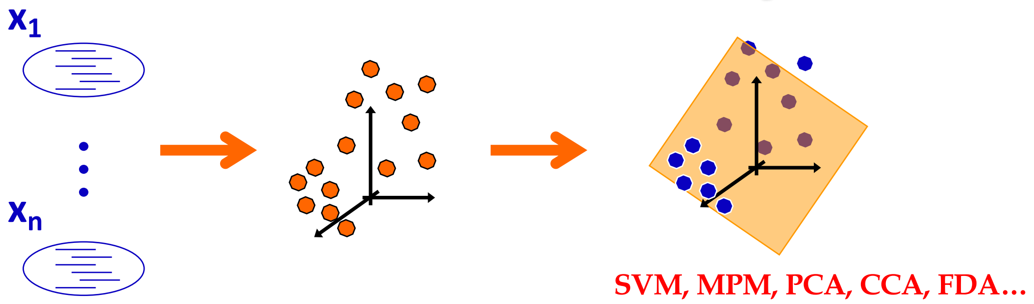
G. Lowe, "Object recognition from local scale invariant features", International Conference on Computer Vision , 1999.

Scale Invariant Features



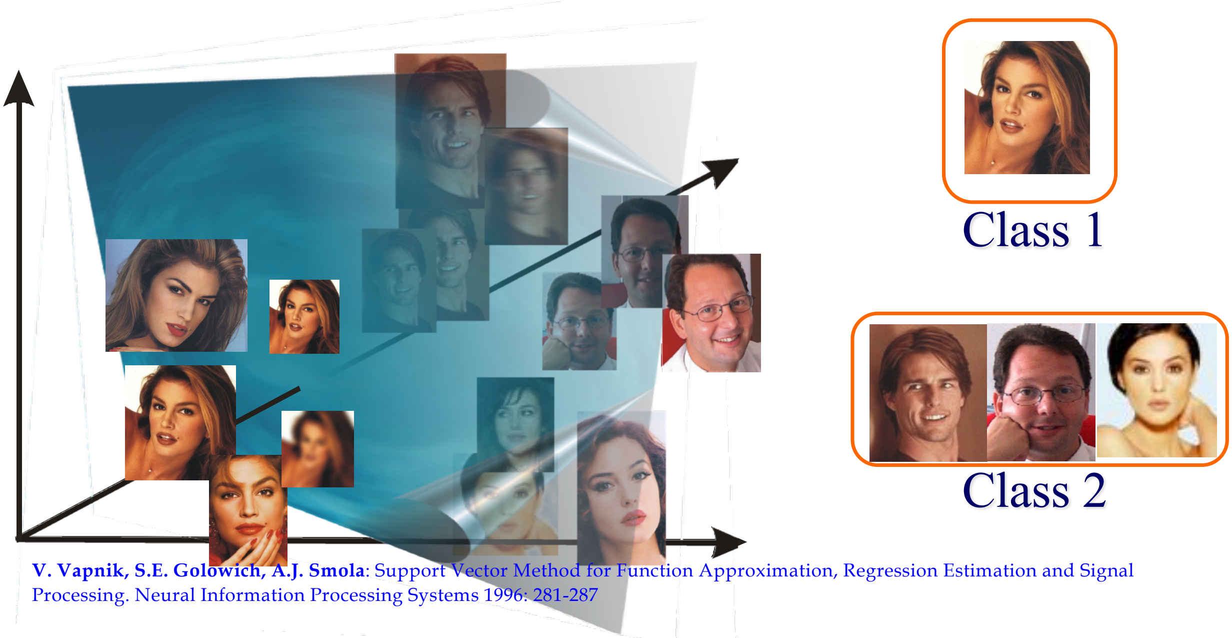
Kernel methods

- **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***



Support Vector Machines

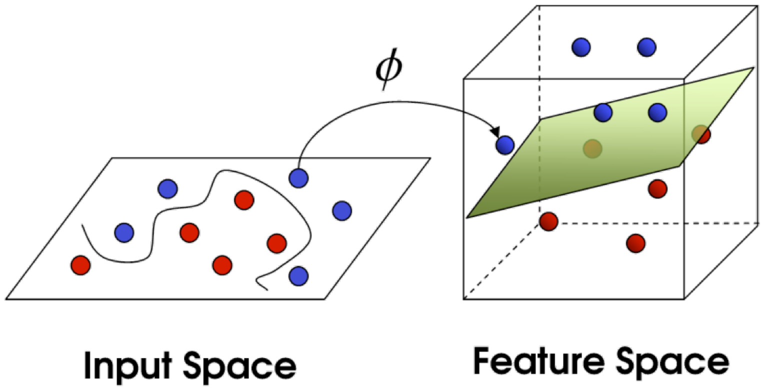
Support Vector Machines are binary classifiers



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

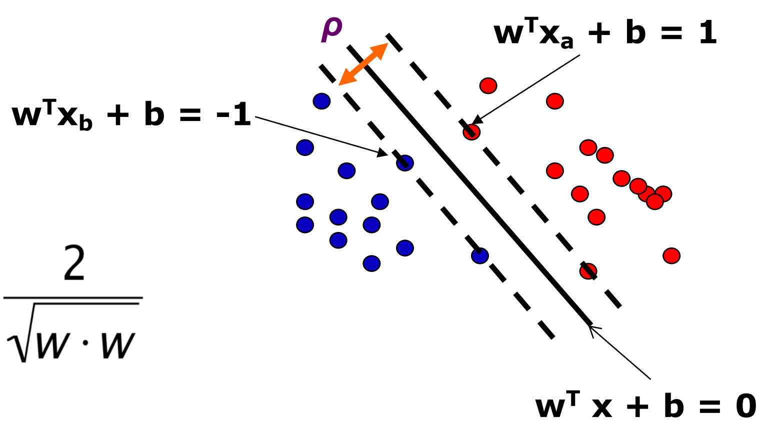
- Solves linearly separable problems
  1. **Data projection**: Input data are transformed
     mapping into higher dimensions



Support vectors

- Solves linearly separable problems
  2. **Training**: find optimal **hyperplane** $\mathbf{w}^T\mathbf{x}_i + \mathbf{b} = 0$
     margin maximisation $\min_{i=1,\ldots,n} |\mathbf{w}^T\mathbf{x}_i + \mathbf{b}| = 1$

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

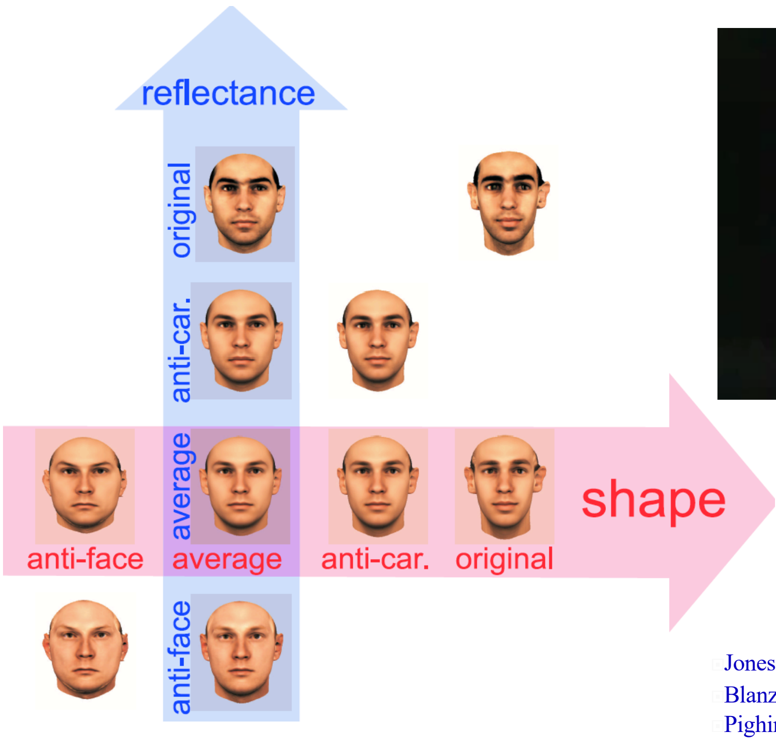


Parametric Morphable Models

3D shape

surface reflectance

Parametric Morphable Models

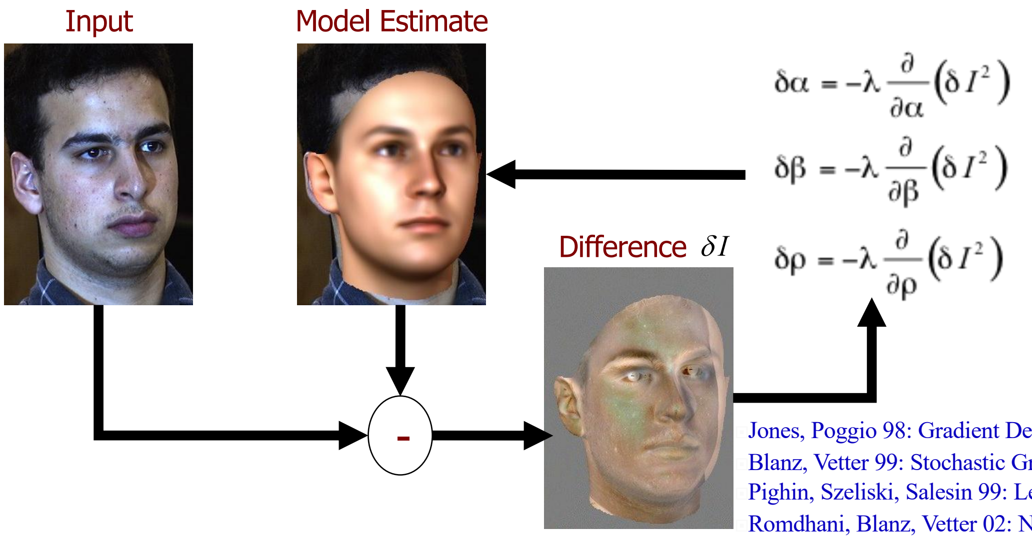


Jones, Poggio 98: Gradient Descent
Blanz, Vetter 99: Stochastic Gradient Descent
Pighin, Szeliski, Salesin 99: Levenberg-Marquardt
Romdhani, Blanz, Vetter 02: Non-linear fitting

Parametric Morphable Models

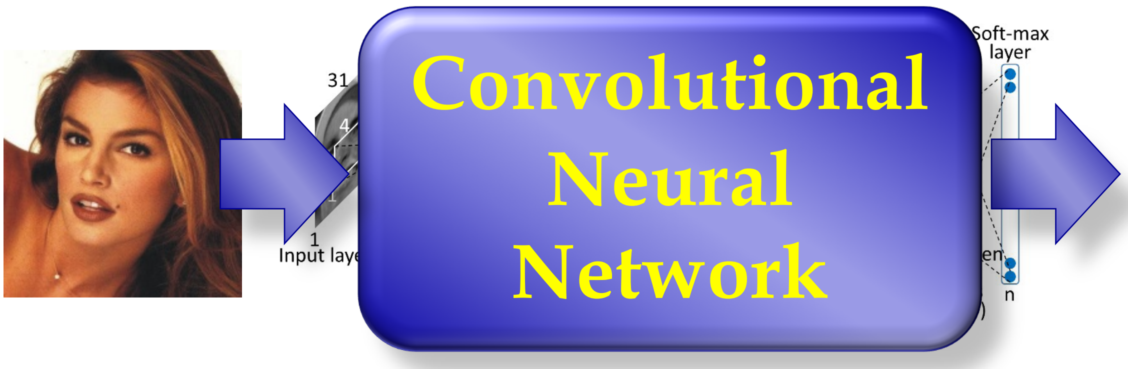
Statistically based analysis (PCA) of :

- laser scans - 3D shape (*x,y,z*) with reflectance (*r,g,b*) maps
- complete correspondence with average
- *Face = deformation from average dx, dy, dz, dr, dg, db*



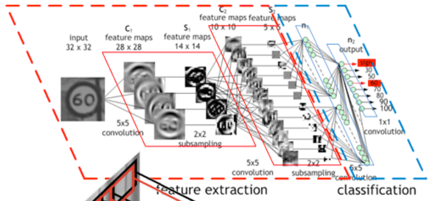
Jones, Poggio 98: Gradient Descent
Blanz, Vetter 99: Stochastic Gradient Descent
Pighin, Szeliski, Salesin 99: Levenberg-Marquardt
Romdhani, Blanz, Vetter 02: Non-linear fitting

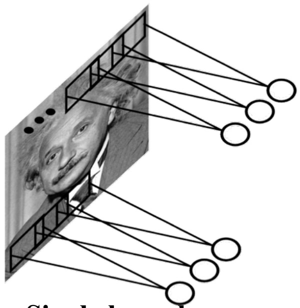
Convolutional Neural Networks



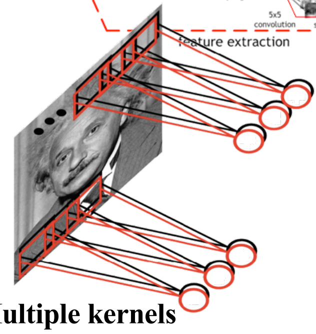
Cindy Crawford

Convolutional Neural Networks

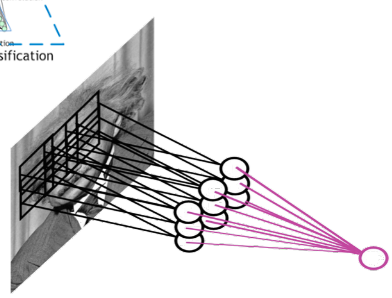




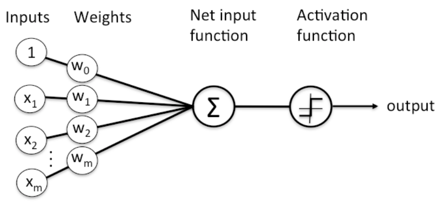
Single kernel Convolution



Multiple kernels Convolution



Spatial Pooling

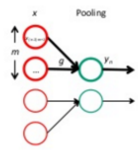


Let m be the size of pooling region, x be the input, and y be the output of the pooling layer. subsample(f, g)[n] denotes the n-th element of subsample(f, g).

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

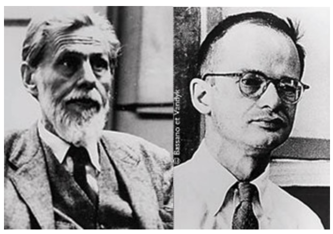
$$y = \text{subsample}(x, g) = [y_n]$$

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

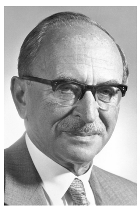


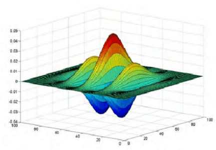
Why CNNs… today?

Neural networks have been proposed since the early '40s:
In **1943**, neurophysiologist **Warren McCulloch** and mathematician **Walter Pitts** wrote a paper on how neurons might work. **They modeled a simple neural network using electrical circuits**.



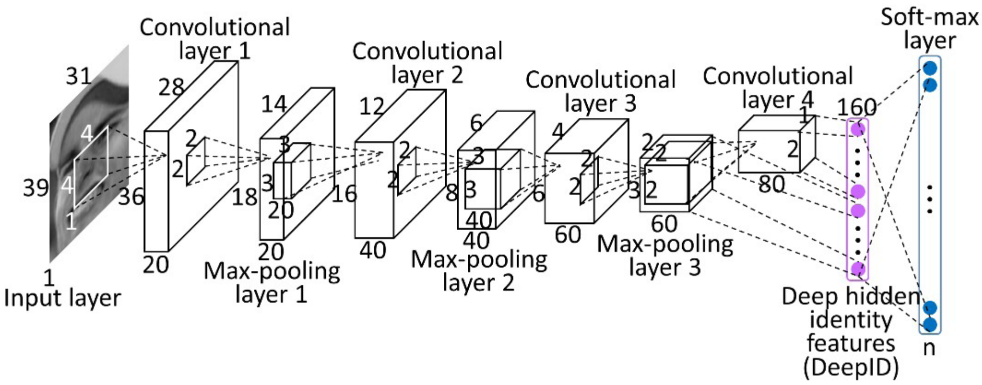
Convolutions or **digital filtering** have been used since the 50's for several vision tasks, including face recognition.





