# Presentation overview

- Eurécom
- MM dept.
- Current projects in FR
- Short overview of our activities in FR
- More about FR from video

# dept. of Multimedia Communications

**Multimedia Indexing and Filtering**

**Prof. B. Merialdo**

**Affective Computing**

**Prof. C. Lisetti**

Video, Searching

Interface

Image

Audio

**Multimedia Imaging**

**Prof. J-L Dugelay**

**Speech and Audio processing**

**Prof. C. Wellekens**

# Current contracts

- ACI BIOMUL
- TECHNOVISION $IV^2$
- NoE BIOSECURE
- RNRT BIOBIMO

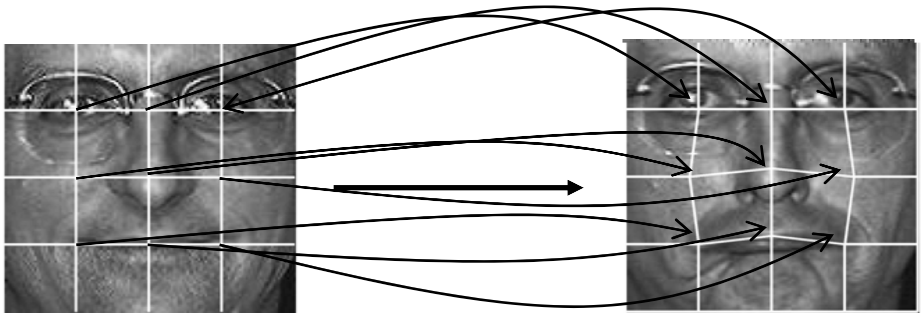
- FT R&D
- Univ. of Salerno

# Towards Novel Facial Biometrics

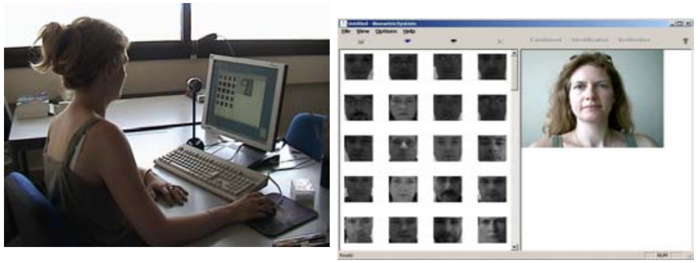
Goal: model the set of possible transformations between face images of the same person

Approximation:

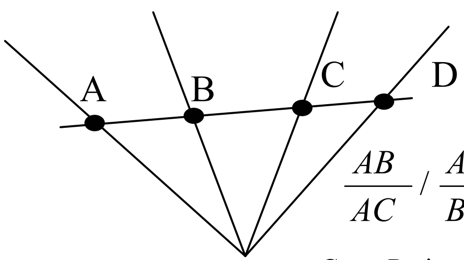
- split global transformation into a set of local transformations
- introduce neighborhood coherence constraints to avoid over-flexibility
- embed the system within the probabilistic framework of a 2-D HMM



Online Face Detection and User Authentication
Tech. demo at ACM Multimedia 2005
Eurécom + FT R&D

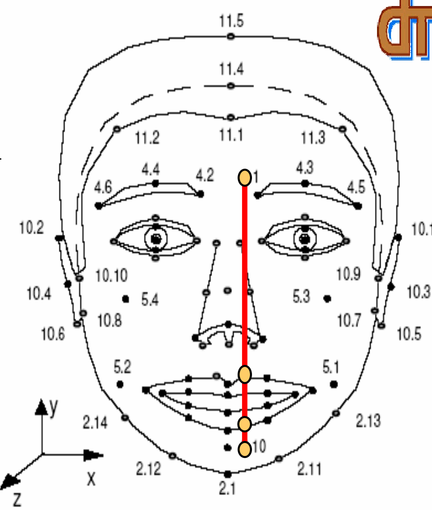


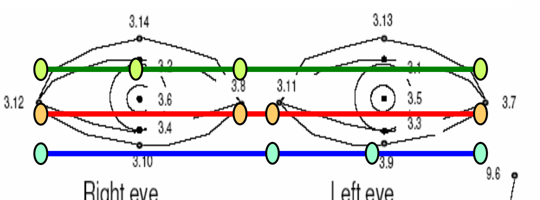
Using Projective Invariants in 3D Face Recognition Ratio




$$\frac{AB}{AC} \Big/ \frac{AD}{BD}$$

Cross Ratio of
4 collinear points




Open question:
Can we obtain better performances by using shape (3D) or shape + apparence (2D).

