A Unified Neural Scheme for Facial Image Understanding

Pr. Christophe Garcia

Christophe.Garcia@liris.cnrs.fr

Laboratoire d'InfoRmatique en Image et Systèmes d'information

LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/Ecole Centrale de Lyon

Université Claude Bernard Lyon 1, bâtiment Nautibus

43, boulevard du 11 novembre 1918 — F-69622 Villeurbanne cedex

http://liris.cnrs.fr

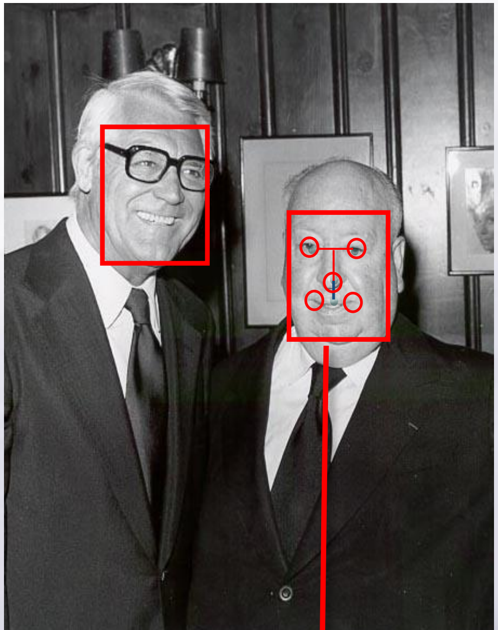
Colleagues

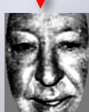
- This work has been performed from 2002 to 2010 in:
- University of Crete, Computer Science Department (2002-2003)
  - with Manolis Delakis
- France Telecom R&D / Orange Labs (2003-2010)
  - with
    - Stefan Duffner,
    - Franck Mamalet,
    - Sébastien Roux

Facial Analysis

- Very active research field
- Challenging and exciting scientific issues
- Numerous possible applications
  - Model-based video coding
  - Image/video indexing and retrieval
  - Surveillance and biometrics
  - Intelligent human-computer interaction

Generic System for Facial Analysis





?

Database

Alfred Hitchcock

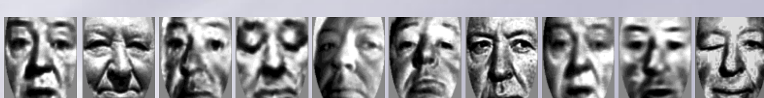


Image / Video

Face Detection

Face counting
Pose estimation
Compression
Advanced HCI

Feature Detection
Face Alignment

Facial feature tracking
Gaze estimation
…
Advanced HCI

Face Coding
Face classification

Surveillance, security
Content-based image indexing
…
Advanced HCI

Identity / Gender / Expression / Emotion

Challenges

Face pattern variability

- size
- orientation, pose
- skin tone
- expressions
- occlusions,…

Scene complexity

- textured background
- uncontrolled illumination
- video blurring effect
- low quality images,…

Principles and Guiding Rules

Generic methods

- directly usable or easily extended to allow analysis of other visual objects
- no use of ad-hoc heuristics (low-level processing or high-level decisions)

Robustness in real-world conditions

- taking into account all the variability factors affecting faces
- without considering a specific application context

Processing speed and compactness

- indexing video and very large image databases
- methods possibly embedded (and paralyzed) on constraint systems

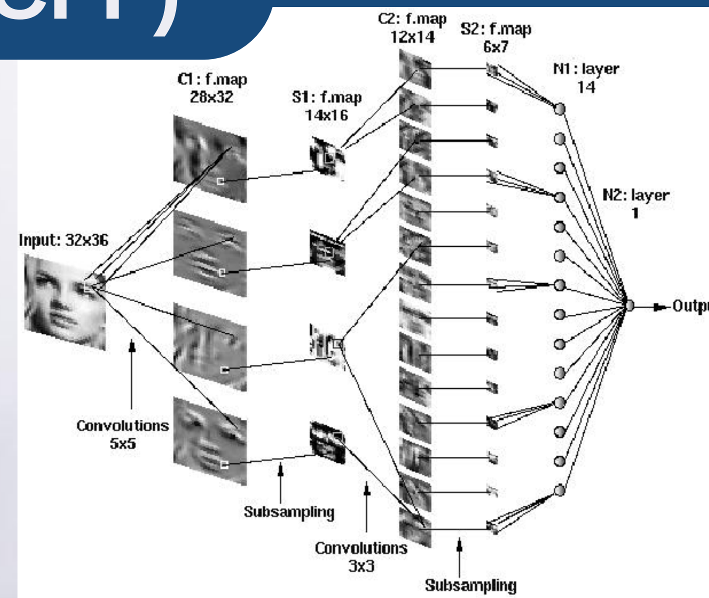
Machine Learning

- automatically infer robust classifiers from many examples, affected by all variability factors
- joint learning of the traditional stages of feature extraction and classification

Convolutional Face Finder (CFF)

Convolutional Neural Networks

- Seminal work of Fukushima (88) and LeCun (90)
- Principles of mammalian visual processing
  - Simple-to-complex cells
  - Cascade of convolutions and subsamplings
  - Local receptive fields
  - Shared weights



Built-in extraction + classification modules

- No hand-crafted extraction or classification rules
- No preprocessing
- Automatically learnt feature extractors and classifiers
- Robustness to pattern variations, occlusions, noise…
- Straightforward fast implementation

Best state-of-the-art results on the CMU database

C. Garcia and M. Delakis, "Convolutional Face Finder: a Neural Architecture for Fast and Robust Face Detection", IEEE transactions on Pattern Analysis and Machine Intelligence (TPAMI), 26(11), November 2004, pp. 1408-1423

CFF Architecture- Learning

Build a function F such that:

F(Image) → 1 if face image

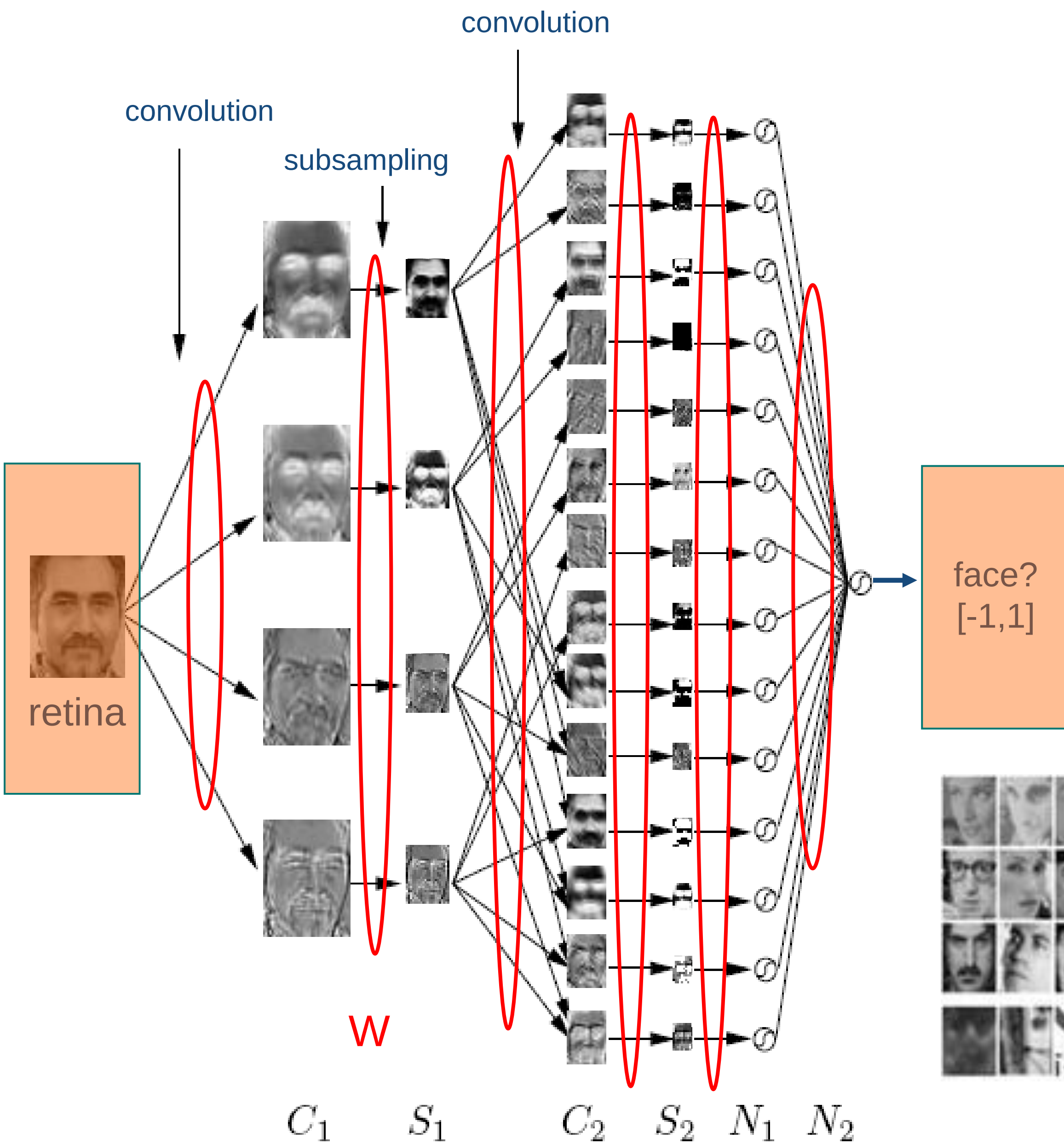
F(image) → -1 otherwise

retina

F

face?
[-1,1]