The progress in the **Theory of Learning** and of **computing power** allowed to implement more efficient and complex neural networks with multiple hidden layers…

Convolutional Neural Networks



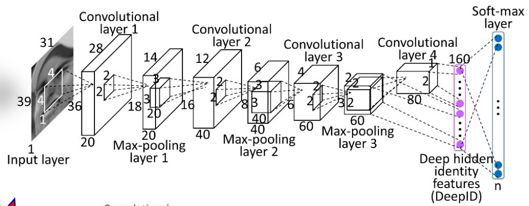
A deep CNN is used to extract a feature vector with relatively high dimension. The network can be supervised by multiclass loss and verification loss

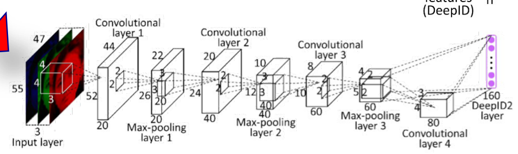
PCA, Joint Bayesian or metric-learning methods are used to learn a more efficient low dimensional representation

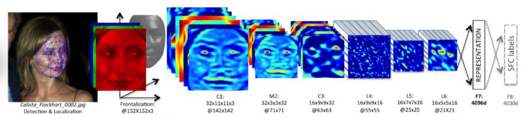
The amount of training data can range from **100K** up to **260M**

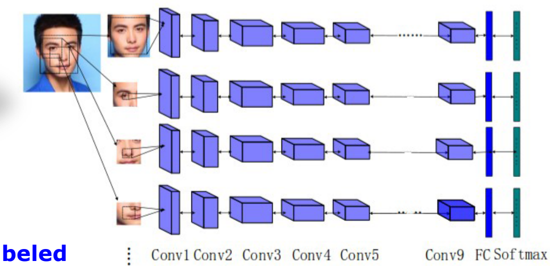
Convolutional Neural Networks

- 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
- VGG (M. Parkhi, A. Vedaldi, A. Zissermann - BMVC 2015)
- 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.

FACE RECOGNITION PERFORMANCE

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

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

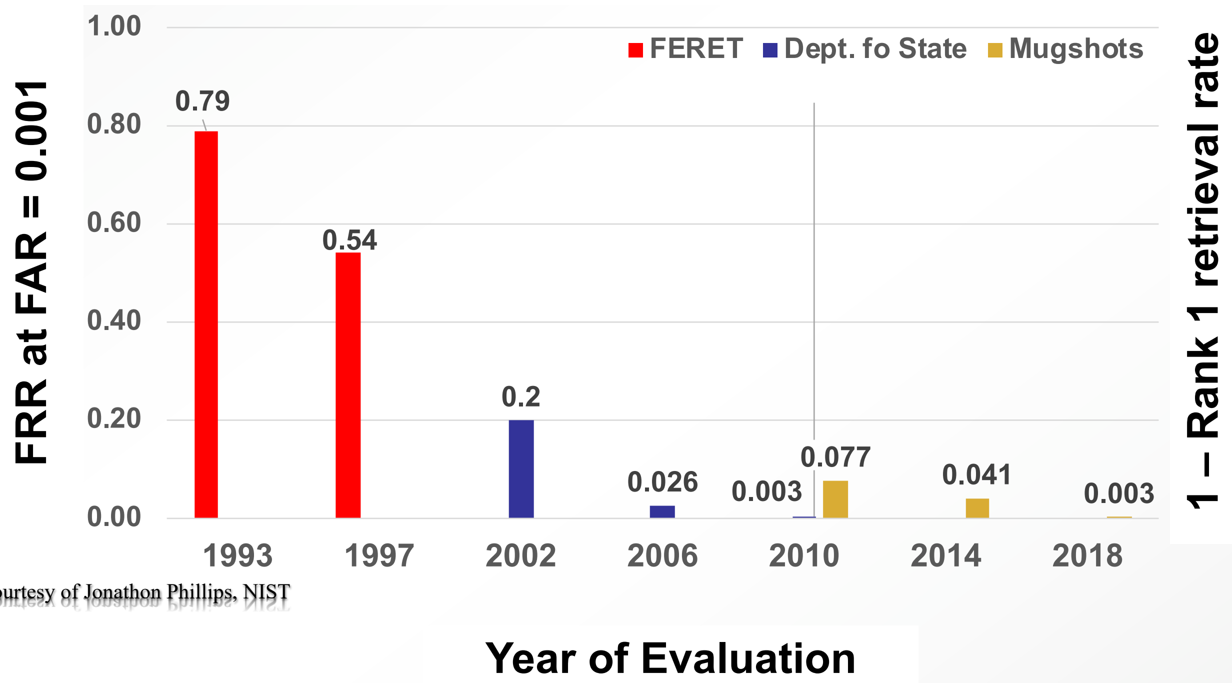
Competitions will be running from January 7, 2013 to March 22, 2013, but database and instructions will be available in late 2012. Each competition will have the opportunity to submit for review a competition summary paper for possible publication into the official proceedings.

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!]

Face recognition performance

Vision Lab

From FERET 1993 to FRVT 2018

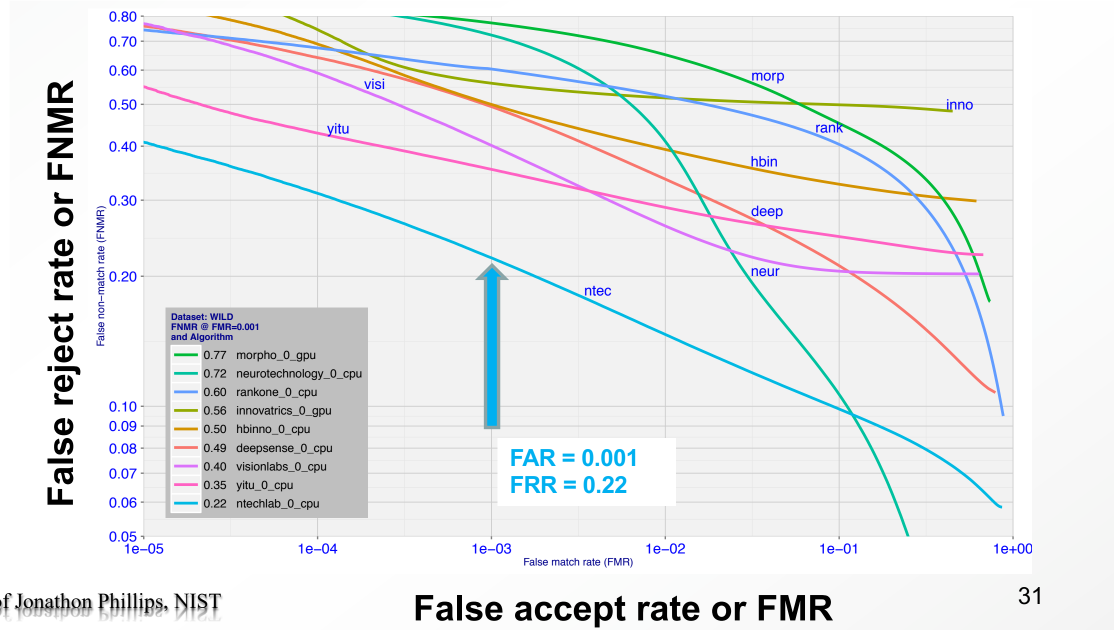


Courtesy of Jonathon Phillips, NIST

Face recognition performance

Vision Lab

FRVT 1:1 Wild-to-wild comparisons
Dataset: IJCB 2017 (sequestered data)

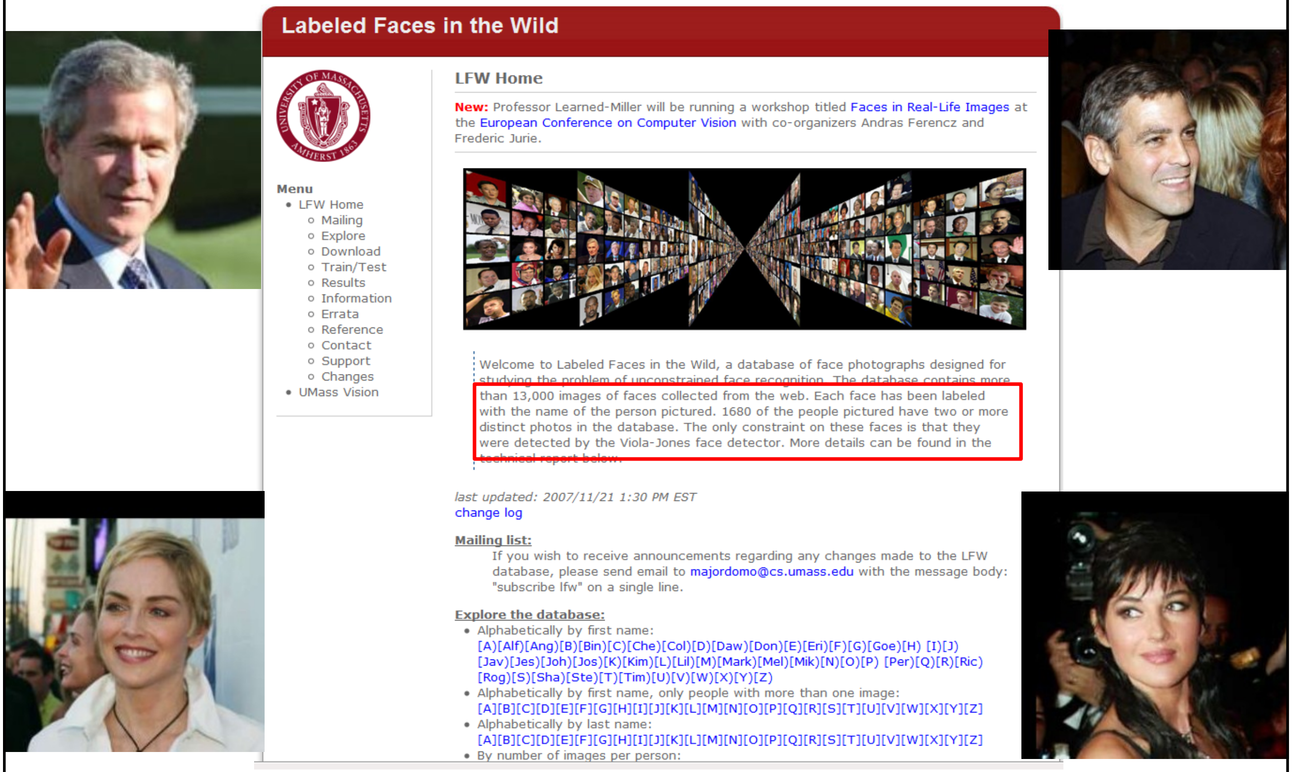


Courtesy of Jonathon Phillips, NIST

Vision Lab

FACE RECOGNITION IN THE WILD

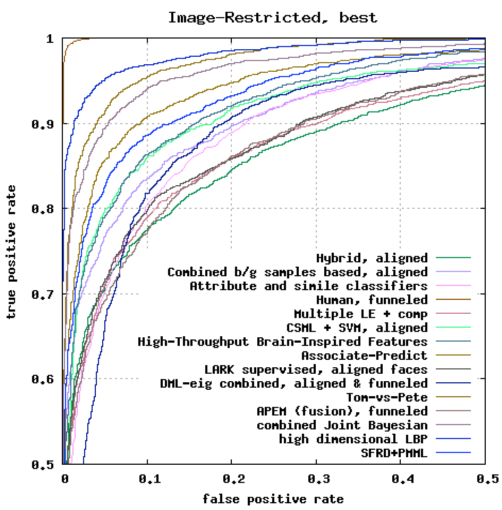
Labeled Faces in the Wild



Labeled Faces in the Wild

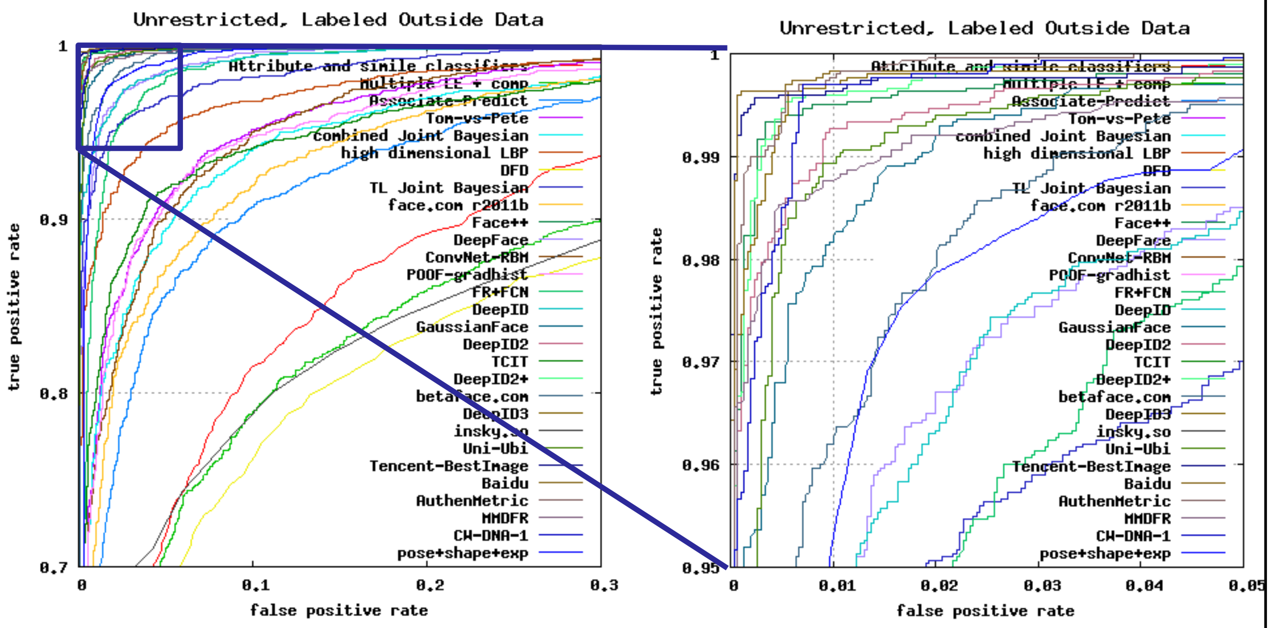
- Images downloaded from WWW (Celebrities)
- Faces detected with Viola-Jones OpenCV
- 13,233 *images*
- 5,749 *people*
- 1,680 with more than 2 images

Labeled Faces in the Wild



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

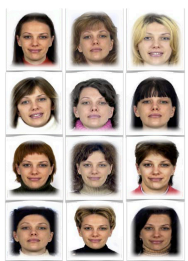
Labeled Faces in the Wild



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

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 thousands 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

The iArpa JANUS Benchmark

The IARPA Janus Benchmark face challenge (**IJB-A/B**) defines several challenges addressing verification, identification, detection, clustering and processing of wild and crowded images.