# Introduction

**Ex. Talking "Hello"**
**Universal**
**Speaker-independent**

**physical**
**behavioral**

**External: conditions of acquisition**
**Internal: level of stress, of …**

**≈ emotion or action ⊕ identity ⊕ context**

**Static (~ appearance)**

**Dynamic (global = pose + local = expressions – mimics)**

**Ex. Smiling**
**Universal**
**Speaker-independent**

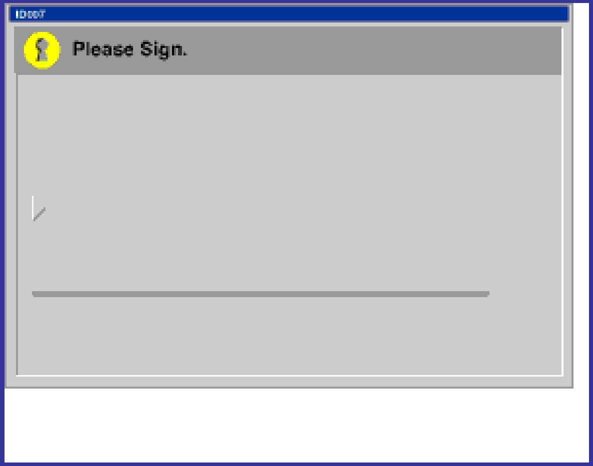
# Biometric systems

- Physical aspects:
  - Face
  - Fingerprint & hand geometry
  - Retina & Iris
  - DNA & veins
- Behavioural aspects:
  - Signature
  - Gait
  - Voice
  - Keystroke dynamics

- Face
  - Considered as physical
    - Appearance
  - Behavioural:
    - Facial mimics:
      - Lips
      - Eyes
      - Mouth
    - Head dynamics:
      - Displacements
      - Pose

# Behavioural approaches

- Recognition using behaviour:
  - Human gait
    - Surprising results
  - Signatures
    - Shape
    - Pressure, inclination
  - Voice
- Question: head dispacements
  - Can be a biometry?

# Facial analysis

- Using images
  - Only physical information
    - Mostly pixel-based systems
- Using videos
  - Physical & behavioural
    - Both possible
  - Spatial information
    - Pixels
  - Temporal information:
    - Motion
    - Evolution
- Video approaches
  - Mostly image generalizations
    - Frames are considered independently as a source of data
    - Lost of temporal information
  - Few real video strategies

# Overview of our system

- Our person recognition system
  - Displacement signals
    - Automatically extracted from videos
    - Work on image plane
  - Tracking of some head features
    - Eyes, nose, mouth
    - Displacements: **d**(t)

# Video analyser

- Head detection
  - Semi-automatic
  - User must click
    - Only on the first frame
- Head tracking
  - Automatic
  - Template matching tecnique
    - RGB color space
    - Simple template update
      - Pixel weighting

# Feature extractor

- Signal representation:
  - Cartesian
- Signal normalization
  - Centering
- Domain:
  - Spatial
- Signal type:
  - Displacements

- Alternatives
  - Polar coordinates
  - Uniform variance
  - Frequence domain
  - Derivatives

# Person classifier

- Bayesian classifier
  - Using a Gaussian Mixture Model (GMM) approximation
    - For modelling likelihoods
    - Trained with head displacements
  - Uniform priors
- Recognition
  - One GMM for each individual
  - Global video score
    - One probability score for displacement
    - Product of individual probabilities
- Verification scores
  - Normalized through Cohort modelling

