# An Introduction to Automatic Face Recognition using Statistical Models

## Winter Workshop on Bayesian Biometrics for Forensics (BBfor2)

**Dr Sébastien Marcel**

Senior research scientist

www.idiap.ch/~marcel

Idiap Research Institute

Martigny, Switzerland

www.idiap.ch

December 14, 2011

# Outline

1. The Idiap Research Institute
2. Biometric Person Recognition at Idiap
3. Biometric Person Recognition (Biometrics for short)
4. Face Processing
5. Pre-requisites
6. Face Recognition
7. Face Recognition using Statistical Models

# The Idiap Research Institute

## Independent not-for-profit research institute

- Founded in 1991
- Around 100 collaborators ($>$ 25 countries)
- Budget: around 10 MCHF
- Centre du Parc in Martigny (2300 m2)
- 37 research programs ($>$ 130 publications/year)
- Affiliated with EPFL (joint development plan) and University of Geneva
- Accredited (and co-funded) by the Swiss Federal Government, State and City, as "part of the "ETH Strategic Domain"
- Host institution of Swiss National Centre of Competence in Research on "interactive multimodal information management" (IM2)

# Idiap Research Activities

## Perceptual and Cognitive Systems

Speech processing, Natural language understanding and translation, Vision and scene analysis, Computational cognitive science

## Human and Social Behavior

Face-to-face communication analysis, Mobile media analysis, Social media analysis

## Information Interfaces and Presentation

User interfaces, Multimedia information systems

## Biometric Person Recognition

Face and speaker recognition, Mobile biometry, Emerging biometrics, Spoofing

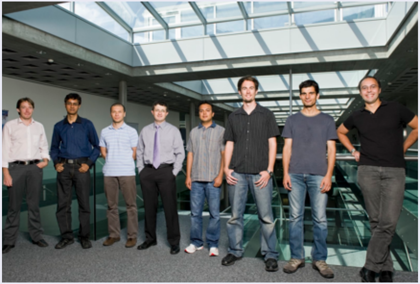
## Machine Learning

Statistical and neural network based ML, Very large families of heuristics, Computational efficiency, targeting real-time applications, Online learning

# Outline

# Research Team

## Current team (4 PostDocs and 4 PhD students)



Dr Christopher McCool, Anindya Roy, Cosmin Atanasoaei, Dr Sébastien Marcel, Venkatesh Bala Subburaman, Dr Roy Wallace, Laurent El Shafey, Dr André Anjos, Ivana Cingovska (not present on the picture) and Manuel Günther (not present on the picture).

## Former PhD students

Dr Anindya Roy (EPFL PhD student 2011), Dr Guillaume Heusch (EPFL PhD student 2009), Dr Agnes Just (EPFL PhD 2006), Dr Yann Rodriguez (EPFL PhD 2006), Dr Fabien Cardinaux (EPFL PhD 2005).

# Projects

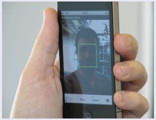
## European and International Projects

- Bayesian Biometrics for Forensic (EU Marie Curie project)
- Trusted Biometrics under Spoofing Attacks (EU FP7 TABULA RASA project)

  Goal: Develop and evaluate countermeasures to direct (spoofing) attacks in face recognition (printed photos, replayed videos, ...), speech, fingerprint, iris, gait, ...
  Partners: IDIAP* (CH), UOULU (FI), UAM (SP), USOU (UK), UNICA (IT), EURECOM (FR), CASIA (CN), STARLAB (SP), MORPHO (FR), KEYLEMON (CH), BIOMETRY (CH), CSSC (IT).
- Mobile biometry (EU FP7 MOBIO project)

  Goal: Develop and evaluate face and speaker recognition algorithms for mobile devices.
  Partners: IDIAP* (CH), UMAN (UK), UNIS (UK), LIA (FR), BUT (CZ), UOULU (FI), VISIDON (FI).

Face recognition on the iPhone 4



Free App on the AppStore (FaceOnIt)

## Projects in Switzerland

- Swiss National Science Foundation: MULTI, GMface
- Hasler Foundation: CONTEXT
- CTI: Replay, A-vision

# Activities

## Research Topics

- Face recognition: face detection, facial feature localization, face identification/verification
- Speaker recognition
- Multi-Modal fusion: early and late fusion

## Research Problems

- robust-to-illumination features for face recognition
- robust-to-noise binary features for speaker recognition
- statistical models for face recognition
- online (unsupervised) model adaptation
- spoofing

# Outline

1. The Idiap Research Institute
2. Biometric Person Recognition at Idiap
3. Biometric Person Recognition (Biometrics for short)
   - Definition
   - Challenges
   - A General Biometric System
4. Face Processing
5. Pre-requisites
6. Face Recognition

# Biometrics

## Definition

Biometric person recognition (or Biometrics for short) refers to the process of automatically recognizing a person using:

- distinguishing behavioral patterns: gait, signature, keyboard typing, lip movement, ...
- physiological traits: face, voice, iris, fingerprint, hand geometry, electroencephalogram (EEG), electrocardiogram (ECG), ear shape, vein, ...