IJB-B 2017
- 1845 subjects
- 21978 images
- 7011 videos
- 10044 non-face images

IJB-A 2015
- 500 subjects
- 5712 images
- 2085 videos

iArpa JANUS Benchmarks – A/B

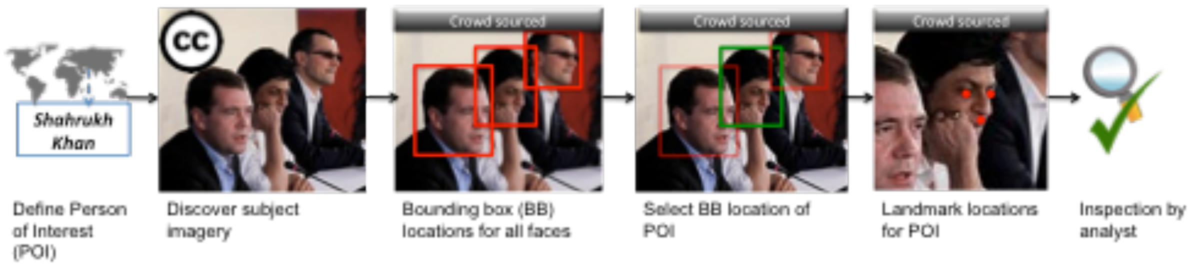
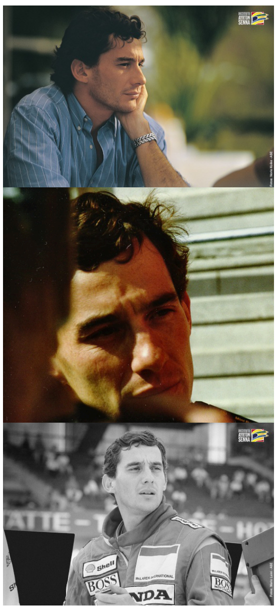


Figure 3: Overview of the data annotation process. Details can be found in [9].

IJB-A 2015	IJB-B 2017
• 500 subjects • 5712 images • 2085 videos	• 1845 subjects • 21978 images • 7011 videos • 10044 non-face images

iArpa JANUS Benchmark - A



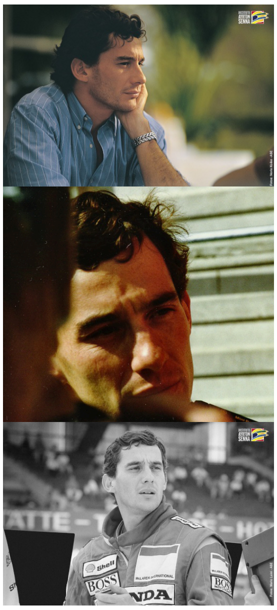
IJB-A is based on fully public data. Similar to the LFW protocol in requiring many pairs of samples to be compared in isolation. This corresponds to recognition tasks like passport verification or forensic comparison, with just a pair of samples and no central database or gallery.

IJB-A 1:1 challenge departs from LFW as follows:

- **Face selection**: LFW contains faces that could be detected with the Viola-Jones face detection algorithm. IJB-A on the other hand, uses manually located and annotated faces.
- **Landmarks**: The IJB-A tests include landmark coordinates (eyes and nose) whereas LFW provides just raw images, and aligned (funneled) images.
- **Multi-image samples**: LFW compared single images. IJB-A uses richer samples containing 1 < K < 202 images, including frames from video sequences.
- **More impostor pairs**: IJB-A 1:1 uses many more impostor comparisons that genuines. In LFW, the ratio was 1 which precluded computation of false match rates at usefully low values.

Brendan F. Klare, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah, and Anil K. Jain. **Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a.** In Proc. IEEE CVPR, June 2015.

iArpa JANUS Benchmark - A



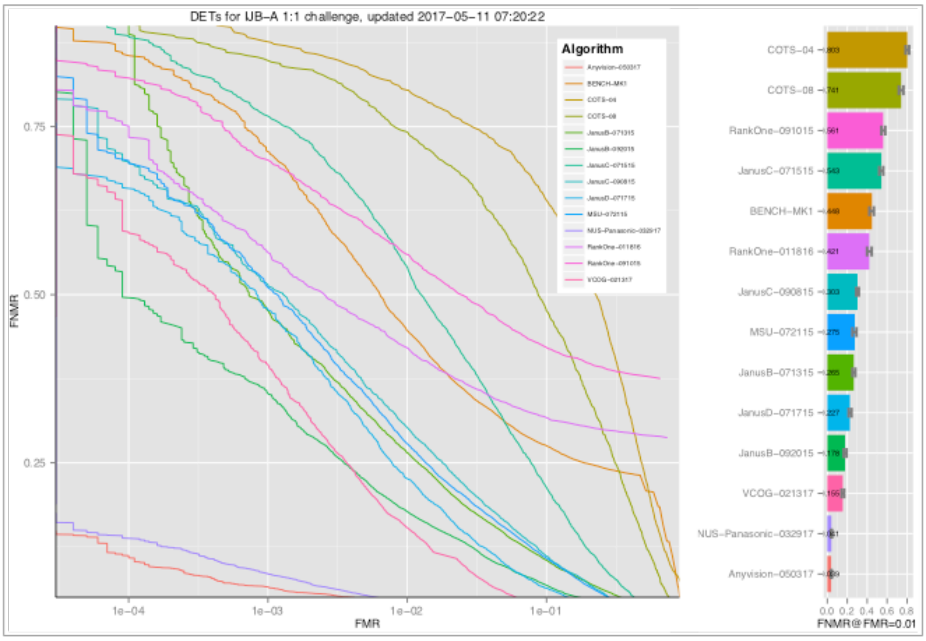


Figure 4: **DETs:** At left, are full error tradeoff characteristics (DETs) for the IJB-A algorithms. At right is a summary statistic corresponding to a vertical slice from the DETs. This is included to ease lookup. While it implies one ranking of the algorithms, it is notable that some DETs cross such that false rejection accuracy depends heavily on the FMR operating point. Algorithm developers can shape the DET characteristics according to a known use-case. Much academic research optimizes at FMR values higher often than is operationally useful. Some algorithms exhibit flatter DETs.

iArpa JANUS Benchmark - B

The IARPA Janus Benchmark-B face challenge (**IJB-B**) defines *eight challenges* addressing **verification**, **identification**, **detection**, **clustering** and **processing of** *crowded images*.

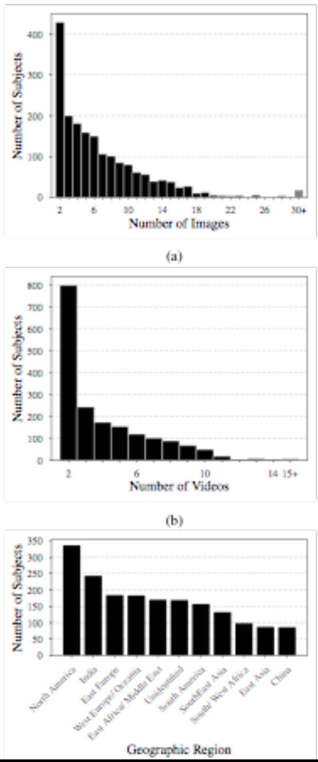
IJB-B set: 67,000 face images - 7,000 face videos - 10,000 non-face images.

Test	Mode	Purpose	"Curation"
1	1:1 Verification	Templates are comprised of mixed media (frames and stills)	All templates
2	1:1 Covariate Verification	Single still-image comparisons to perform failure analysis. *e.g. yaw difference*	All templates
3	1:N Still Image Search	Open set 1:N protocol using only still images as probe	Probe and G1/G2
4	1:N Mixed Search	Open set 1:N protocol using mixed media (frames, stills) as probe	G1/G2
5	1:N Video Search	Open set 1:N protocol using full motion video as probe. Frame and frame index with bounding box of the first occurrence of the subject of interest provided as input	G1/G2
6	Face Detection	Pure detection, without recognition. Includes pile of non-face imagery	–
7	Clustering	Cluster cropped faces. Bounding box and landmark annotations are provided to delineate the subject of interest.	–
8	Detection and Clustering	Same as 7 but requires bounding box association to identify the subjects of interest. Detect all faces in a pile of media and cluster them accordingly. This test also includes a set of non-face media as distractors.	–

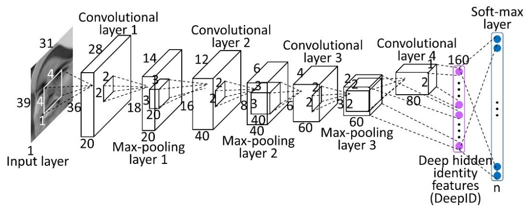
iArpa JANUS Benchmark - B

		Num. of Subjects	Num. of Templates	Pieces of Media
1:N Identification	Gallery-S1	931	931	3,081
	Gallery-S2	914	914	3,376
	Image	1,845	8,104	5,732
	Mixed	1,845	10,270	60,758
	Video	1,845	7,110	7,011
1:1 Verification	Verify	1,845	12,115	66,780
	Verify-Cov	1,845	68,195	66,780
Face Detection	Face Partition	> 1,845	N/A	66,780
	Non-face Partition	0	N/A	10,044
Clustering	Clustering-32	32	1,026	1,026
	Clustering-64	64	2,080	2,080
	Clustering-128	128	5,224	5,219
	Clustering-256	256	9,867	9,860
	Clustering-512	512	18,251	18,173
	Clustering-1024	1,024	36,575	36,092
	Clustering-1845	1,845	68,195	66,780

Table 2: Overview of the different protocols developed for IJB-B. Verify-Cov refers to Covariate Verification; Clustering-X denotes clustering with "X" number of subjects.



Convolutional Neural Networks

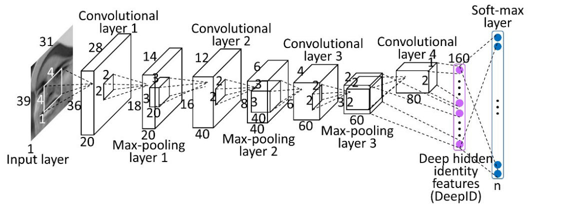


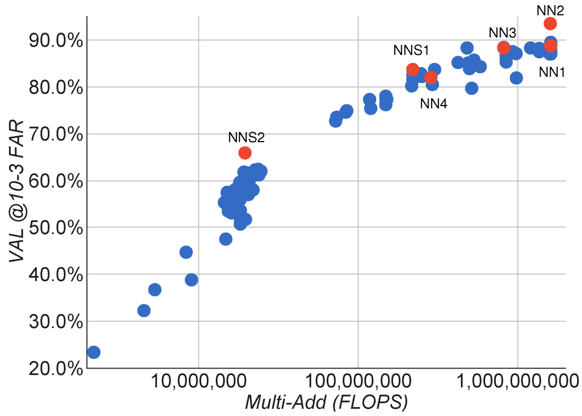
Method	Net. Loss	Outside data	# models	Aligned	Verif. metric	Layers	Accu.
DeepFace [97]	ident.	4M	4	3D	wt. chi-sq.	8	97.35±0.25
Canon. view CNN [115]	ident.	203K	60	2D	Jt. Bayes	7	96.45±0.25
DeepID [92]	ident.	203K	60	2D	Jt. Bayes	7	97.45±0.26
DeepID2 [88]	ident. + verif.	203K	25	2D	Jt. Bayes	7	99.15±0.13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

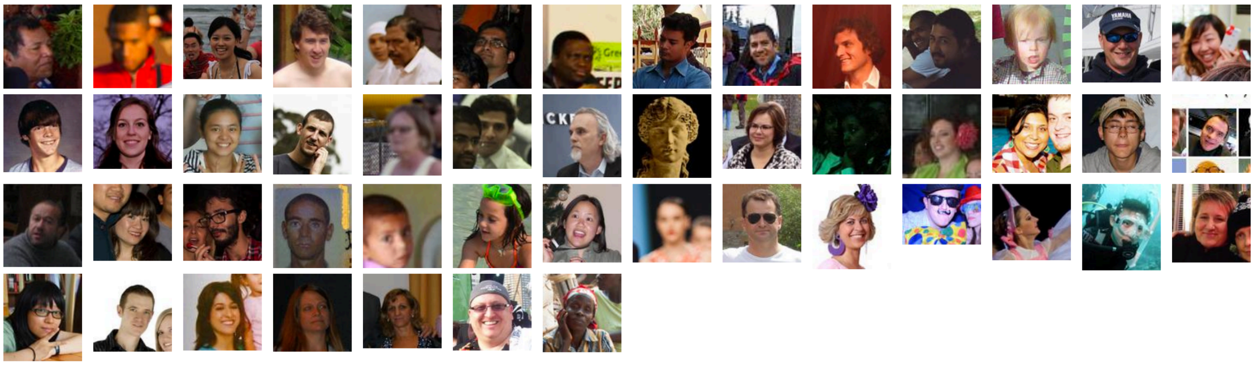
Mega faces

Dataset	Available	#Photos and #people
LFW	Public	13K of 5K people
CelebFaces 2014	Private	202K of 10K people
CASIA-WebFace 2014	Public	500K of 10K people
FaceScrub 2014	Public	100K of 500 people
YouTube Faces	Public	3425 videos of 1595 people
DeepFace (Facebook) 2014	Private	4.4 Million of 4K people
FaceNet (Google) 2015	Private	100-200 Million of 8M people
MegaFace	**Public**	**1 Million**

Figure 2: Representative sample of face recognition datasets that were created in the recent years (in addition to LFW). All the public datasets are small scale, and all the large scale datasets are mainly used for training rather than testing and are not publicly available. MegaFace (this paper) is the first large scale unconstrained dataset. It is collected from Flickr and will be available publicly.

Miller et al. (2015) Mega-Face: A million faces for recognition at scale.

Mega faces



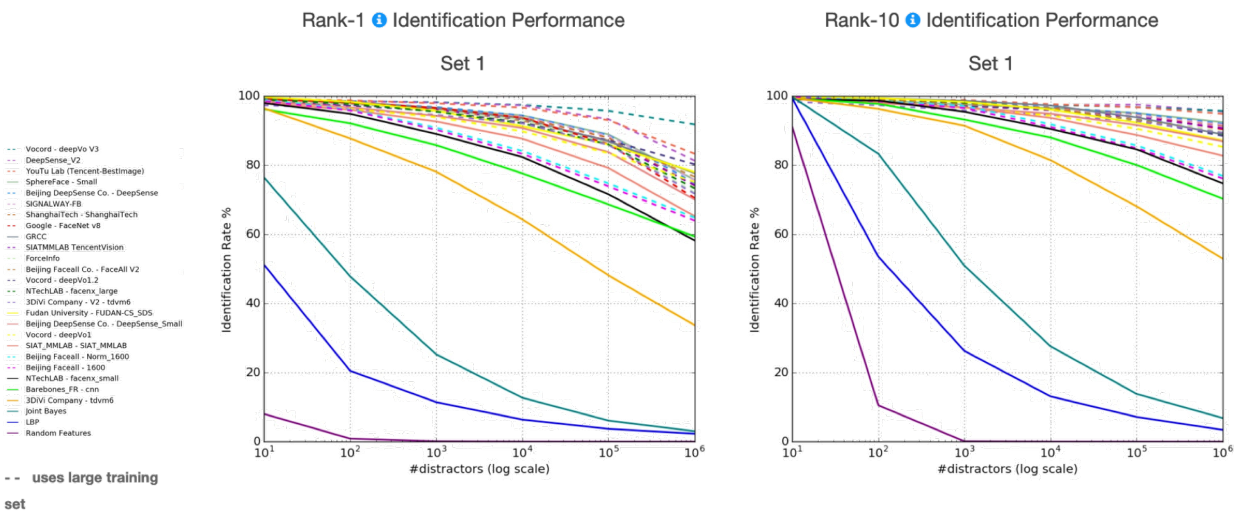
Distractors	Training Set	Test Sets
1 Million Photos	**4.7 Million** Photos	**FaceScrub** Celebrities
690,572 Unique Users	**672,057** Unique Identities	**FGNet** Age-invariant non-celebrities
	7 Mean photos / person (3 min, 2469 max)	

http://megaface.cs.washington.edu

Kemelmacher-Shlizerman, A. Seitz, D. Miller, E. Brossard, "**The MegaFace Benchmark: 1 Million Faces for Recoignition at Scale**", CVPR 2017.