CFF Architecture- Learning



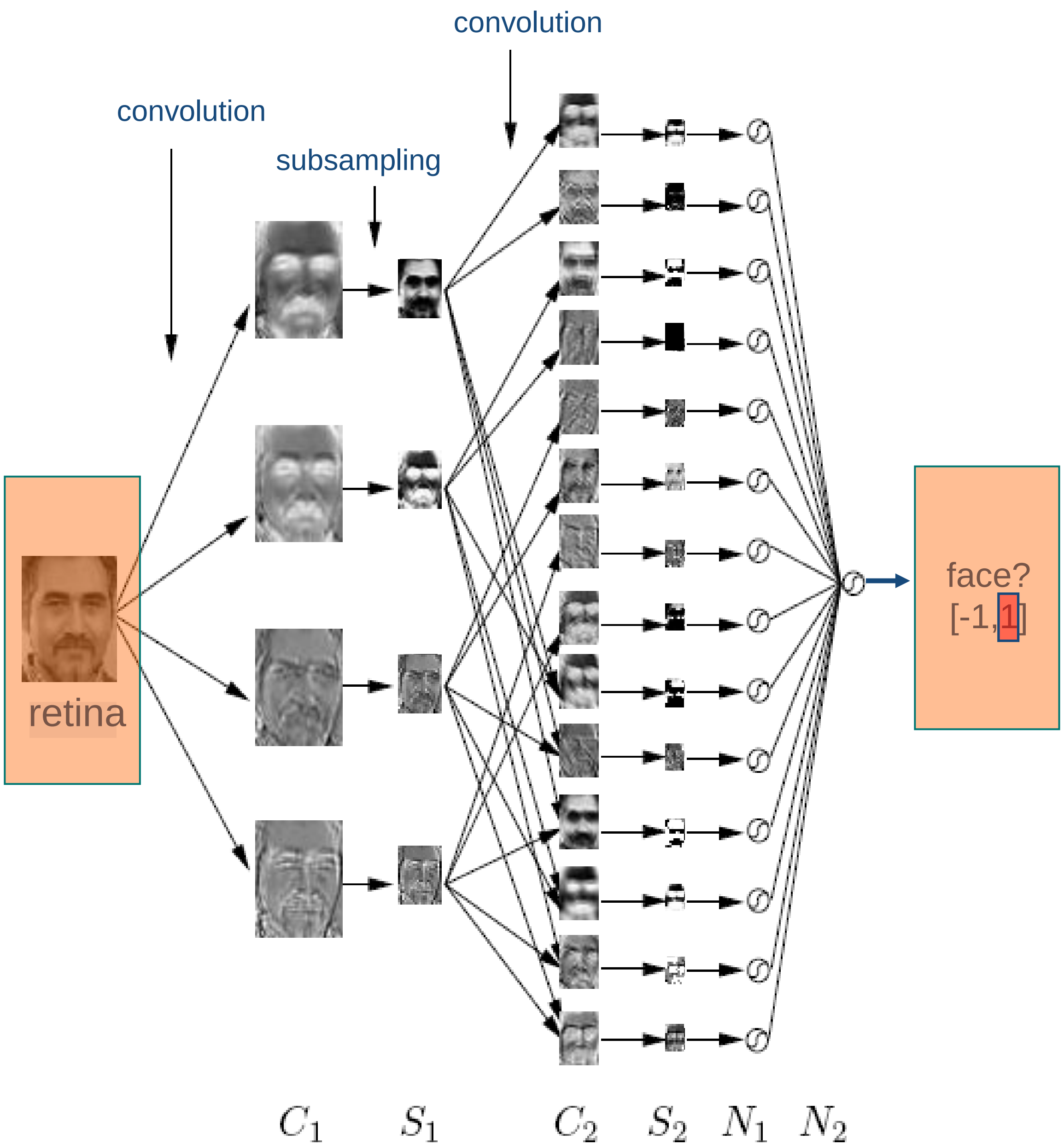
Find synaptic weights W that minimize MSE

$$\sum \left(F(\text{ image, } W \text{) – class} \right)^2$$

on all training examples

Class = 1 for face image
Class = -1 otherwise

CFF Architecture- Learning

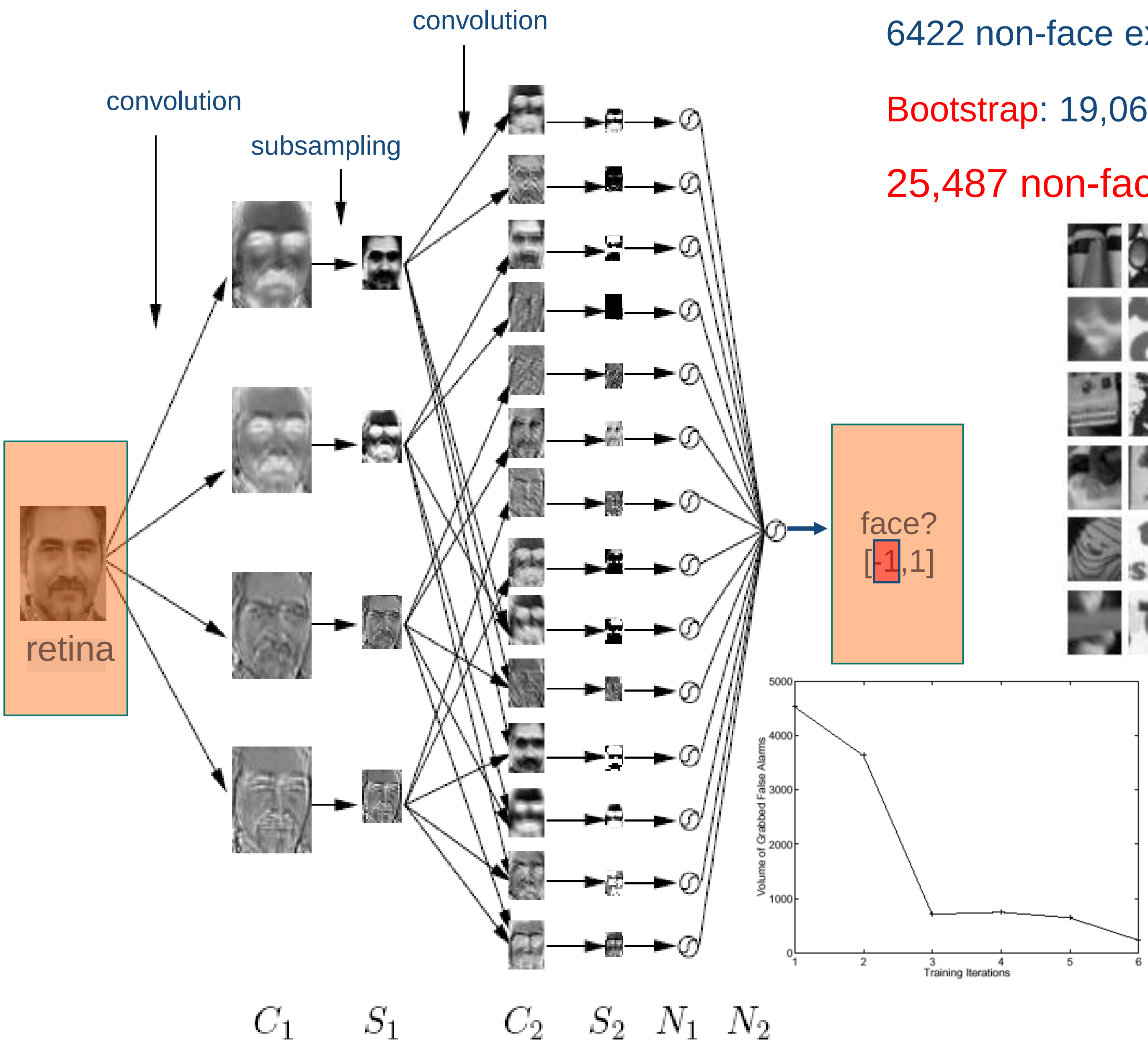


3702 face images (32x36)
(web, scan,...)