# Data collection

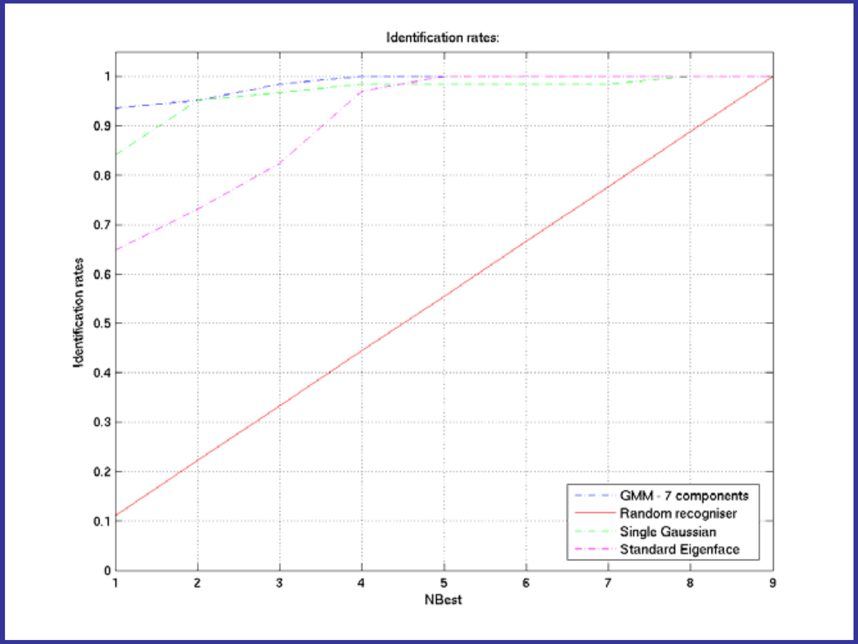
- Video databases
  - Really few standard ones
  - XM2VTSDB
    - Some seconds in high quality
- Minutes needed!
  - For extracting temporal information
    - Training & testing

- Our database
  - 9 individuals
    - Small!
    - 3 minutes per person
  - Low quality
    - Size: 352c x 288r
    - Rate: 24 frames/second
  - Compressed
    - 300 Kbits/second
  - Real case!
    - Real videos

# Comparison: Eigenfaces

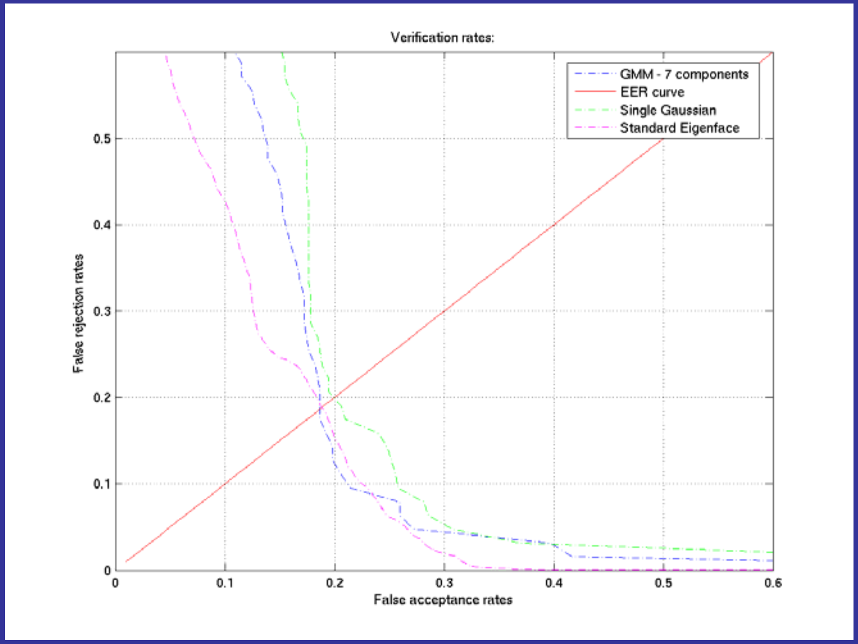
- Original Turk & Pentland
  - Distances in face space
    - Dimension: 25
  - Same individuals for:
    - Space computation
    - Enrollment & testing
- Implementation details
  - Light preprocessing
  - Loose face normalization
    - Head position & size

# Identification scores




- Good identification scores!
  - 94.7% & 98.4%
  - Better than single Gaussian...
    - 84.1% & 96.8%
  - ...and Eigenface
    - 64.8% & 82.4%

# Verification scores




- Not so bad verification results
  - EER: 0.191
  - Better than single Gaussian...
    - EER: 0.206
  - ... and close to Eigenface
    - EER: 0.188

# Future works

- Multimodality: physical
  - Include physical analysis
    - Pixel-based image/video techniques
- Multimodality: behavioural
  - Eye blinking, mouth opening, lips...
    - Suited database?
- Tracking: robust
  - More precise signals
  - Although...
    - Our case: few errors
- Features: more local
  - Capture the instant motion
    - Big learning space... enough data?

# General Conclusion & perspectives

Being able to identify people thanks to head motion can be a good test to evaluate our ability to estimate pose in videos;

But pose is one parameter only among many others that are required to fully understand a video of a speaker.

- Unsolved Problem...
  Absence of suited database

- Short term
  - Limited expression & emotion
  - Performed by head motion
    - E.g Yes/No movement
- Medium term
  - Basic Local analysis
    - Eye blinking
    - Speaking or not
    - Mouth motion