Mega faces

MegaFace and MegaFace 2:
Million-Scale face recognition



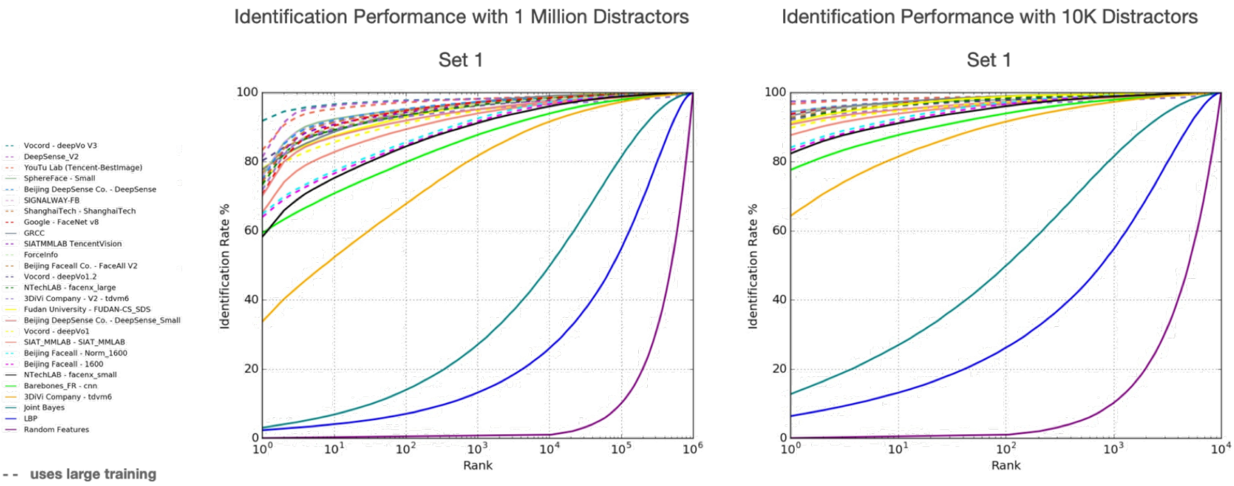
http://megaface.cs.washington.edu

Kemelmacher-Shlizerman, A. Seitz, D. Miller, E. Brossard, "**The MegaFace Benchmark: 1 Million Faces for Recoignition at Scale**", CVPR 2017.

Mega faces

Vision Lab

MegaFace and MegaFace 2:
Million-Scale face recognition



- - uses large training set

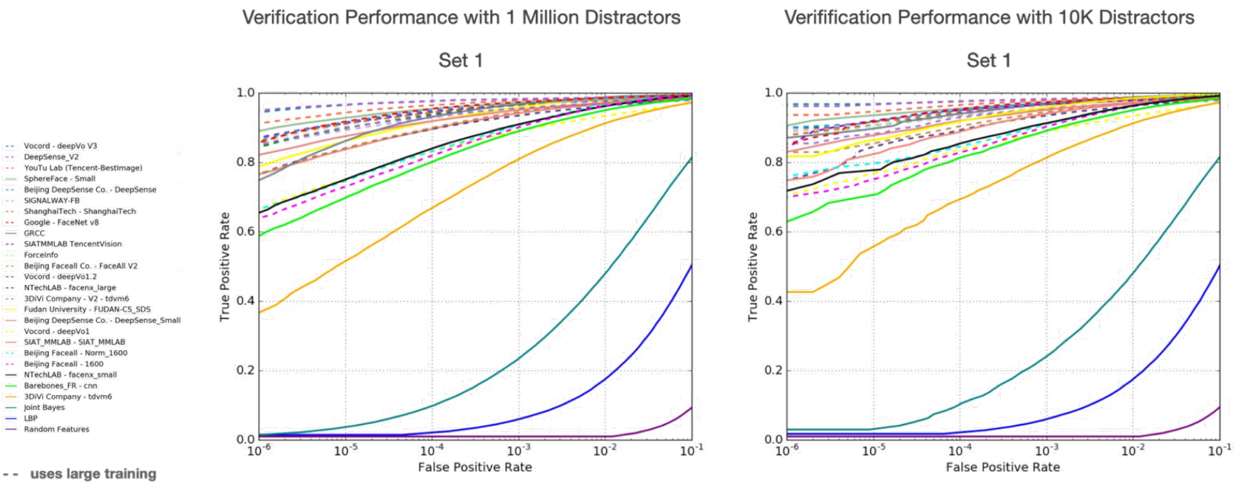
http://megaface.cs.washington.edu

Kemelmacher-Shlizerman, A. Seitz, D. Miller, E. Brossard, "**The MegaFace Benchmark: 1 Million Faces for Recoignition at Scale**", CVPR 2017.

Mega faces

Vision Lab

MegaFace and MegaFace 2:
Million-Scale face recognition



- - uses large training set

http://megaface.cs.washington.edu

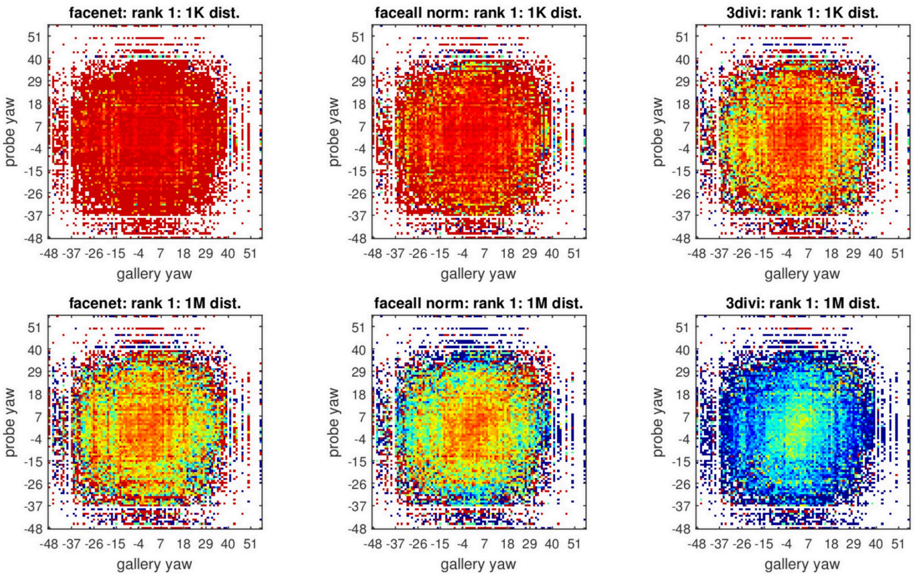
Kemelmacher-Shlizerman, A. Seitz, D. Miller, E. Brossard, "**The MegaFace Benchmark: 1 Million Faces for Recoignition at Scale**", CVPR 2017.

Mega faces

Vision Lab

MegaFace and MegaFace 2

Rank-1 Identification for varying poses

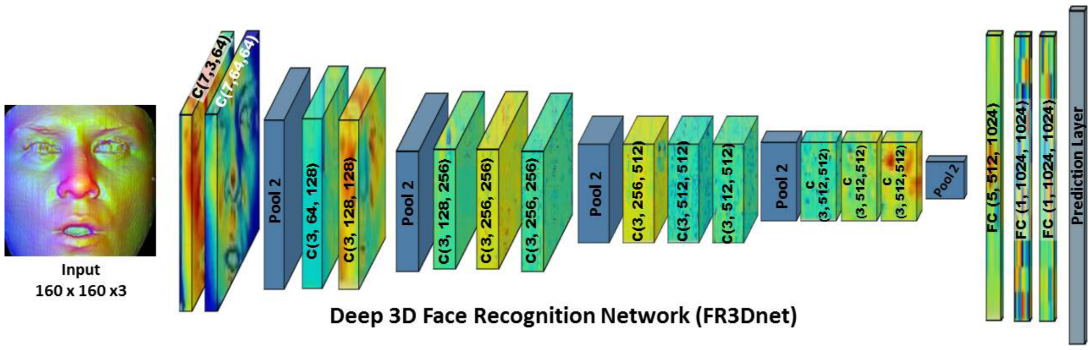


on.edu

Kemelmacher-Shlizerman, A. Seitz, D. Miller, E. Brossard, "**The MegaFace Benchmark: 1 Million Faces for Recoignition at Scale**", CVPR 2017.

Millions of 3D faces

Vision Lab



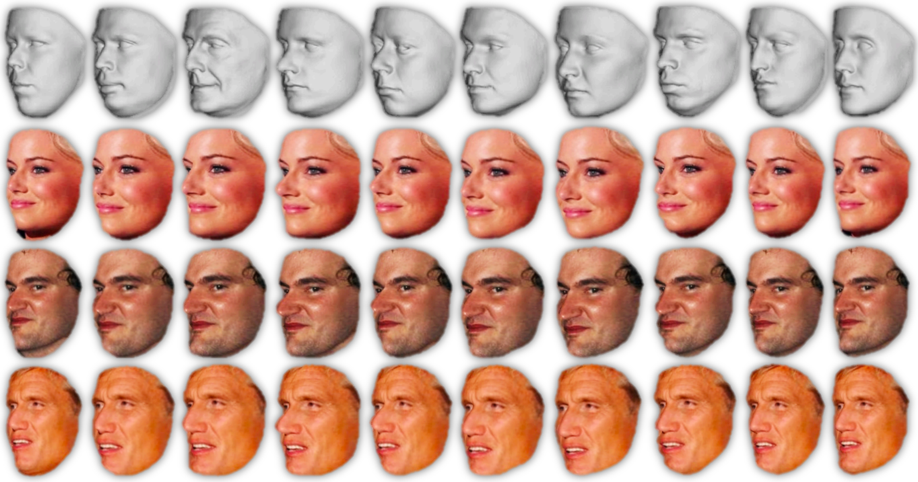
Deep 3D Face Recognition Network (FR3Dnet)

3.1M 3D scans of 100K identities

Method	Model \ Technique	Modality	Gallery of LS3DFace: LS3DFace This paper	FRGC [45]	BU3DFE [59]	BU4DFE [58]	Bosphorus [47]	CASIA [57]	GavabDB [39]	TexasFRD [22]	3D-TEC [56]	UMBDB [11]	ND-2006 [15]
CNN	GoogleNet [53]	RGB	53.97	21.51	50.76	65.41	63.44	85.91	-	53.08	79.95	65.78	24.14
	Resnet152 [23]	RGB	15.05	13.53	8.04	9.64	7.05	52.85	-	20.94	72.66	34.08	10.92
	VGG-Face [41]	RGB	90.85	87.92	97.68	96.51	96.39	94.18	-	99.73	83.30	81.54	82.86
	GoogleNet [53]	3D	38.66	35.54	46.56	41.88	26.81	50.81	66.56	67.59	67.29	47.66	30.81
	Resnet152 [23]	3D	12.49	14.40	5.80	10.13	3.84	25.34	44.26	16.25	60.98	22.20	12.08
	VGG-Face [41]	3D	61.20	62.42	71.16	53.17	48.14	71.95	77.38	85.58	78.04	67.48	60.81
Conventional	MMH [34]	3D + 2D	83.08	89.37	88.50	84.93	85.10	85.24	86.64	85.67	80.85	77.32	86.71
	3D Keypoint [35]	3D	81.76	86.59	85.14	82.50	82.64	81.38	84.41	84.99	75.63	71.68	82.30
	R3DM [17]	3D	82.89	87.50	87.13	83.21	86.06	84.51	85.60	85.47	78.27	77.11	84.84
	K3DM [18]	3D	84.67	89.50	89.24	86.05	88.60	85.35	87.90	86.13	79.55	78.64	87.77
CNN	**FR3DNet**	3D	95.51	97.06	98.64	95.53	96.18	98.37	96.39	100.00	97.90	91.17	95.62
	FR3DNet$_{FT}$	**3D**	**98.75**	**99.88**	**99.96**	**98.04**	**100.00**	**99.74**	**99.70**	**100.00**	**99.12**	**97.20**	**99.13**

S.Z. Gilani, A. Mian, , "**Learning from Millions of 3D Scans for Large-scale 3D Face Recognition**", CVPR 2018.

Efficient CNN Learning

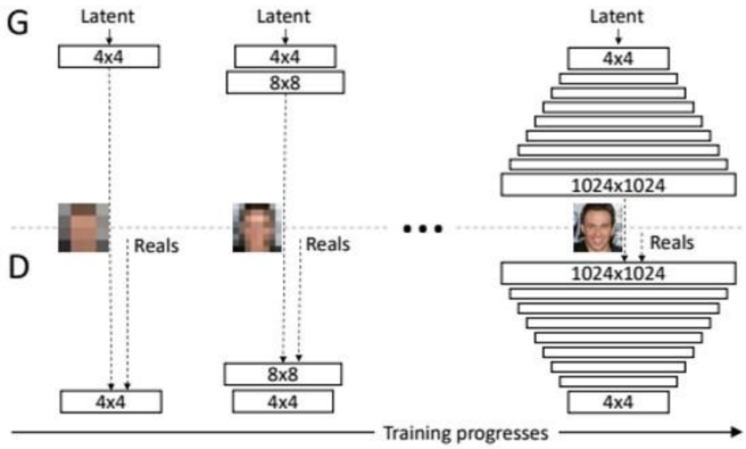


Augmenting faces by using different generic 3D models for rendering

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.

Generating Faces



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, producing images of unprecedented quality, e.g., CelebA images 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

Vision Lab

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, producing images of unprecedented quality, e.g., CelebA images 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 in the city

Vision Lab

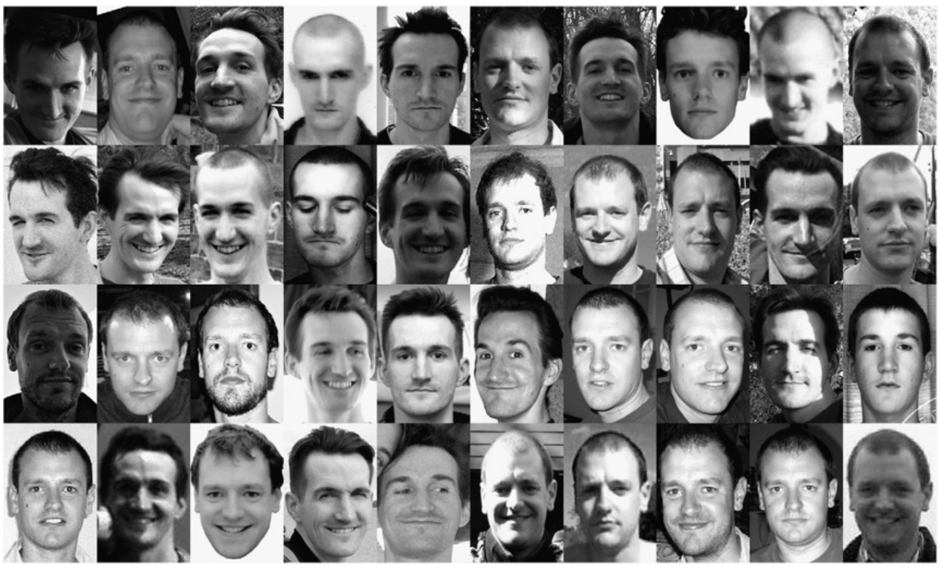
Driver and valid Id identification: the Shenzhen Police installed thousands of cameras with a face recognition system from the **AI company Intellifusion** to catch drivers without a valid driver license.

Face recognition in the city

Detection of jaywalkers: Pedestrians crossing on red are caught by high resolution surveillance cameras and displayed on large screen together with their identity matched on a massive database. Shenzhen said it has shamed almost 14,000 jaywalkers in 10 months. This was not the case in Shanghai as a famous business woman, Dong Mingzhu was recognized from an advertisement on the side of a bus.