## A truly inter-disciplinary research field

Biometrics offer a wide range of challenging fundamental and concrete problems in signal (image/audio) processing, computer vision, pattern recognition and machine learning.

# Biometrics

## Challenges

- In most real life applications, the environment is not known *a-priori* and the system should be fully automatic. A biometric system has to deal with variabilities such as:
  - Lighting variations,
  - Background noise (audio),
  - Alignment (head pose),
  - Deformation (facial expressions) or Corruption (occlusion),
  - Aging.
- Typically these variability are classified as:
  - *extra-personal* variabilities: variations in appearance between different identities,
  - *intra-personal* variabilities: variations in appearance of the same identity.

# Biometrics

## A General Biometric System

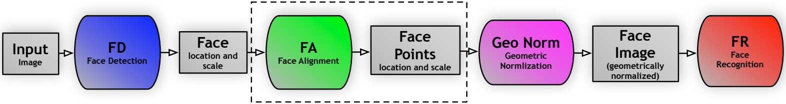
1. data capture,
2. segmentation (such as face detection, silence detection),
3. geometric normalization (such as alignment),
4. sample normalization (such as illumination normalization),
5. feature extraction,
6. enrollment and classification:
   - *enrollment*: building a template (or model) of an identity,
   - *classification*: identity recognition from models and features.

# Outline

# Overview

## 3 main tasks

1. detection (and tracking)
2. alignment
3. recognition



# Applications

## Security

- Video surveillance (public places, restricted areas),
- "Biometrics"
  - access control (computer or mobile device log-in, building gate control, digital data protection),
  - forensics (police investigations),

## Multimedia information management

content-based image/video indexing and retrieval.

## Human computer interaction

video games, personal settings identification.

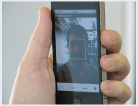
# Biometrics Applications

## Transactions and services applications

- micro payment services,
- phone card reloading,
- remote purchase,
- telephone banking, ...


## Embedded applications

- PIN code replacement,
- lock/unlock device,
- personal data protection.



# Content-based Image/Video Indexing and Retrieval Applications

## Multimedia information management applications

- automatic annotation of personal photos and home videos,
- automatic indexing by image/video content in addition to (Exif) metadata (time, location via GPS tags, ...),
- tools to facilitate search in large (photo) collections (QBE, QBT, ...).

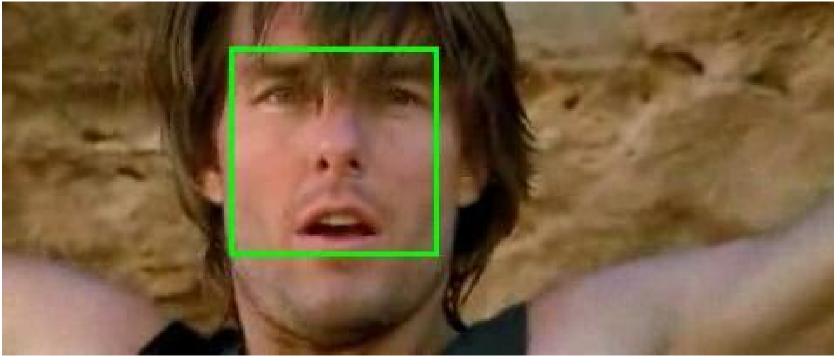
# Content-based Image/Video Indexing and Retrieval Applications

## Examples

- Nevenvision: face processing (acquired by Google in 2006), i
- Riya (www.riya.com): face processing and text recognition,
- Like (www.like.com): Visual shopping (acquired by Google in 2009),
- PolarRose (www.polarrose.com): face processing (acquired by Apple in 2010),
- PittPatt (http://www.pittpatt.com/): face processing (acquired by Google in 2011),
- Google (www.google.com) with Picasa (picasa.google.com),
- Apple (www.apple.com) with iPhoto'09 (www.apple.com/ilife/iphoto).