Artificially transformed
(rotation, translation, contrast, smoothing, ...)

25,212 face examples

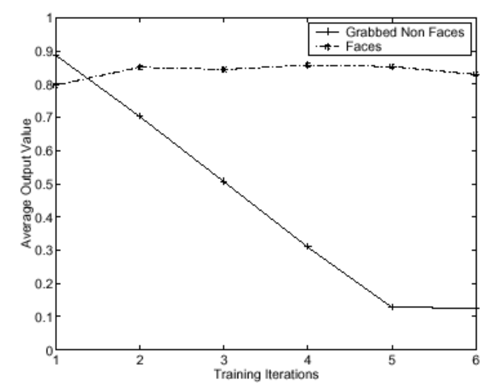
CFF Architecture- Learning



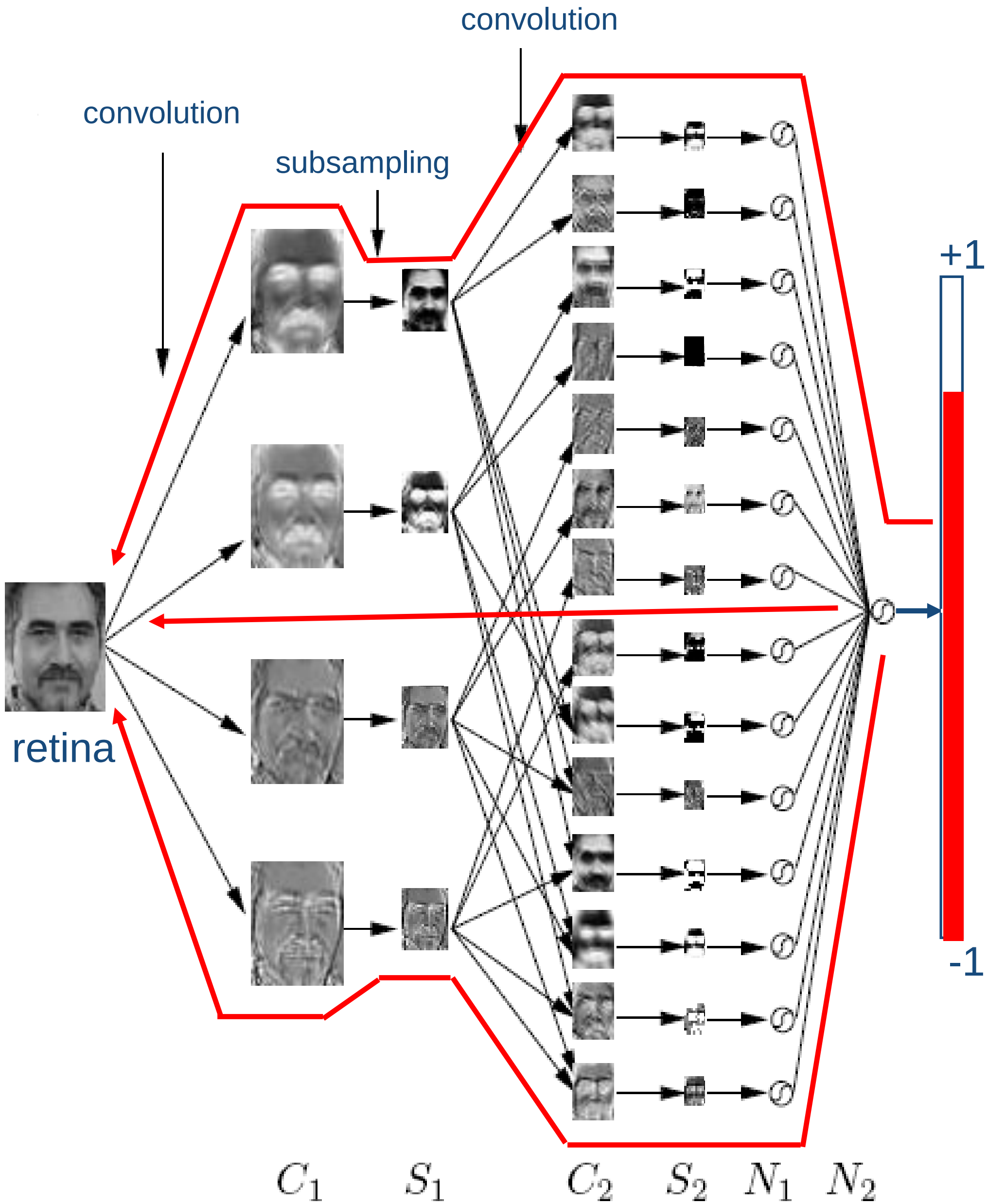
6422 non-face examples

Bootstrap: 19,065 false alarms

25,487 non-face examples



CFF Architecture- Learning



Δ = (output – class)

face?

[-1,1]

Output > 0 → face!

- ✓ 131.475 synaptic weights
- ✓ 951 weights W to learn
- ✓ via gradient backpropagation

Searching for faces…

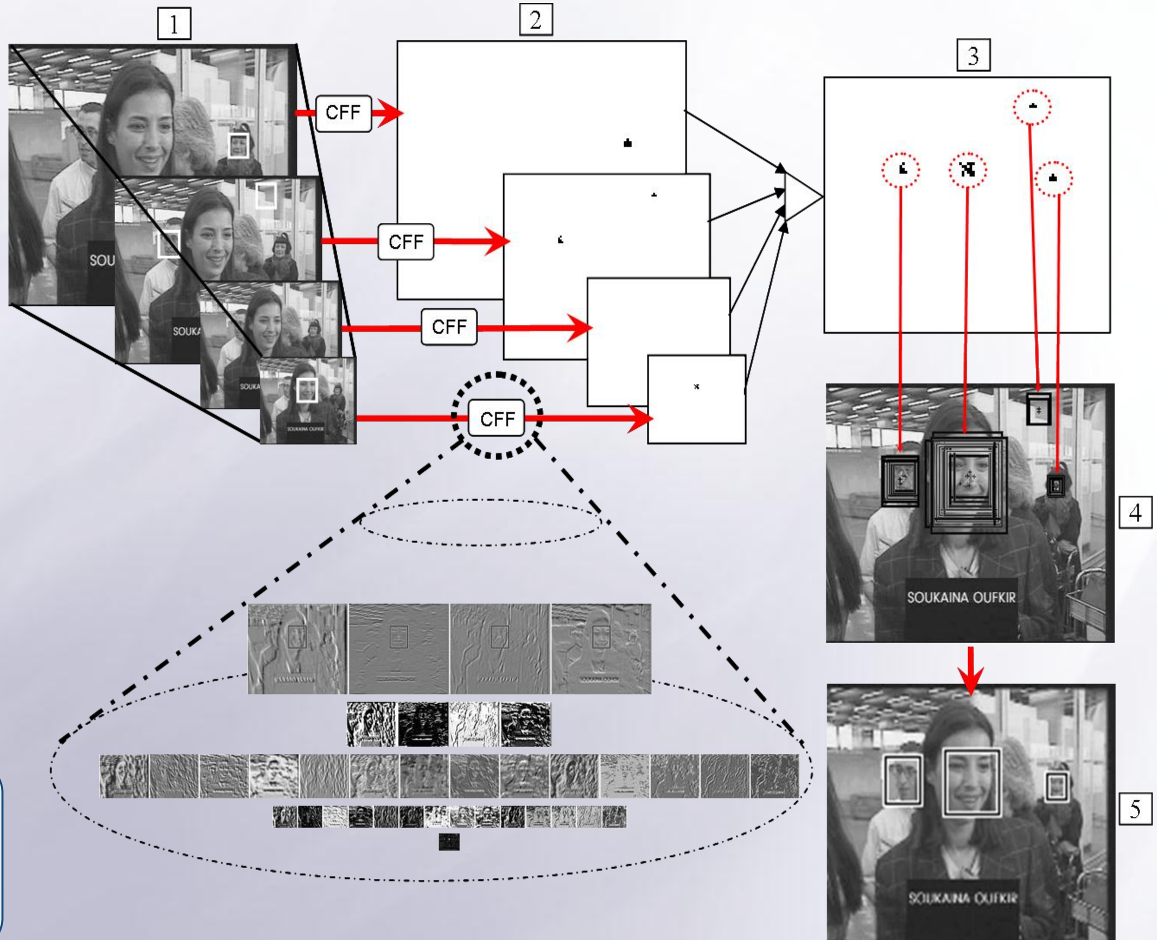
1 Produce a multi-scale Pyramid (factor 1.2)