CNN Performance

- **How do humans perform in recognizing faces?**



Jenkins, White, Burton (2011)

CNN Performance

Vision Lab

- **How do machines perform in recognizing faces?**

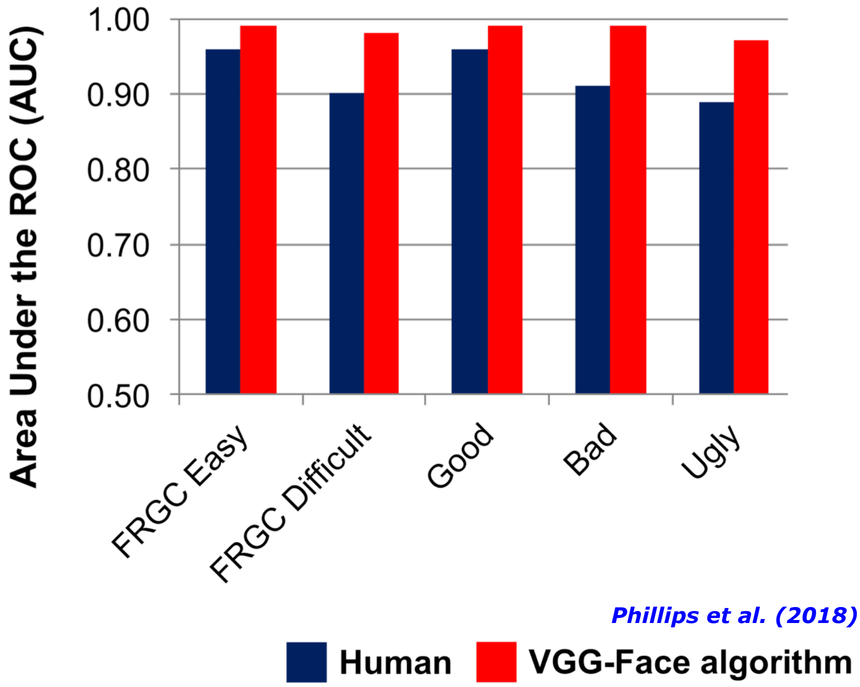
Good | Challenging | Very Challenging

Phillips et al. (2018)

CNN Performance

Vision Lab

- **How do machines perform in recognizing faces?**

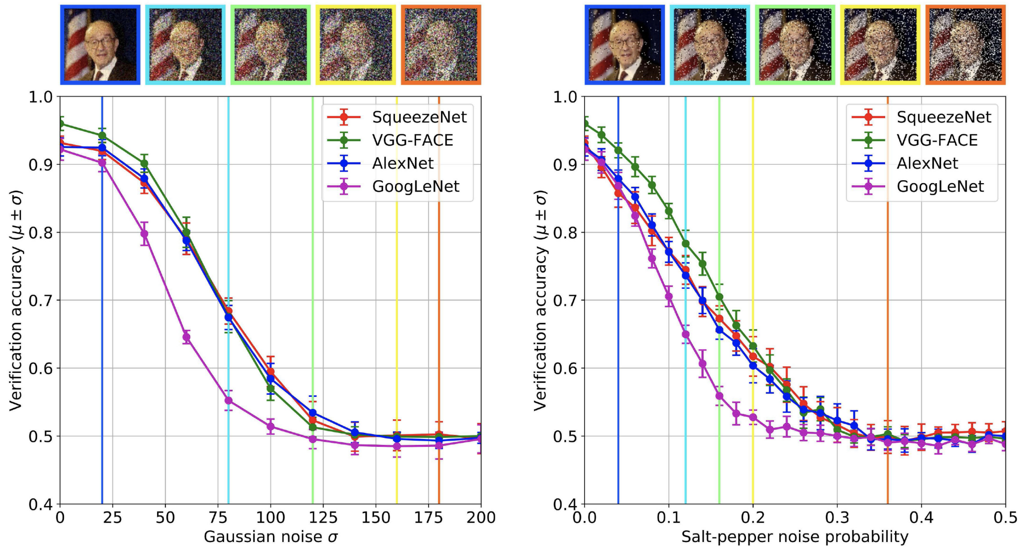


Phillips et al. (2018)

CNN Performance

Vision Lab

- **However, we're not done yet...**

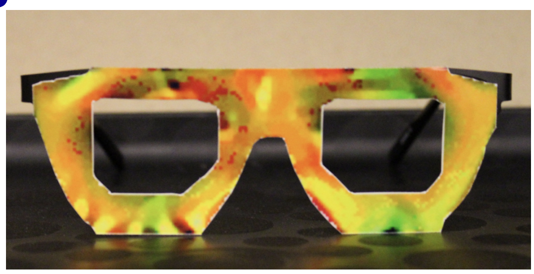


K. Grm , V. Štruc, A. Artiges, M. Caron, H. K. Ekenel, "**Strengths and weaknesses of deep learning models for face recognition against image degradations**" IET Biometrics, 7(1):81-89, 2018

CNN Performance

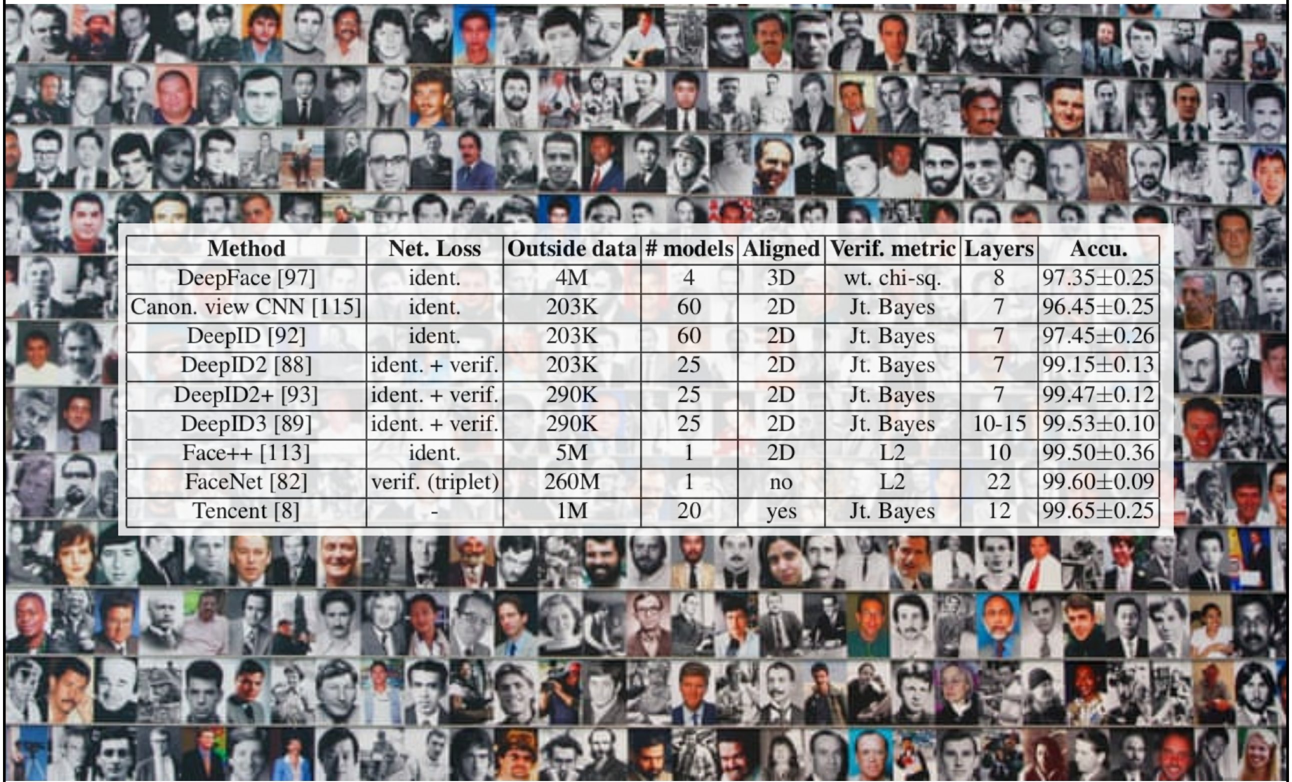
Vision Lab

- **The "magic glasses"**



M. Sharif , S. Bhagavatula, L. Bauer, M. K. Reiter, "**Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition**", CCS'16 October 24-28, 2016, Vienna, Austria

The "curse of training"



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

Is this *you*?

Face recognition under ban

CNN BUSINESS

San Francisco just banned facial-recognition technology

By Rachel Metz, CNN Business

Updated 2315 GMT (0715 HKT) May 14, 2019

Microsoft CEO says self regulation needed with new technologies

US Steel announces temporary layoffs

Company is growing steak without the cow

TOP STORIES

What we learned from one of Jeffrey Epstein's final interviews with a...

A 3-year-old was found alone and adrift in a boat in Texas. A man's...

San Francisco (CNN Business) — San Francisco, long one of the most tech-friendly and tech-savvy cities in the world, is now the first in the United States to prohibit its government from using facial-recognition technology.

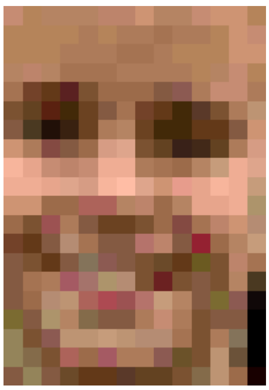
The ban is part of a broader anti-surveillance ordinance that the city's Board of Supervisors approved on Tuesday. The ordinance, which outlaws the use of facial-recognition technology by police and other government departments, could also spur other local governments to take similar action. Eight of the board's 11 supervisors voted in favor of it; one voted against it, and two who support it were absent.

…The ordinance adds yet more fuel to the fire blazing around facial-recognition technology. While the technology grows in popularity, it has come under increased scrutiny as concerns mount regarding its deployment, accuracy, and even where the faces come from that are used to train the systems.

https://edition.cnn.com/2019/05/14/tech/san-francisco-facial-recognition-ban/index.html

Human face perception

How many pixels to detect/recognize a face?

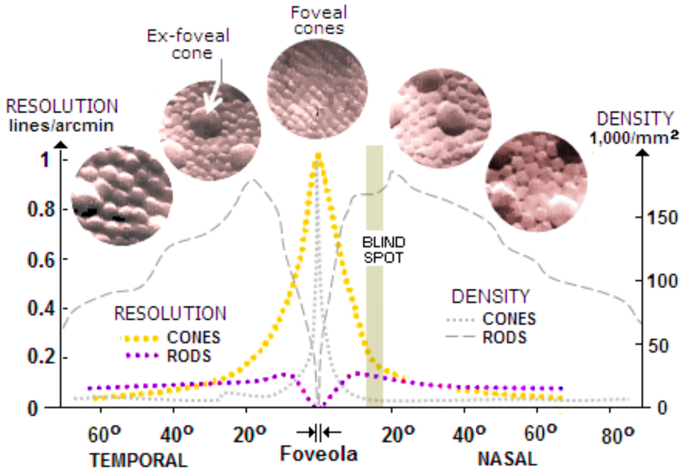


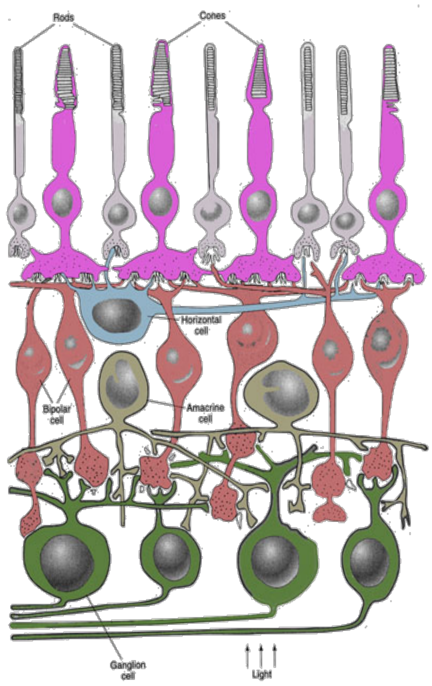
… Not many … *(20x14)*

It's more a question of **spatial distribution** and …proper **frequency tuning**

The human retina

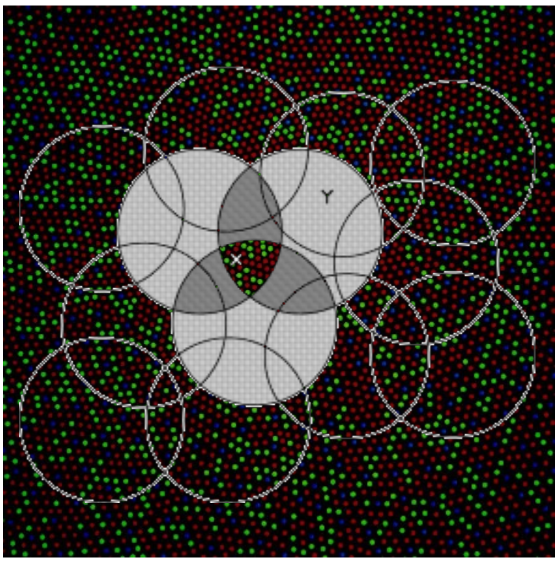
Vision Lab

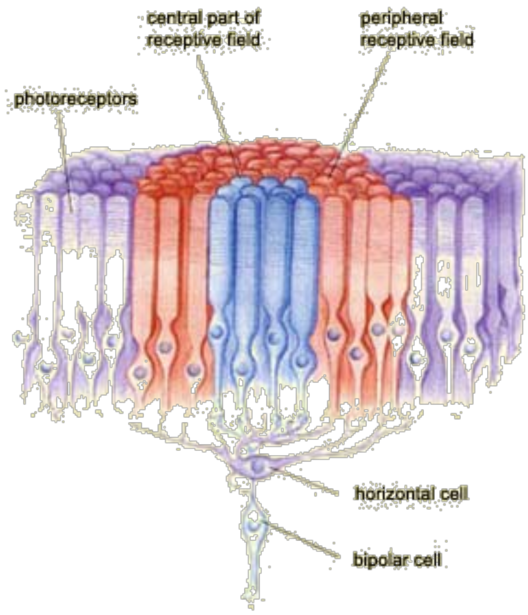




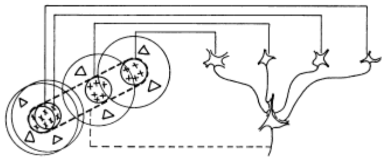
Receptive fields

Vision Lab

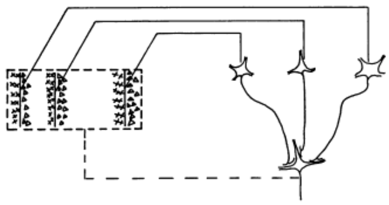




Hubel & Wiesel 1962



Text-fig. 19. Possible scheme for explaining the organization of simple receptive fields. A large number of lateral geniculate cells, of which four are illustrated in the upper right in the figure, have receptive fields with 'on' centres arranged along a straight line on the retina. All of these project upon a single cortical cell, and the synapses are supposed to be excitatory. The receptive field of the cortical cell will then have an elongated 'on' centre indicated by the interrupted lines in the receptive-field diagram to the left of the figure.



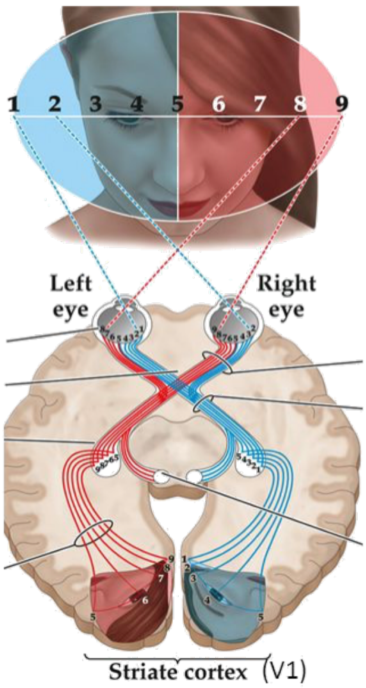
Text-fig. 20. Possible scheme for explaining the organization of complex receptive fields. A number of cells with simple fields, of which three are shown schematically, are imagined to project to a single cortical cell of higher order. Each projecting neurone has a receptive field arranged as shown to the left: an excitatory region to the left and an inhibitory region to the right of a vertical straight-line boundary. The boundaries of the fields are staggered within an area outlined by the interrupted lines. Any vertical-edge stimulus falling across this rectangle, regardless of its position, will excite some simple-field cells, leading to excitation of the higher-order cell.