# Example

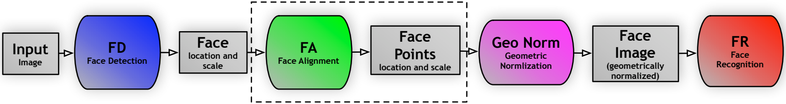
How to do this will NOT be explained during this presentation :-(



# Face Processing

## 3 main tasks

1. detection (and tracking)
2. alignment
3. **recognition**



# Outline

# Pre-requisites

**Who knows ?**

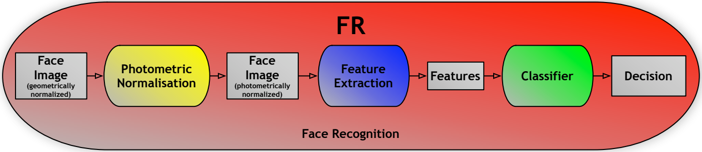
PCA, LDA, NN, MLP, SVM, Gabor wavelet, LBP, FFT, DCT, GMM, HMM, BN, EER, HTER, ROC, DET, EPC.

# Outline

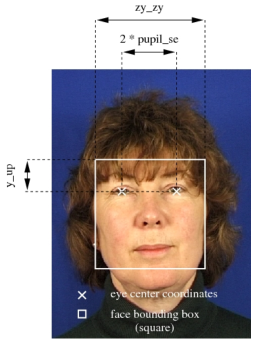
# Face Recognition

## A General Face Recognition System

- ○ segmentation (face detection and alignment),
- ⋆ geometric normalization (face cropping),
- ⋆ sample normalization (illumination normalization),
- ⋆ feature extraction,
- ⋆ enrollment and classification:
  - *enrollment*: building a template (or model) of an identity,
  - *classification*: identity recognition from models and features.
- ○ score normalization (Z-norm, T-norm).



# Geometric Normalization




## Goal

1. align eye centers,
2. compensate for in-plane rotation.

# Illumination Normalisation

## Adini et al. PAMI 1997

image variations due to lighting changes are more significant than the one due to different personal identities !

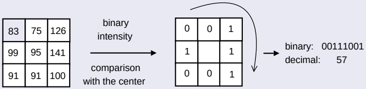
## Many techniques

- Image enhancement: Histogram Equalization, gain correction,
- Lambertian reflectance, Illumination cone,
- Land's Retinex Theory: $I(x, y) = L(x, y) \cdot R(x, y)$
  - Single-Scale and Multi-Scale Retinex,
  - Self-Quotient Image (SQI).
- Gross and Brajovic,
- Tan and Triggs (T&T),
- Local Binary Patterns

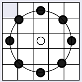
# Local Binary Patterns

## Definition

- Original LBP operator: 3x3 kernel which summarizes the local spatial structure of an image.




- $LBP_{P,R}$: $P$ equally spaced pixels on a circle of radius $R$.

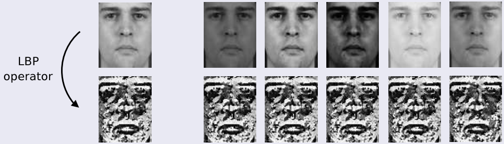


- $LBP^{u2}_{P,R}$: only uniform patterns (at most two bitwise 0 to 1 or 1 to 0 transitions)
- Other variants: Improved LBP, Extended LBP.

# Local Binary Patterns

## Properties

- Very low computational cost
- Powerful texture descriptor
- Invariant to monotonic gray-scale transformation



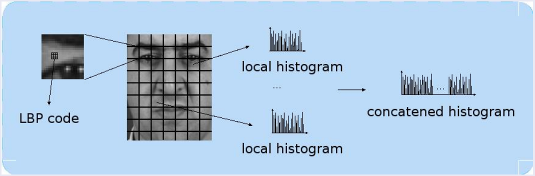

LBP is becoming a very popular technique due to its simplicity and its interesting monotonic grayscale invariant property.

# Feature Extraction

## Many techniques, to name a few:

- Principal Component Analysis (PCA) so called Eigenfaces,

- Linear Discriminant Analysis (LDA) so called Fisherfaces,
- Gabor transform,
- Local Binary Pattern histograms,




- Discrete Cosine Transform (DCT),
- and any combination of the above techniques.