2 Convolve each pyramid image
Positive answers correspond to candidate face locations

3 Project the candidate faces back to the original image scale and fuse them

4 Apply the neural filter in a finer pyramid centered at each candidate face locations

5 Classify considering the local volume of positive answers (*ThrVol*)



Processing Speed (Pentium IV) : 10 fps (384x288)

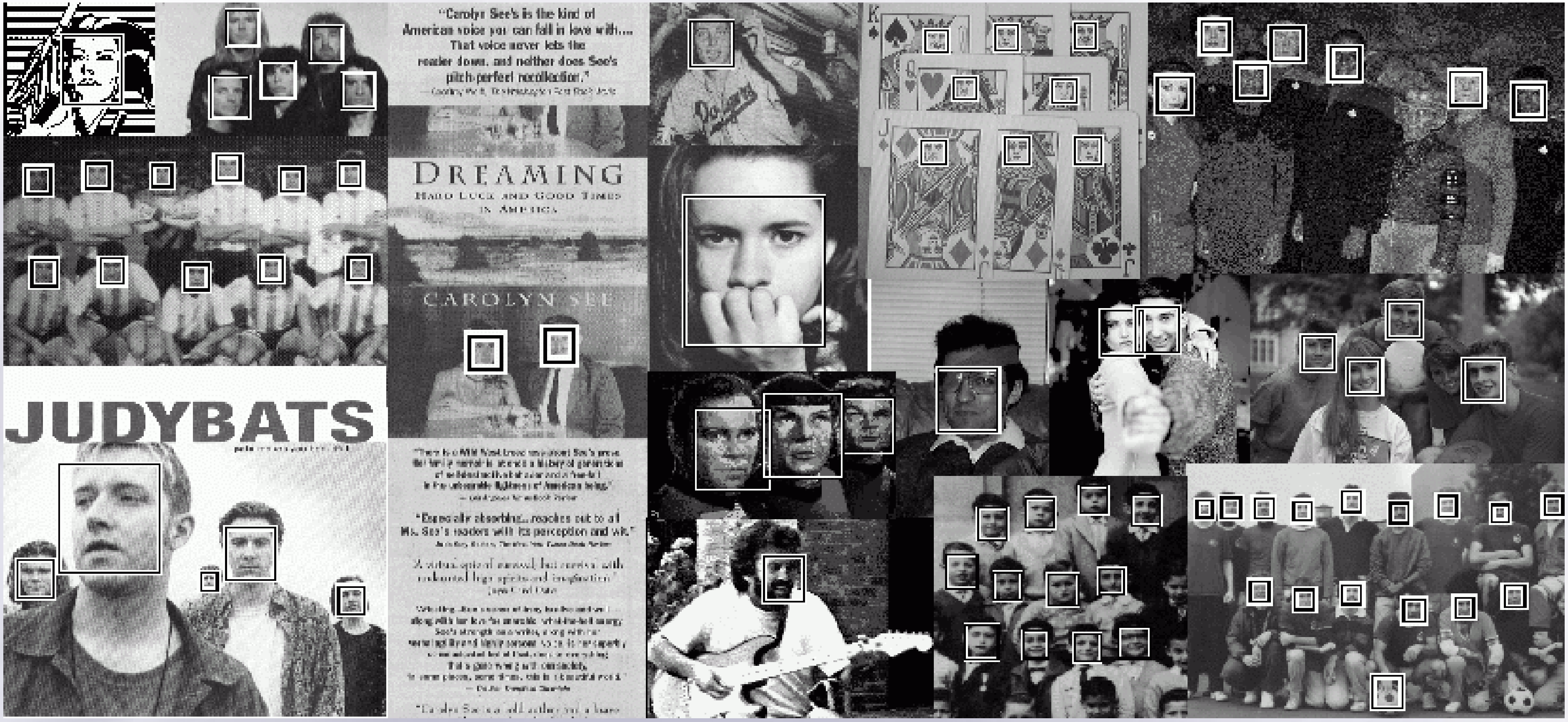
Versus State-of-the-Art Approaches

CMU Face database

Methods	False Alarms				
	0	10	31	65	167
Rowley *et al.* [RBK98a]	-	83, 2%	86, 0%	-	90, 1%
Schneiderman *et al.* [SK98]	-	-	-	94, 4%	
Li *et al.* [LZZ+02]	-	83, 6%	90, 2%	-	-
Viola - Jones [VJ01]	-	76, 1%	88, 4%	92, 0%	93, 9%
Osadchy *et al.* [OLM07]	-	-	-	83, 0%	88, 0%
Garcia - Delakis [GD04]	88, 8%	90, 5%	91, 5%	92, 3%	93, 1%

Results: CMU test set

Rowley *et al*. (CMU) : 130 images, 507 faces: 90.3% / 8 false alarms



Results: Cinema test set

Garcia and Delakis (UoC): 162 images, 276 faces: 90.2% / 3 false alarms

Embedded CFF

Optimization : code, memory, algorithm

- Fixed-point coding of CFF weights (16bits) without perfomance loss
- Fusion of convolution and subsampling operations, parallelization
- Implementation on ARM (Xscale, IMX21), DSP (Starcore) and SPVM 3000 Orange

The first prototype of robust face detection on FPGA (Field-Programmable Gate Array)

- N. Farrugia, F. Mamalet, F. Yang, M. Paindavoine (university of Bourgogne)

F. Mamalet, S. Roux and C. Garcia: *Real-Time Video Convolutional Face Finder on Embedded Platforms*, **EURASIP Journal on Embedded Systems**, 2007

Performances (QCIF images (176x144) **)**

	PXA27x @624MHZ	SC140 @275MHZ	SPVM3000 @200MHZ	Pentium IV @3.2GHZ
Original	0,3 fps			10 fps
Optimized	6,5 fps	13 fps	2,3 fps	58 fps
Tracking	16,5 fps	35 fps	5.0 fps	180 fps

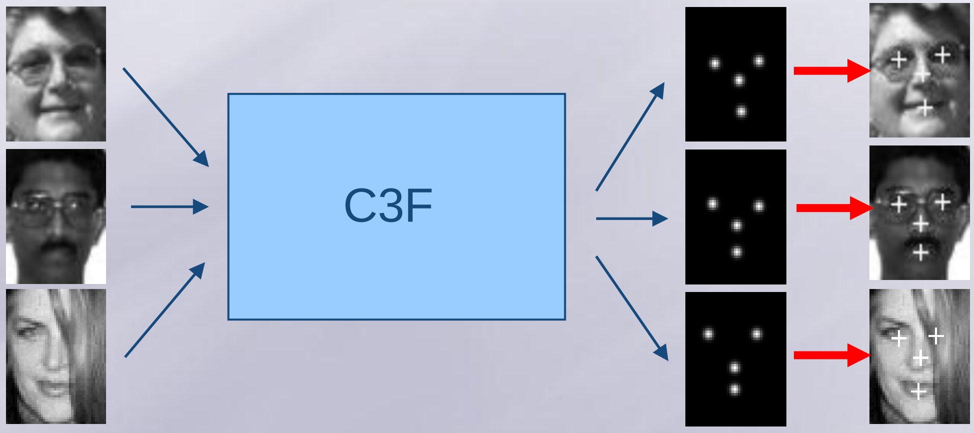
Convolutional Facial Feature Detector (C3F)

- **In non constraint images**
  - noise, resolution, blurring, occlusions

•**Automatically** learns and applies:

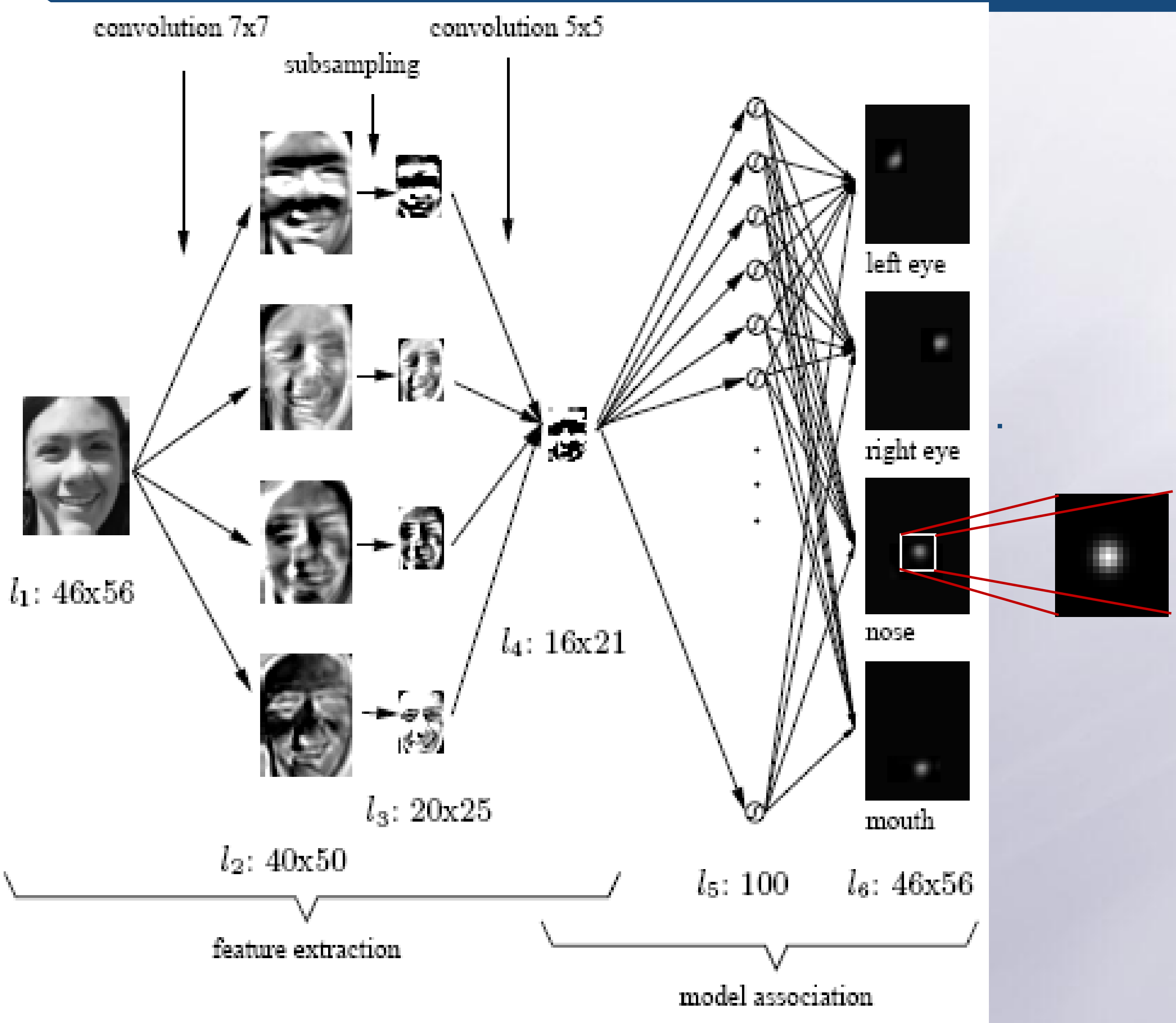
local feature extractors and classifiers
and constraints encoding the face model

- **Robust** to pattern variations, **occlusions**, noise…



S. Duffner, C. Garcia, "Robust Facial Feature Detection via Autoassociative Filters", invited talk, *The Learning Workshop 2005*, Snowbird, Utah, USA, april 2005

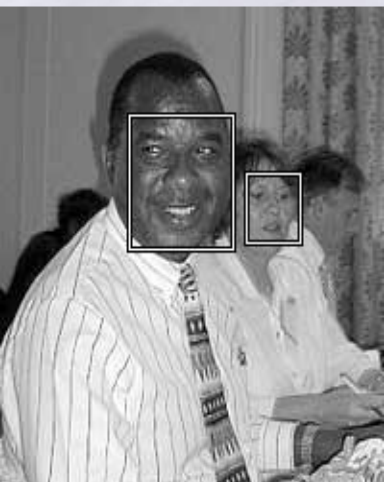
Convolutional Facial Feature Detector (C3F)

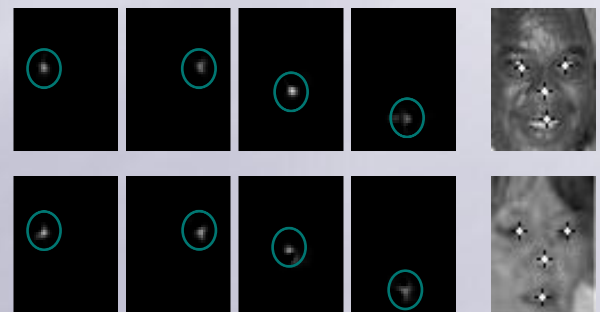


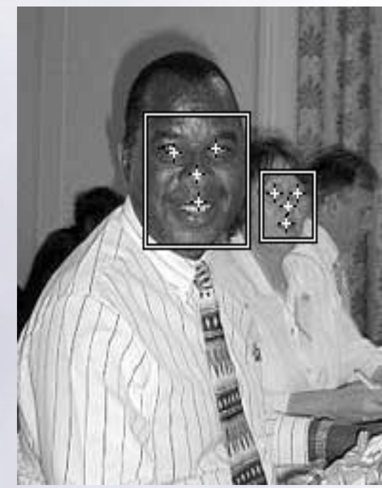
S. Duffner, C. Garcia, "Robust Facial Feature Detection via Autoassociative Filters", invited talk, *The Learning Workshop 2005*, Snowbird, Utah, USA, april 2005

Convolutional Facial Feature Detector (C3F)

Face Detection by CFF → Face extraction and rescaling → C3F + search for global maxima → Backprojected to the original image



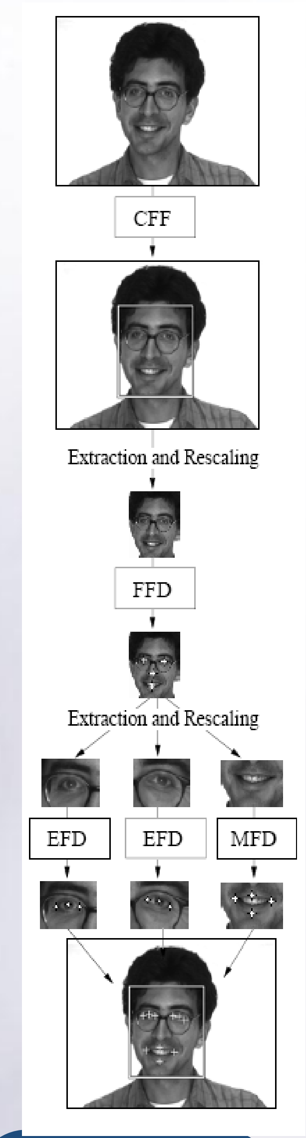




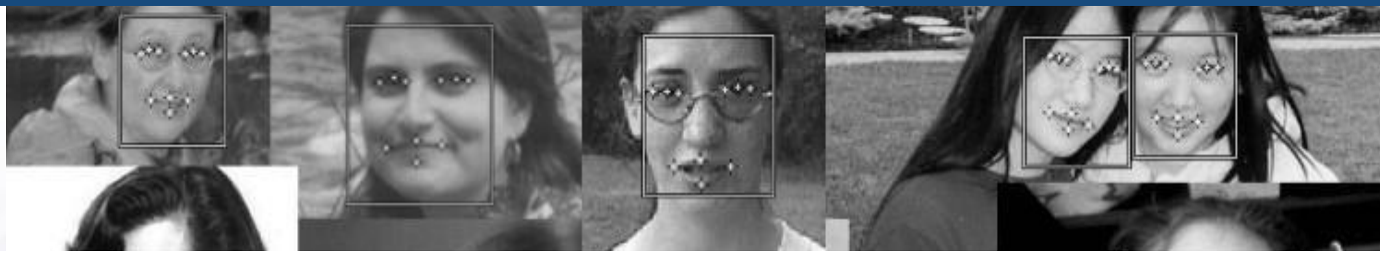
Convolutional Facial Feature Detector (C3F)