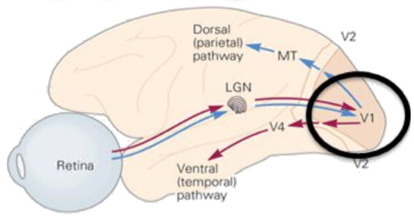
Simple and Complex cells

Hubel DH & Wiesel TN (1962). "**Receptive fields, binocular interaction and functional architecture in the cat's visualcortex**". JPhysiol160, 106–154

Retinotopic mapping

V1 retinotopic maps

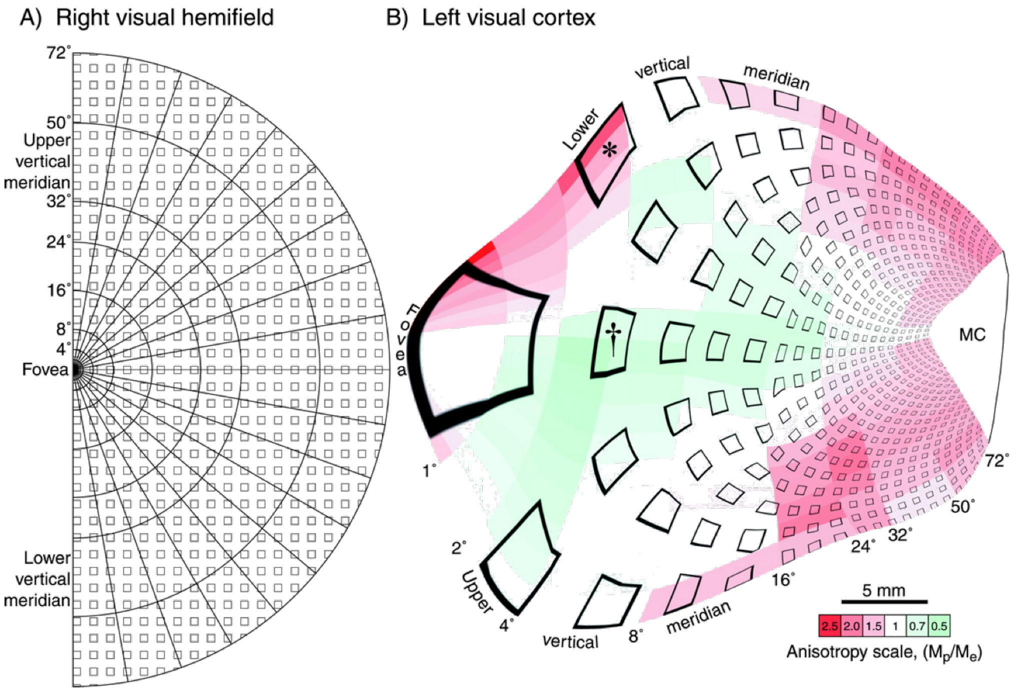




- Each point of the visual field maps on to a local group of neurons in V1.
- Retinotopy = Remapping of retinal image onto cortical surface
- Foveal region uses more of V1 (greater magnification factor)

Retinotopic mapping

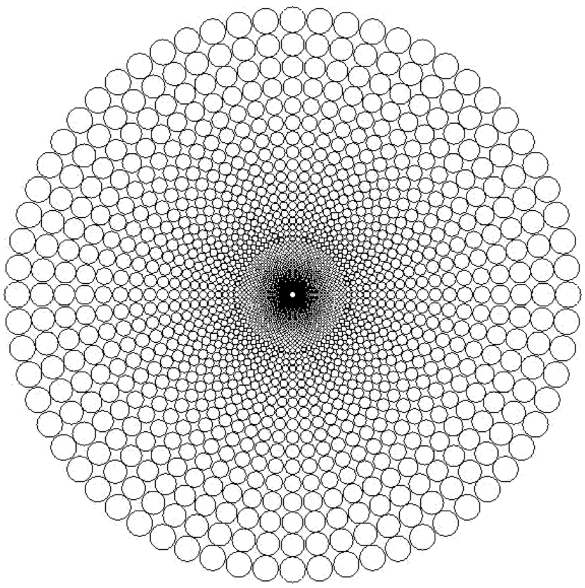
Vision Lab



Log-Polar mapping

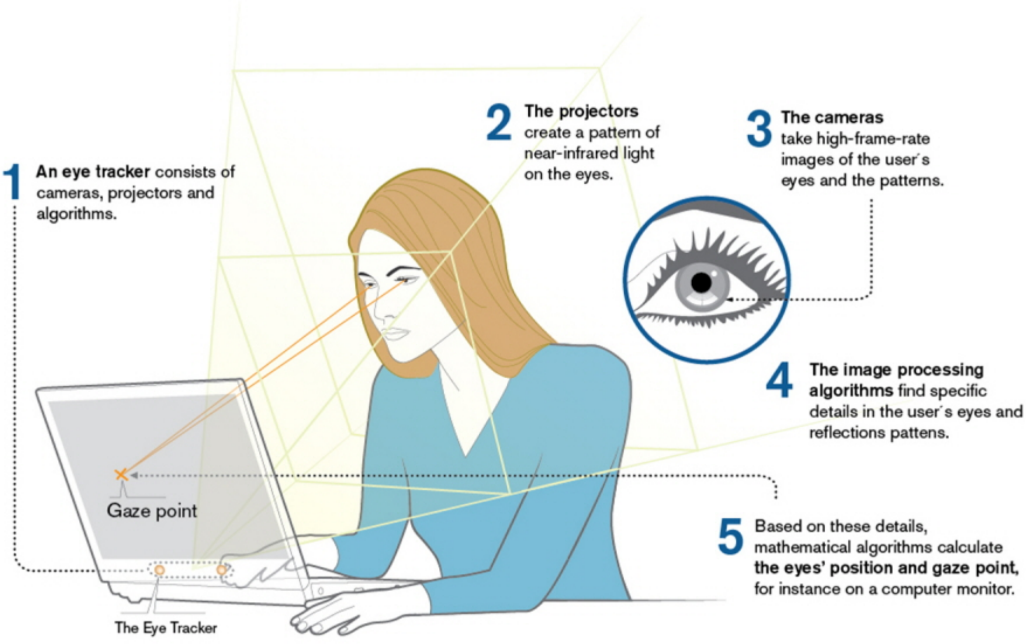
Vision Lab

The **complex log-polar transform** is a good approximation of the retinal sampling

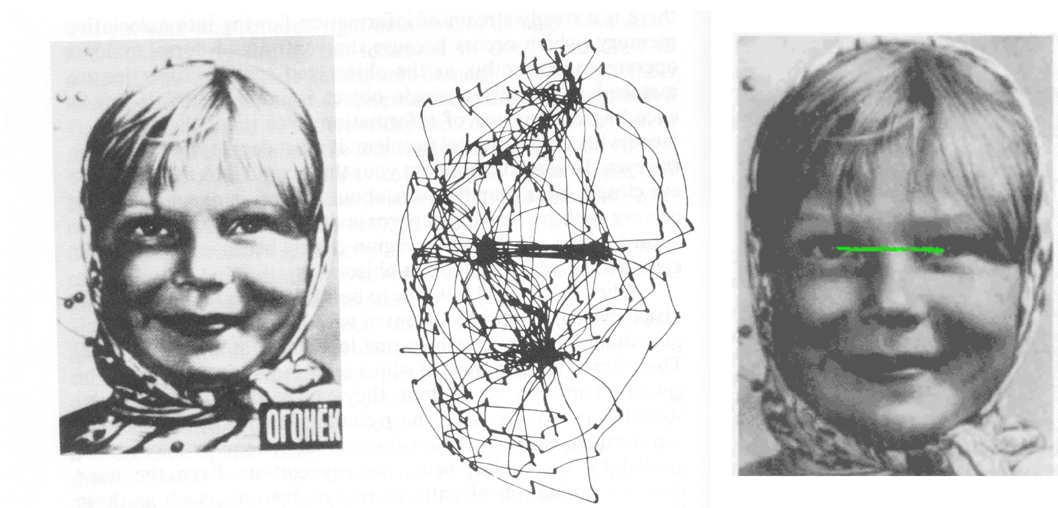


$$\begin{cases} x = \rho \sin\theta \\ y = \rho \cos\theta \end{cases} \qquad \begin{cases} \xi = \log_a\left(\rho / \rho_0\right) \\ \eta = q\theta \end{cases}$$

Visual attention



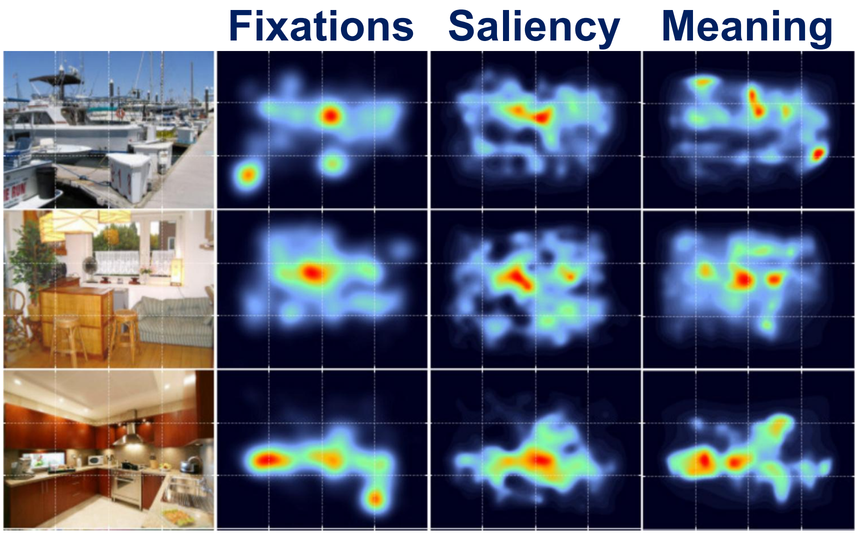
Visual attention



Eye movements while watching a girl's face

A.L. Yarbus, "**Eye Movements and Vision**", Plenum Press, 1967

Visual attention



- Attention is driven by **utilitarian features** related to the **objects' meaning**

J.M,. Henderson, T.R. Hayes, "**Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps**", Journal of Vision 18(6):1-18, June 2018

Visual attention

- Body parts have a **meaning**

Visual attention

- Body parts have a **meaning**

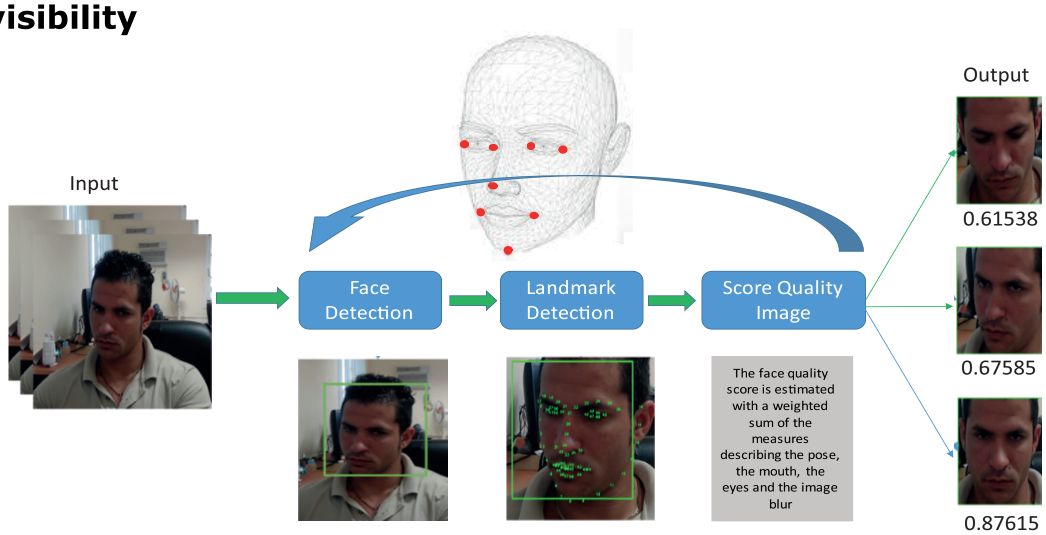
Visual attention

- Body parts have a **meaning**

Visual attention

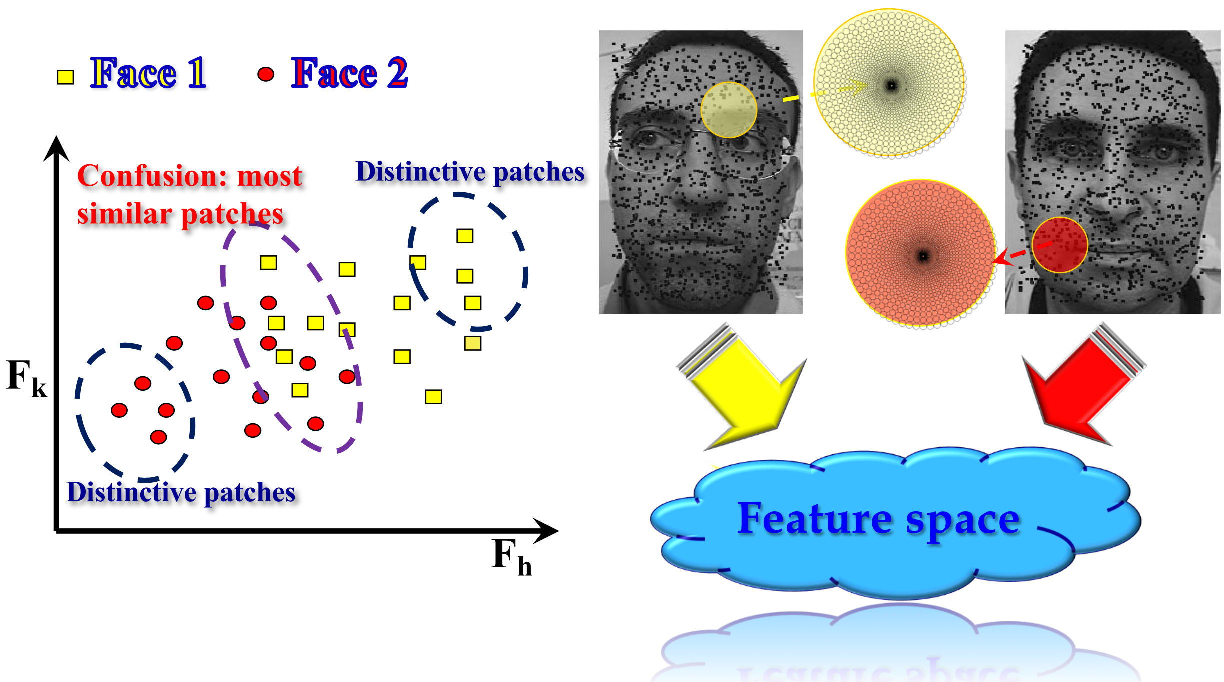
Vision Lab

- ***Meaningful* facial regions** are extracted according to the position of facial landmarks
- Images are clustered in different categories, according to the approximate **head rotation** along the vertical axis.
- Regions are associated to each pose category according to their **visibility**



Facial landmarks

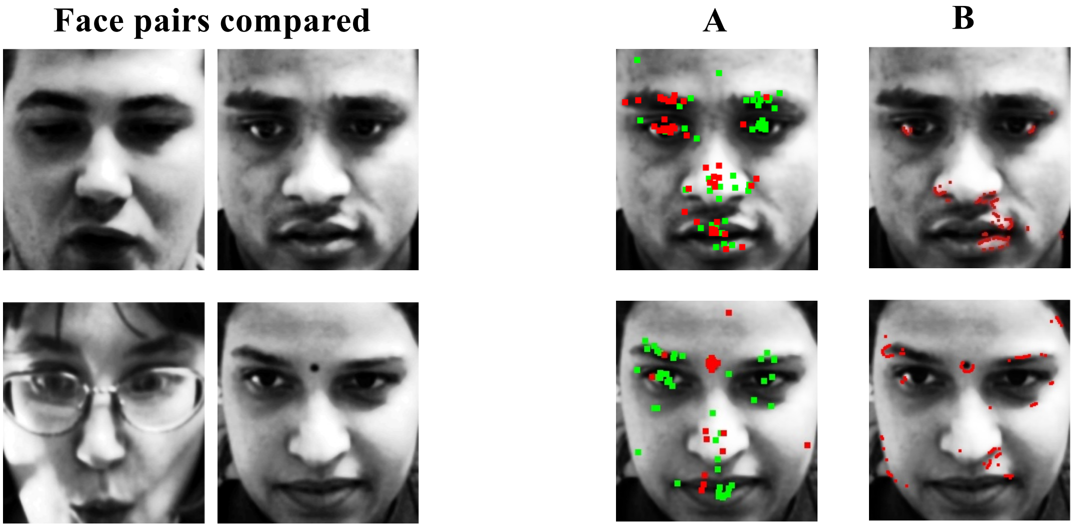
Vision Lab



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.

Facial landmarks

Vision Lab

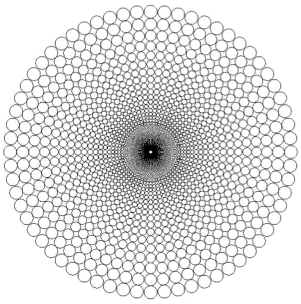


(A) perceptual and (B) computational results of saliency of local facial features, demonstrate the relevance of *non-standard* facial landmarks

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.

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

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

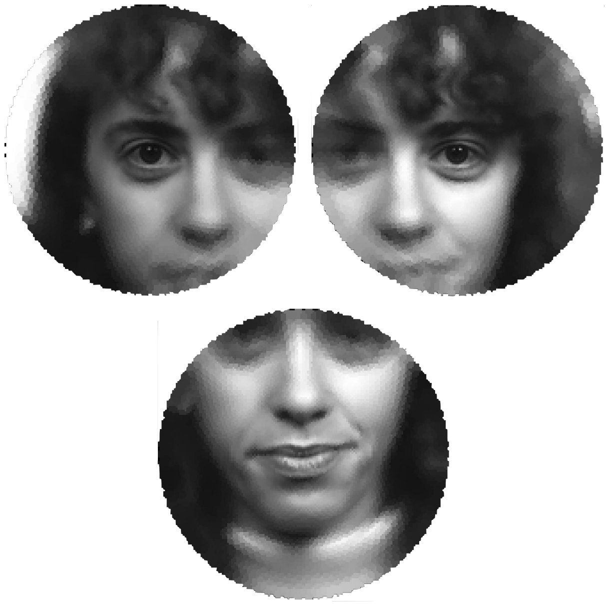
Space-variant imaging

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

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

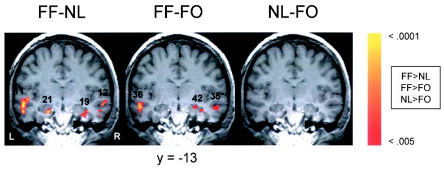
Tistarelli, M. and Grosso, E. (1997) "**Active face recognition with an hybrid approach**" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997

Tistarelli, M. and Grosso, E. (2000) "**Active vision-based face authentication**" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Functional MRI

Vision Lab





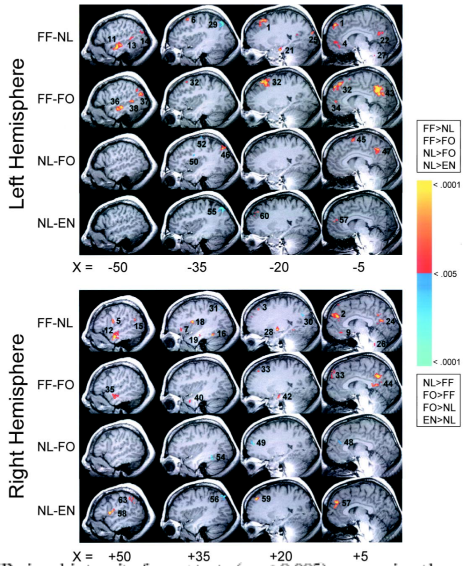


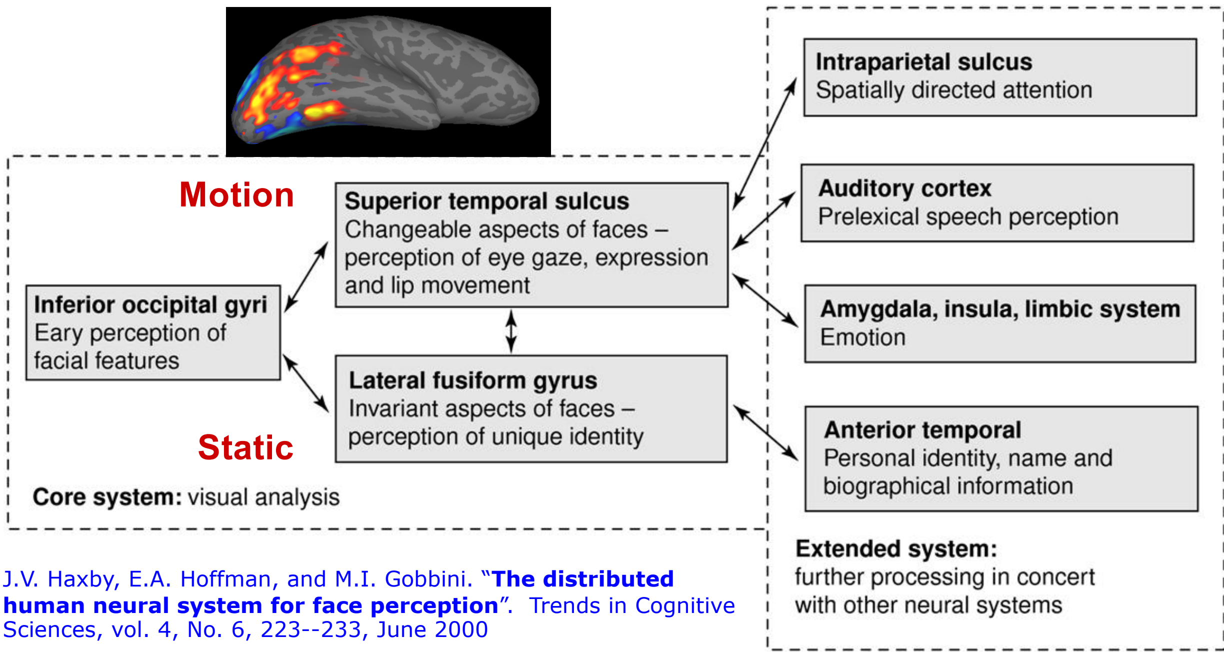
Figure 2. Areas of significantly increased (*red–yellow* scale) and decreased (*blue–cyan* scale) MR signal intensity from *t* tests ($p < 0.005$) comparing the three conditions: FF minus NL, FF minus FO, and NL minus FO. Numbers below each image represent millimeters from the interhemispheric fissure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

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

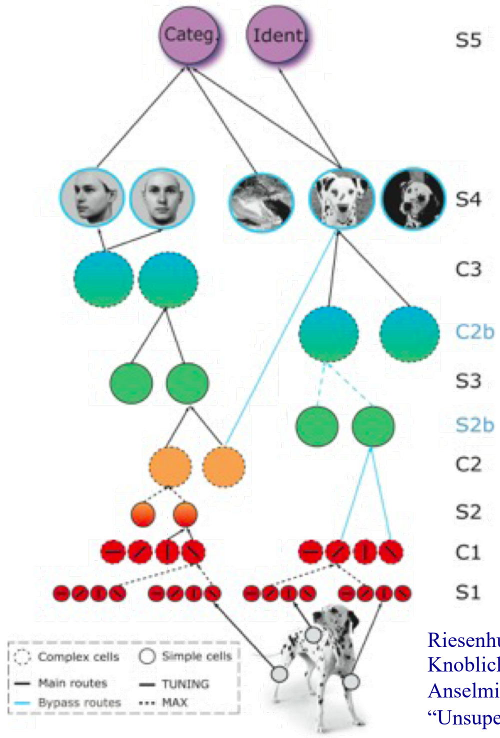
Brain models

Vision Lab



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

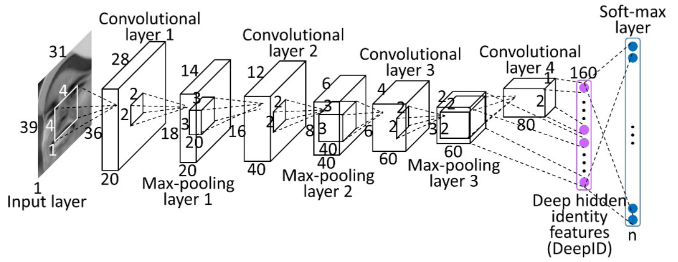


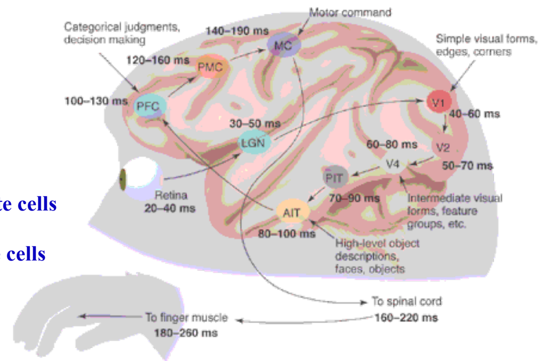
C2 **Complex composite cells**

S2 **Composite feature cells**

C1 **Complex cells**

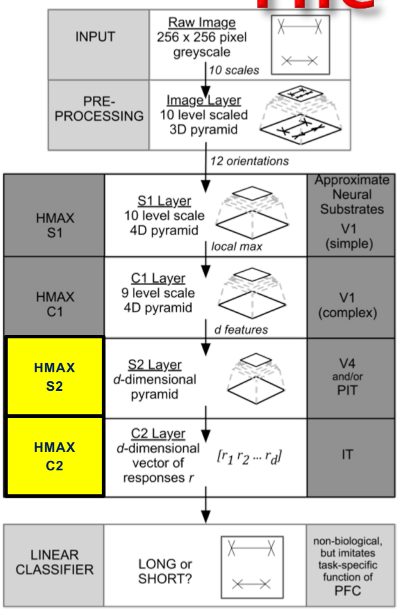
S1 **Simple cells**

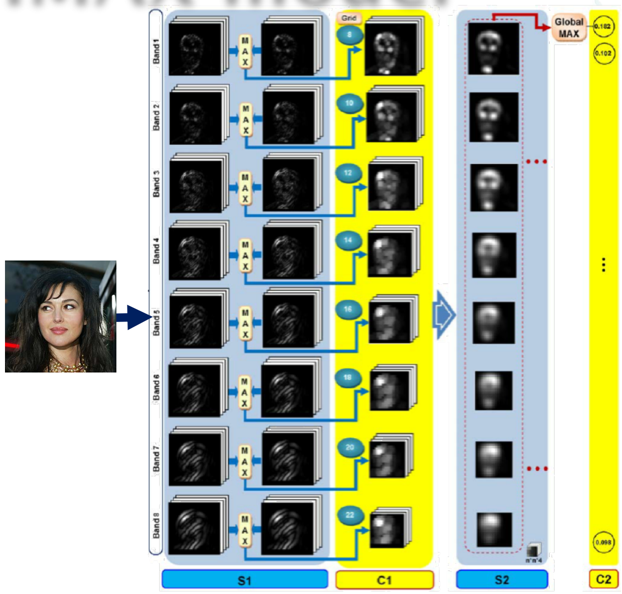




Riesenhuber & Poggio 1999, 2000; Serre Kouh Cadieu
Knoblich Kreiman & Poggio 2005; Serre Oliva Poggio 2007
Anselmi, F., Leibo, J. Z., Rosasco, L., Mutch, J., Tacchetti, A., and Poggio, T.,
"Unsupervised learning of invariant representations", Theoretical Computer Science, 2015.

The HMAX model

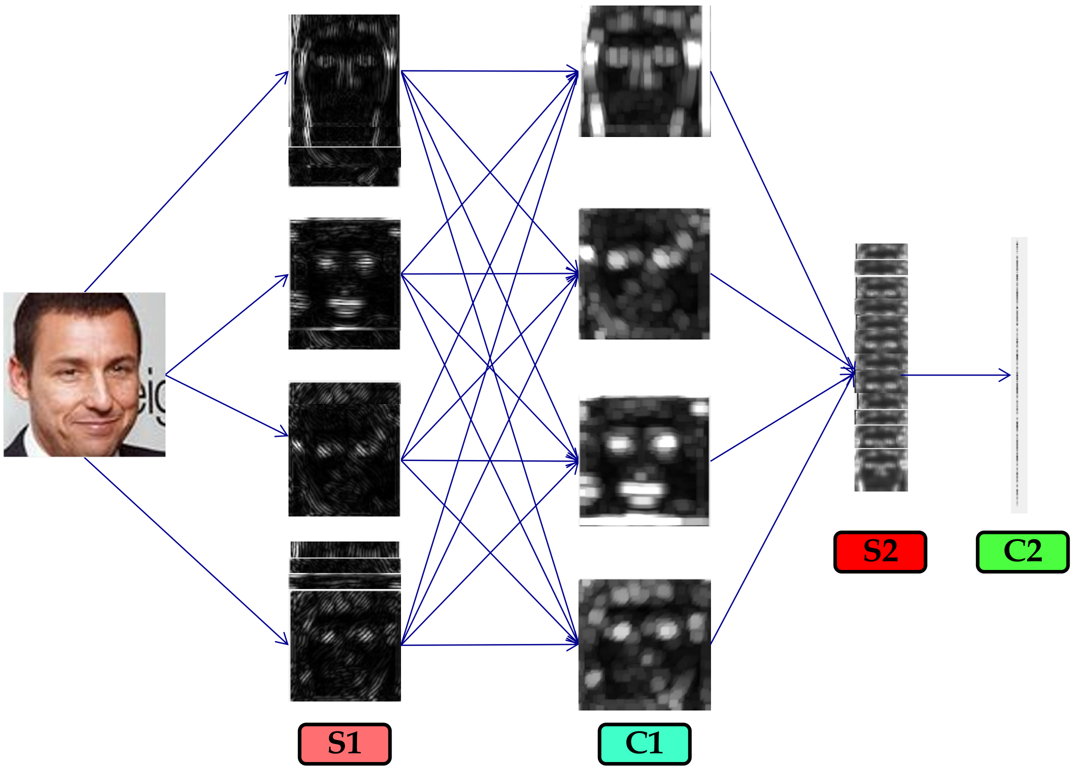




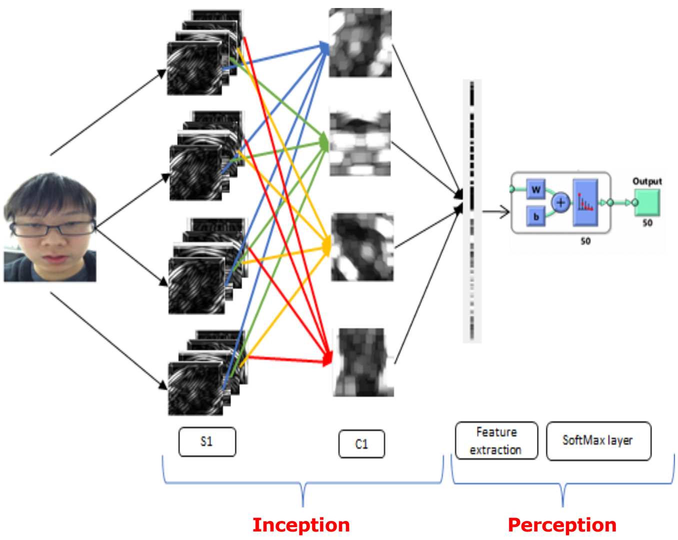
Riesenhuber, M. & Poggio, T. (1999). Hierarchical Models of Object Recognition in Cortex. Nature Neuroscience 2: 1019-1025

(S1) In this layer an input image is analyzed with a pyramid of filters (16 filter sizes×4 orientations=64)
(C1) In this layer, the local maximum between 2 adjacent scales with the same orientation is taken.
(S2) The Euclidean distances between stored prototypes, which are obtained in the learning stage, and new input is computed. This process occurs for all bands in C1 and as a result, S2 maps are obtained.
(C2) The global maximum is computed over all S2 responses in all positions and scales in this layer.