- Best State-of-the-Art Precision : mean Localization error < 4% of interocular distance
- Strong robustness to partial occlusions

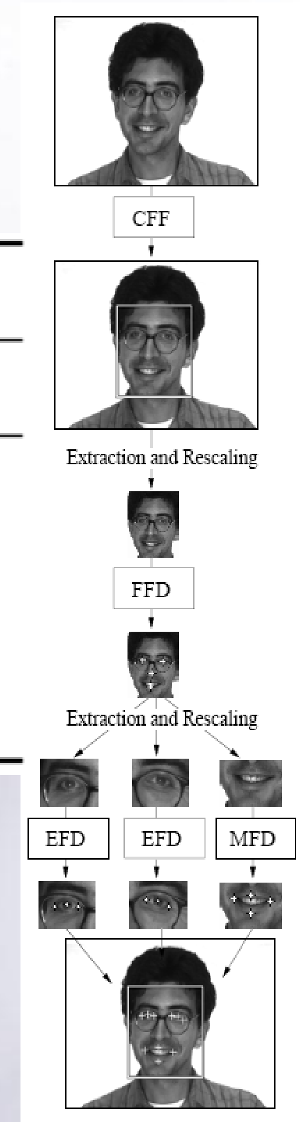
Convolutional Facial Feature Detector (C3F)



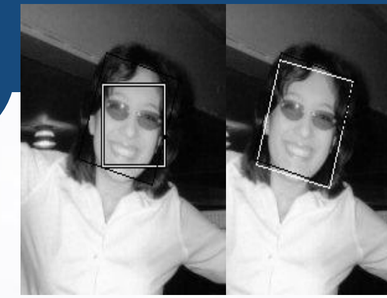
Convolutional Facial Feature Detector (C3F)



Methods	Allowed Error (eye centers)		
	5%	10%	15%
Jesorsky *et al.* [JKF01]	40%	79%	93%
Hamouz *et al.* [HKK^{+}04]	50%	66%	70%
Cristinacce *et al.* [CCS04]	60%	96%	97%
solution (C3F)	79%	92%	98%



Convolutional Face Aligner (CFA)



- To automatically align the face bounding boxes
  - **Conjointly learn affine transforms** (Tx, Ty, rotation, zoom)
- Traing database of 30.000 face images
  - **Translations Tx and Ty: ± 12%, rotation angle α: ± 30 degrees, scale factor Sc: 90% to 110%**