The HMAX model



Face recognition with HMAX

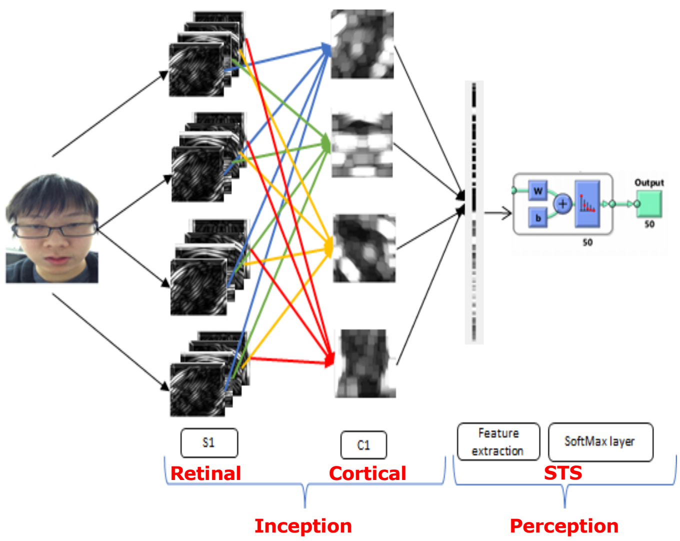


The architecture is inspired by the concept of ***Inception-Perception***.

The ***Inception*** part is implemented by the **S1** and **C1** layers of the **HMAX** network, followed by a down-sampling operator to build the feature vectors.

The ***Perception*** part is implemented by a **SoftMax** layer.

Face recognition with HMAX



The **Gabor** and **max pooling** layers encode the face images based on a biologically-inspired chain running from the **retinal** stage to the **V1 cortex**.

The connections between the **V1 cortex** and the **Superior Temporal Sulcus**, the face-selective area, is simulated by a network whose neurons are activated by a ***SoftMax*** function.

Feature extraction and fusion

- The **S1** and **C1** layers in the HMAX are used.
  - The **S1** layer performs a band-pass filtering with a bank of Gabor kernels.
  - At the local invariance layer (**C1**), a local maximum is computed for each orientation.
- The final feature vector is built by down-sampling the output by 8, obtaining a 256-dimensional feature vector.
- The feature vectors, extracted from different facial regions, are concatenated into a single feature vector of fixed size, given by the head pose. For example, the feature vector for the second category pose is:

$$F = [F_{le}; F_m; F_c; F_a]$$

F_{le} ; F_m; F_c *and* F_a are the feature vectors obtained from the face regions extracted from the left eye, mouth, chin and forehead.

Classification

- During the learning phase, a neural network, with a ***SoftMax*** activation, is trained from a subset of the available sample data (disjoint from the test data).
- The loss function for the ***SoftMax*** layer is based on the computation of the ***crossentropy***:

$$L_i = -\log\left(\frac{e^{f_i}}{\sum j e^{f_j}}\right)$$

Where $\boldsymbol{f_j}$ is the *j-th* element of the feature vector representing subject ***f***, while $\boldsymbol{L_i}$ is the full loss over the training examples.

- The concatenated feature vectors are fed to the classification network. The scores obtained from each image group are fused by applying a mean rule.

Experimental results

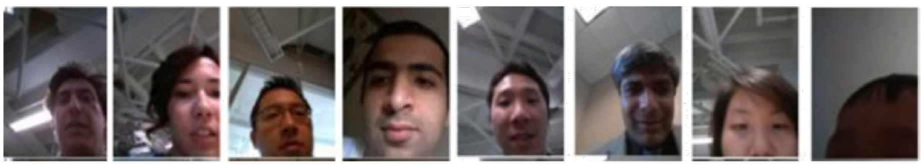


TABLE VI
RECOGNITION RATE OBTAINED BY FUSING THE FEATURES EXTRACTED FROM DIFFERENT FACIAL REGIONS.

Session	Best frames	Average frames	Bad frames
1	**96.36**	72.73	58.18
2	**87.27**	34.55	74.55
3	**80.00**	50.91	54.55

- In this experiment different regions are fused from each frame category.
- The features extracted from the fiducial regions are concatenated into a single feature vector for classification.

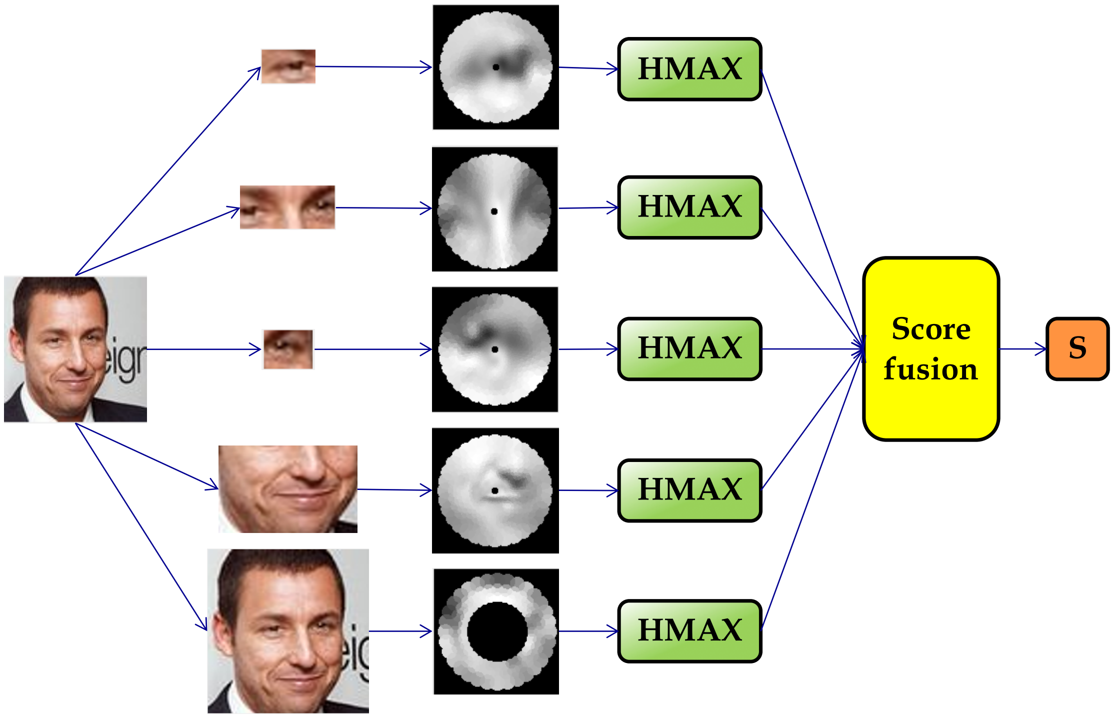
- In this experiment the protocols defined for the UMDAA database were applied.
- Performances are compared with Fisher Faces (FF), Sparse Representation based classification (SRC) and Mean-Sequence SRC (MSSRC) and the VGG deep network model.

TABLE VII
COMPARISON WITH THE METHODS DESCRIBED IN [1] AND FOLLOWING THE TESTING PROTOCOL 1. THE RECOGNITION RATE FOR THE PROPOSED METHOD WAS OBTAINED FROM THE THREE HEAD POSE GROUPS.

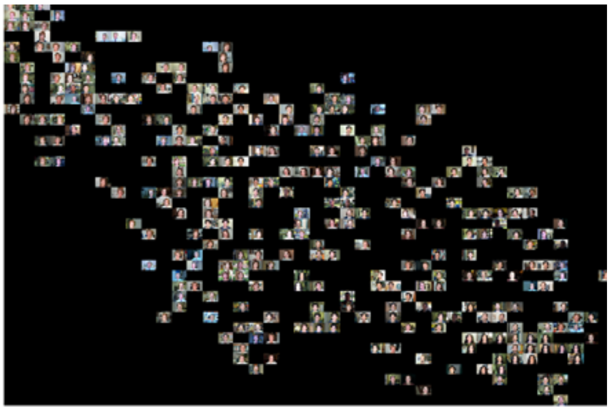
Training	Testing	FF	SRC	MSSRC	S1C1	VGG	Proposed ap-proach
1	2	54.48	52.79	47.21	7.27	**67.27**	61.82
1	3	45.27	51.18	46.15	16.36	49.09	**52.73**
2	1	25.52	44.18	43.06	20.00	50.91	**65.45**
2	3	56.8	58.58	60.36	52.73	38.18	**72.73**
3	1	24.77	17.64	17.64	20.00	**47.27**	36.36
3	2	**56.01**	51.95	45.85	51.82	33.64	50.91

Khellat Kiehl, S, Lagorio, A and Tistarelli, M (2019) "**A Biologically-Inspired Attentional Approach for Face Recognition**" *Proc. of IAPR/IEEE Int.l Workshop on Biometrics and Forensics* – IWBF 2019, Cancun, Mexico, May 2019.

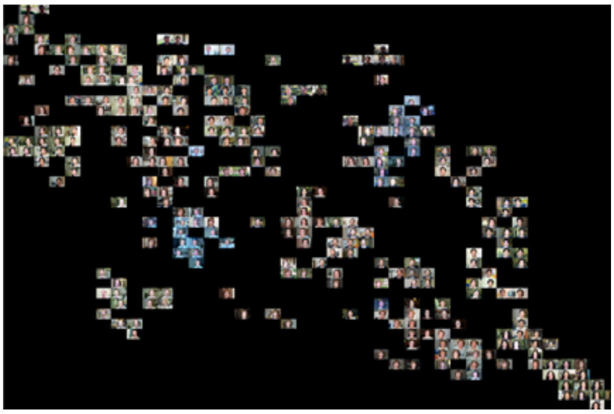
Foveated face recognition



Foveated face recognition

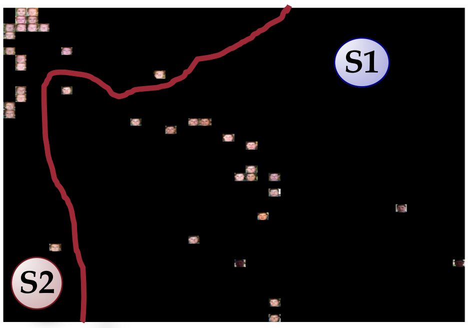


HMAX Space representation on uniformly sampled face images

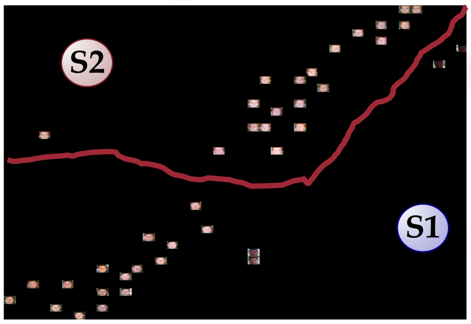


HMAX Space representation on log-polar sampled face images

Foveated face recognition



Uniform resolution



Log-polar mapping

Training	Testing	FF	SRC	MSSRC	VGG	Outer face	Ocular regions	Fusion
Lab[1] light	Dim[2] light	54.48	52.79	47.21	**62.27**	53.15	33.33	54.95
Lab[1] light	Sun[3] light	45.27	51.18	46.15	49.09	94.31	91.87	**95.12**
Dim[2] light	Lab[1] light	25.52	44.18	43.06	50.91	56.76	66.67	**78.38**
Dim[2] light	Sun[3] light	56.80	58.58	60.36	38.18	84.68	73.87	**84.68**
Sun[3] light	Lab[1] light	24.77	17.64	17.64	47.27	48.78	73.17	**73.98**
Sun[3] light	Dim[2] light	**56.01**	51.95	45.85	33.64	48.65	31.53	50.45

Performances are compared with Fisher Faces (FF),
Sparse Representation based Classification (SRC), Mean-Sequence SRC (MSSRC) and VGG deep CNN.

S. Khellat Khiel, A. Lagorio, M. Tistarelli. "**Face Recognition 'On the Move' Combining Incomplete Information**". Proc. of 6th Int.l Workshop on Biometrics and Forensics, June 7,8 2018, Alghero, Italy. IEEE 2018.

S. Khellat Khiel, A. Lagorio, M. Tistarelli. "**Foveated vision for biologically-inspired continuous face authentication**". In A. Rattani Ed. *Selfie Biometrics: Methods and Challenges*, Springer 2019.

Conclusion

- Deep neural architectures provide the current state of the art performance of face recognition *in the wild*
  - The large number of layers requires a huge amount of data for training to reach a stable configuration of the neural connectivity
  - They are sensitive to unexpected changes in the spatial frequencies of the input patterns.
- Simple biological networks allow to perform very complexvisual tasks
- In biological systems **attention** drives **recognition**.
  - A space-variant **scale-space decomposition** of the input signal is performed.
- The **S1C1** neural architecture, derived from the ***HMAX*** model, with face quality, outperforms the deep VGG model.
  - The peripheral area of the face (face outline and hair dressing) proved to be very distinctive for recognition.

Future of Face Recognition

Vision Lab

… Go to ICB, IJCB, CVPR 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**… CNNs are just tools…

New paradigms for learning
- Learn from interactions

17th Int.l Summer School for Advanced Studies on Biometrics for secure authentication:

Vision Lab

"Biometrics, Forensics and Identity Science for Human-Centered Applications"

Alghero, Italy - June, 8 – 12 2020

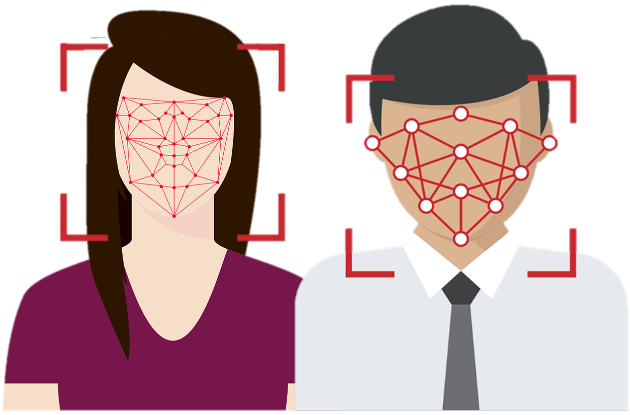
http://biometrics.uniss.it

Contact: tista@uniss.it

IAPR

IDENTITY

IDEMIA augmented identity



IEEE

EURASIP

450 UNISS

Vision Lab

THANK YOU FOR YOUR ATTENTION

…AND PATIENCE