Tx,Ty= 6,6
α = 30
Sc = 0.9

Tx,Ty= 6,0
α = 0
Sc = 1

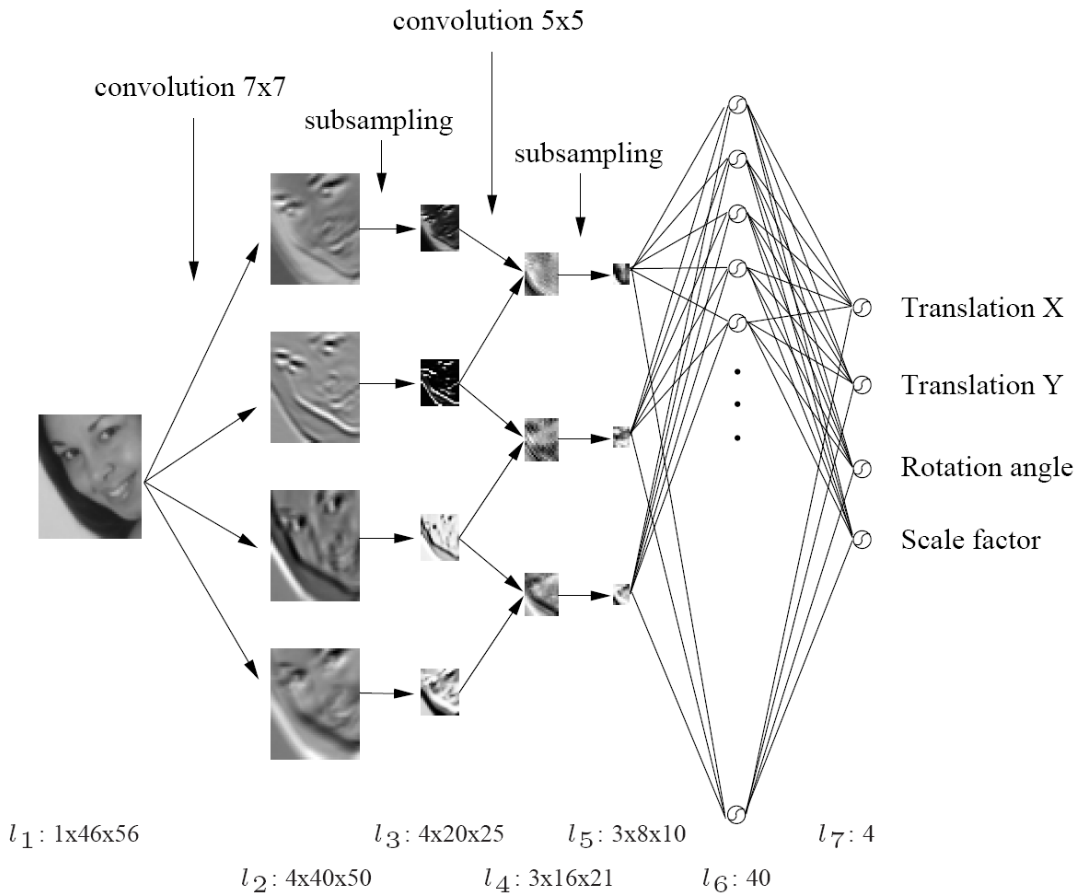
Tx,Ty= 0,6
α = 0
Sc = 1

Tx,Ty= -6,3
α = -20
Sc = 1.1

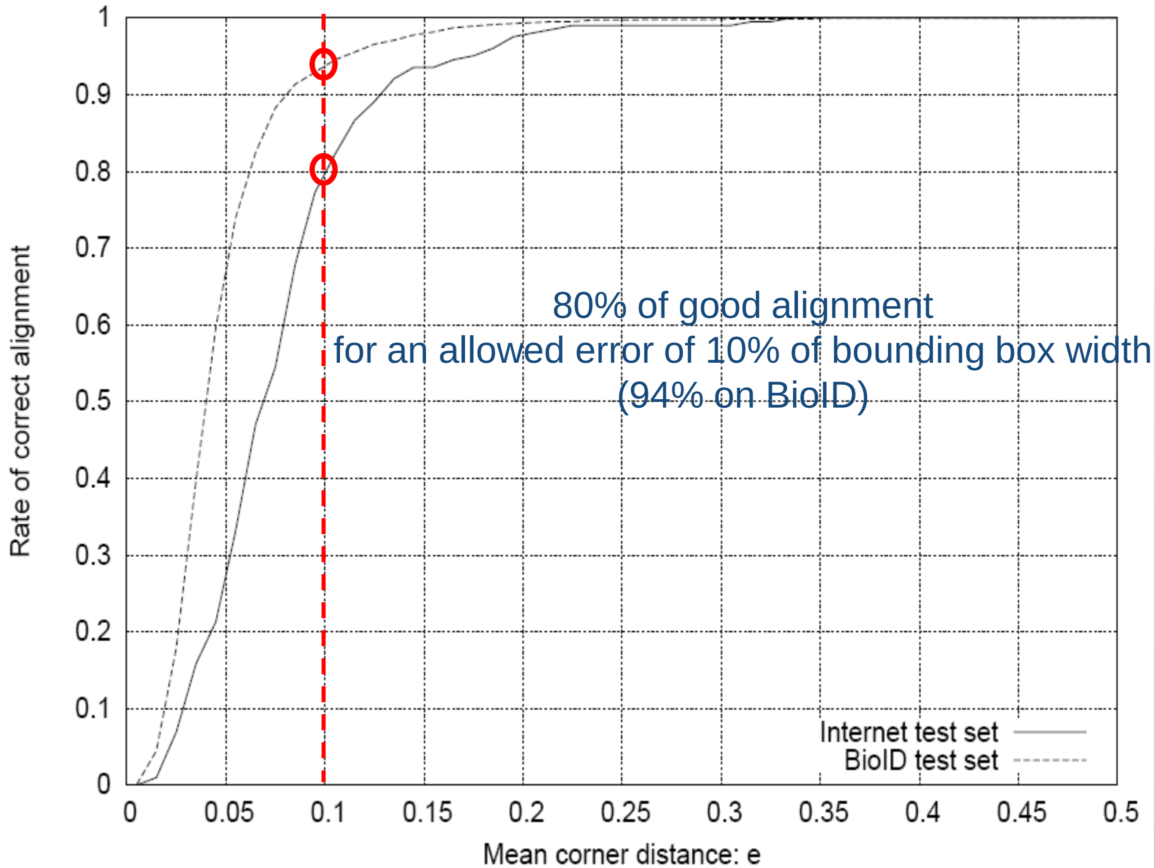
Tx,Ty= 0,0
α = 30
Sc = 1.1

Tx,Ty= 0,0
α = 0
Sc = 0.9

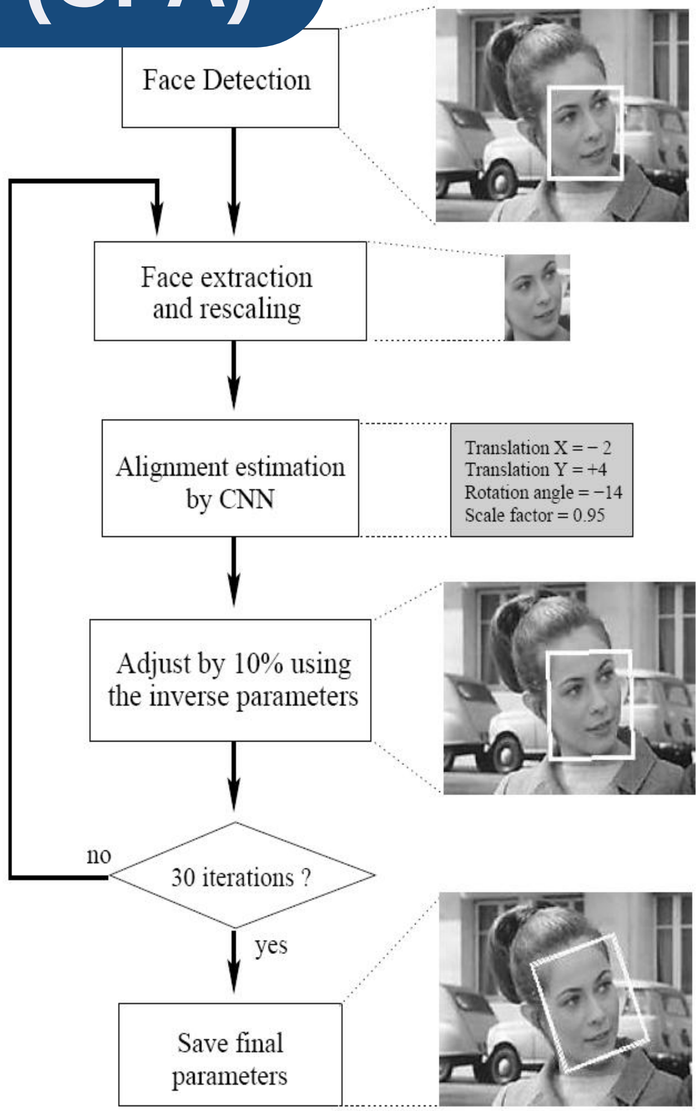
Tx,Ty= 0,0
α = 0
Sc = 1



Convolutional Face Aligner (CFA)

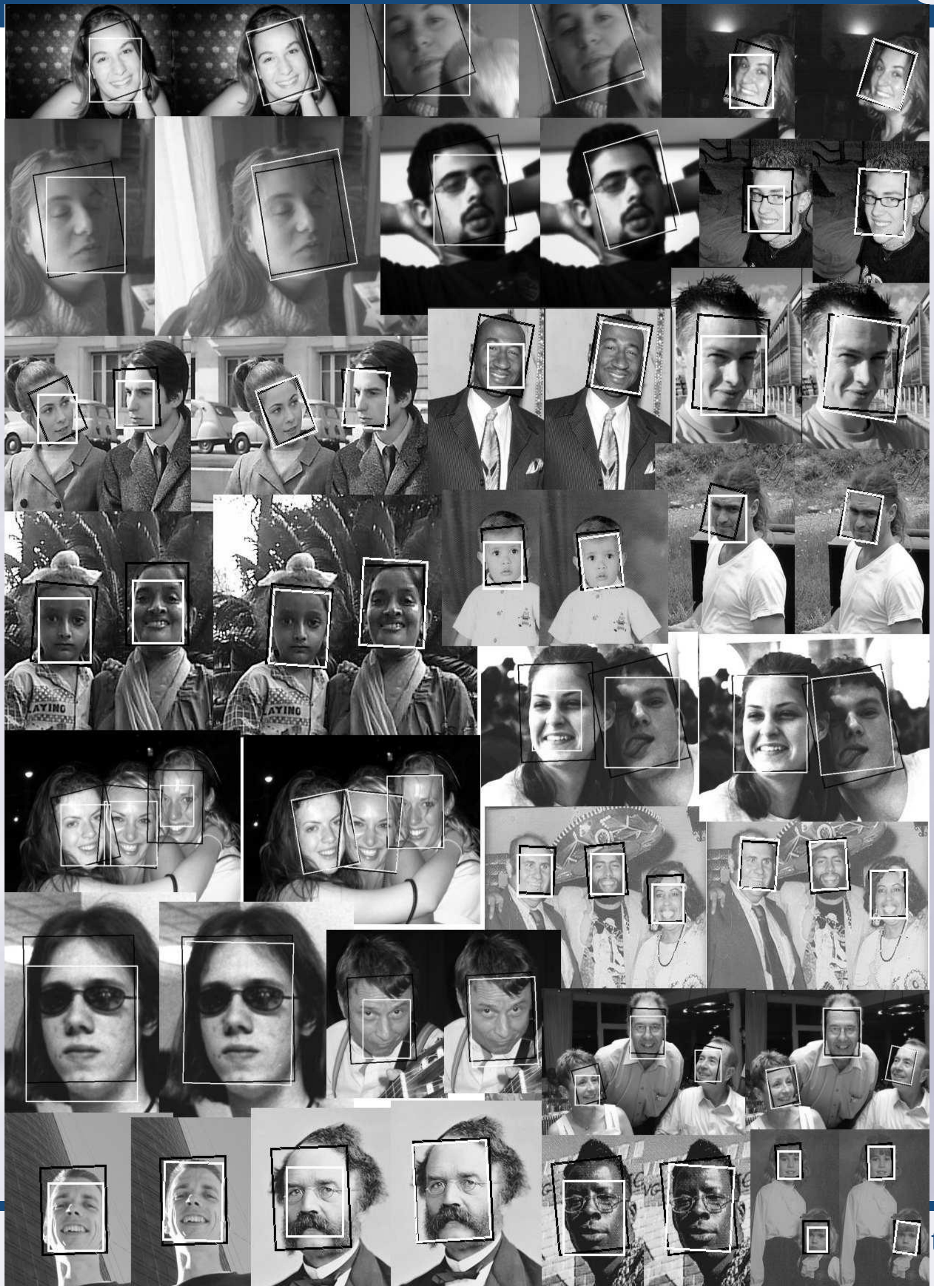


80% of good alignment
for an allowed error of 10% of bounding box width
(94% on BioID)



Duffner S., Garcia C., "Robust Face Alignment Using Convolutional Neural Networks", *International Conference on Computer Vision Theory and Applications (VISAPP 2008)*, Funchal, Portugal, january 2008.

Convolutional Face Aligner (CFA)

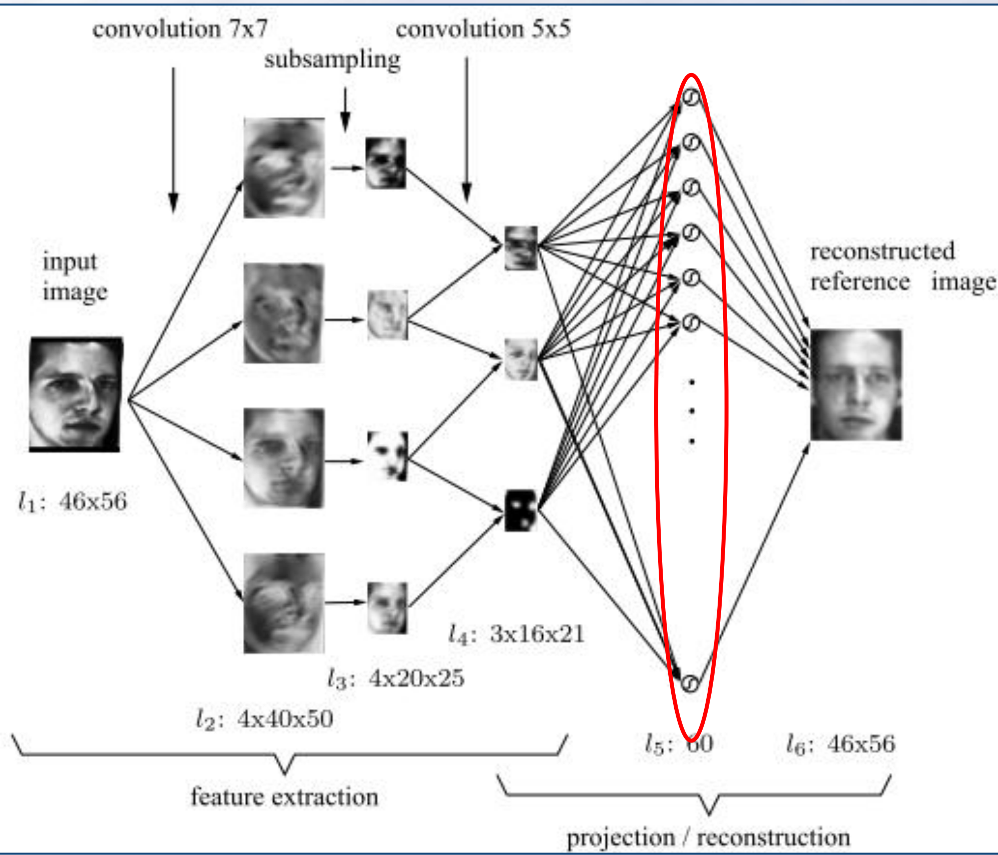


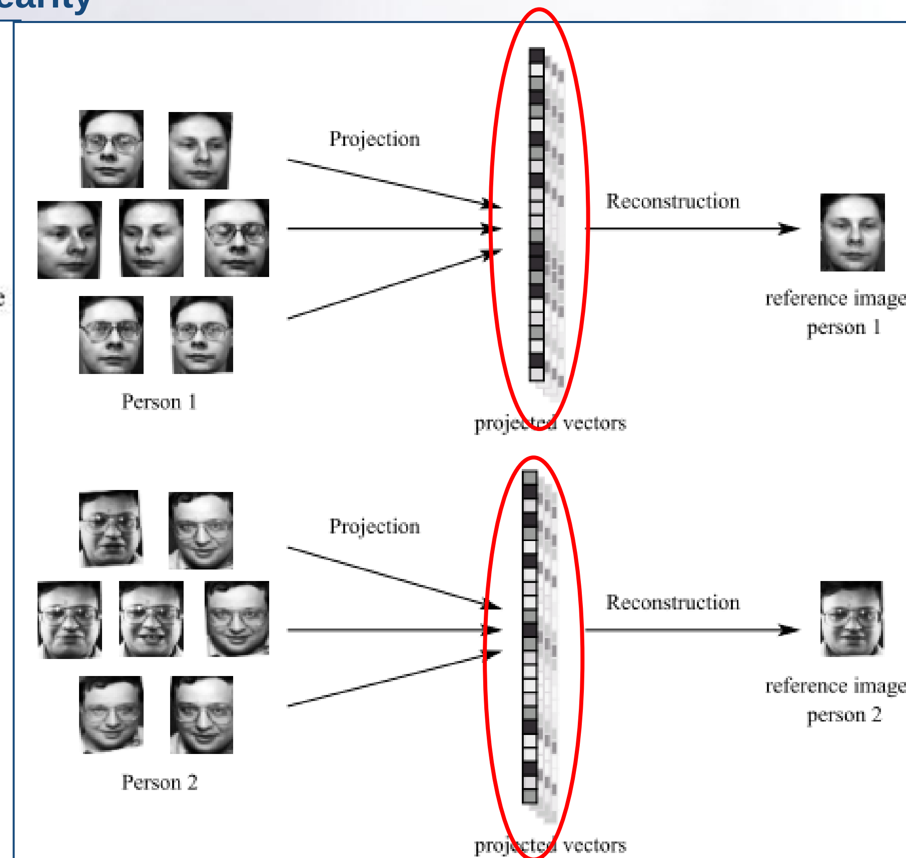
Alignment error < 10% of face width, on more than 3000 images

Convolutional Face Recognizer (CFR)

Non-linear Projection via an auto-associative neural network

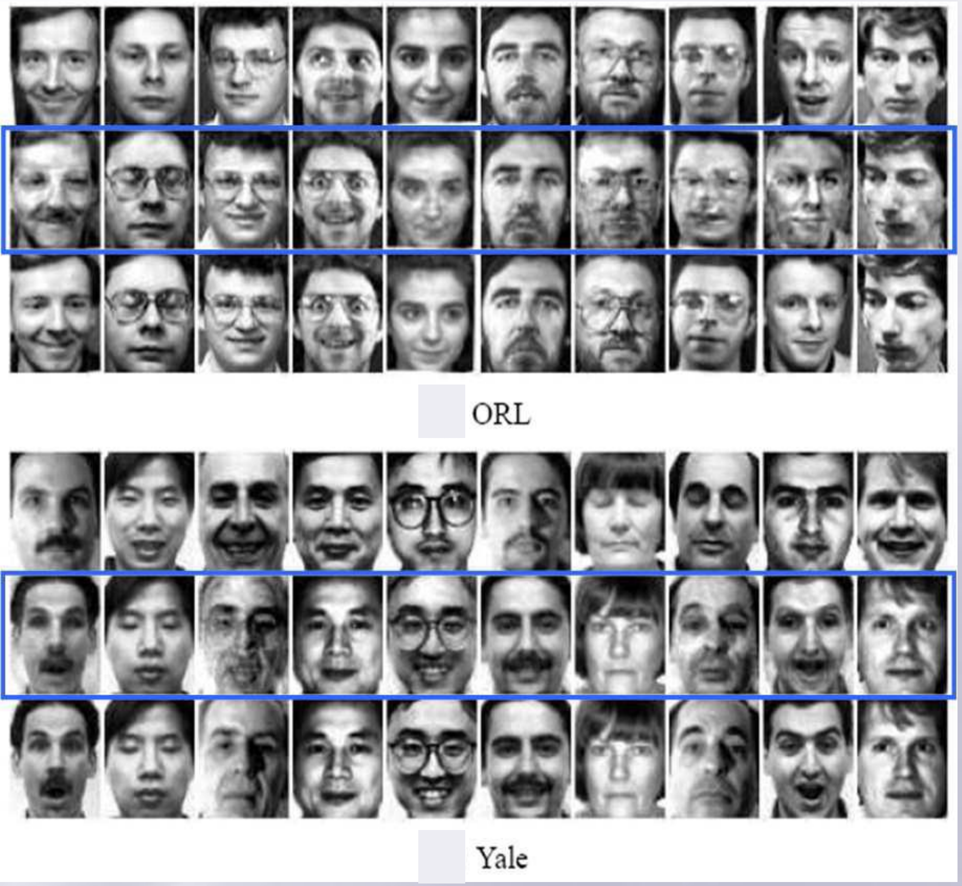
- Signature extraction allowing reconstruction of a reference image per individual
- Minimizing intra-class variance
- Taking into account 2D information and non-linearity





S. Duffner, C. Garcia, Face Recognition Using Non-Linear Image Reconstruction, invited paper, Int. Workshop on Advances in Pattern Recognition (IWAPR), Southampton, 2007

Convolutional Face Recognizer (CFR)



Test images

Reconstructed images

Reference images

92.6 % on ORL database

93.3 % on Yale database

Conclusion

- A set of robust methods for analyzing ***wild*** faces (from detection to identification)

- General principles
  - Propose **generic** approaches for analyzing a **specific object**
  - **Learn** from examples, with **no** use of ad-hoc **heuristics** or specific pre-processing

- In the current research trends
  - Convergence of **signal processing**, statistics and **machine learning**

- Embedded versions have been integrated with success in operational **audio-visual services**

- **Perspectives** : address 3D faces + fusion 2D-3D