# Classification

## Classification

- Classification consists of attributing a label to the input data and differs according to the specific task (closed or open set identification, verification)
- All system provides a score $\Lambda(X; \Theta_I)$ corresponding to an opinion on the probe feature set $X$ to be the identity $I$.
  - verification: the label is true (client) or false (impostor)
  - closed set identification: the label is the identity
  - open set identification: the label is the identity or *unknown*

# Classification

## Classification

- verification: given a threshold $\tau$, the claim is accepted when $\Lambda(X;\Theta_I) \geq \tau$ and rejected when $\Lambda(X;\Theta_I) < \tau$
- closed set identification: we can recognize identity $I^*$ corresponding to the probe feature set $X$ as follows

$$I^* = \arg\max_I \Lambda(X;\Theta_I) \tag{1}$$

- open set identification: the recognized identity $I^*$ corresponding to the probe is found as follows

$$I^* = \begin{cases} \text{unknown} & \text{if } \Lambda(X;\Theta_I) < \tau \ \forall\ I \\ \arg\max_I \Lambda(X;\Theta_I) & \text{otherwise} \end{cases} \tag{2}$$

# Classification

## To summarize

1. determine a feature extraction to produce $X$
2. choose a model of the features $X$ (with parameters $\Theta$)
3. compute a score $\Lambda(X;\Theta)$ for classification

# Modelling and Scoring

## $\Theta_l = X_l^e$

- Modelling: no modelling, the model $\Theta_l$ is actually $X_l^e$ or a set of $X_l^e$ (template).
- Scoring $\Lambda(X^t; \Theta_l)$: $L_1$ norm, $L^2$ norm (Euclidean), Cosine similarity, *Elastic Graph Matching* [Lades:1993] and *bunch graph* [Wiskott:1997] using Gabor filters and a labeled graph.

## $\Theta_l$ are derived from $X_l^e$

- Modelling: $\Theta_l$ represents the parameters from a statistical model[a].
- Scoring $\Lambda(X^t; \Theta_l)$: computing a probability $P(X^t|\Theta_l)$.

[a] We will consider here only statistical generative models as opposed to discriminative models such as MLP or SVM.

# Outline

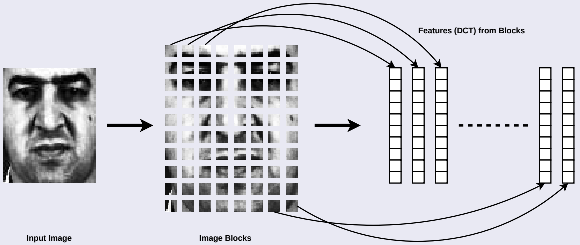
# Statistical Model based Approaches

## Discriminant vs Generative

- *Discriminant models* such as Multi-Layer Perceptrons or Support Vector Machines:
  - training dataset of $l$ pairs $(X_i, y_i)$ where $X_i$ is a vector containing the pattern, while $y_i$ is the class of the corresponding pattern,
  - we train one model per identity, $y_i$ being coded as $+1$ for patterns corresponding to this identity and as $-1$ for patterns corresponding to an other identity,
  - Drawback: difficulty to train them with a small training dataset.
- **Generative models** estimate the likelihood of the face image being a specific identity using models representing identities.

# Feature extraction

## using local (parts-based) features




*the face image is decomposed into parts (blocks)*

## as opposed to holistic features

*the face image is processed as a whole*

# Simple to complex models

$\Lambda_I(X;\Theta_I) = P(X|\Theta_I)$

- Gaussian Mixture Models (GMM) [Sanderson:2003],
- 1D Hidden Markov Models (1D-HMM) [Eickeler:2000],
- Pseudo-2D Hidden Markov Models (P2D-HMM) [Nefian:1999],
- Bayesian Networks (BNface) [Heusch:2009].

# Statistical Generative Models

## Enrollment

- using the Maximum Likelihood (ML) criterion via the Expectation Maximization (EM),

  A lot of data is required to properly estimate model parameters.

- using a well trained generic (non-person specific) model as the starting point for ML training,

  ML training still produces poor models.

- using Maximum *a Posteriori* (MAP) training [Gauvain:1994] (also called *MAP adaptation*).

  This approach derives a client specific model from a generic model and circumvents the lack of data problem.

# Statistical Generative Models

## Scoring (log-likelihood ratio)

- Let us denote the parameter set for client $C$ as $\lambda_C$ and the parameter set describing a generic face (non-client specific) as $\lambda_{\overline{C}}$.
- Given a claim for client $C$'s identity and a set of $T$ feature vectors $X = \{\vec{x}_t\}_{t=1}^{T}$ supporting the claim (extracted from the given face).
- We find an opinion on the claim using the log-likelihood ratio: $\Lambda(X) = \log P(X|\lambda_C) - \log P(X|\lambda_{\overline{C}})$

# Statistical Generative Models

## Scoring and Decision

- We find an opinion on the claim using $\Lambda(X) = \log P(X|\lambda_C) - \log P(X|\lambda_{\overline{C}})$ where:
    - $P(X|\lambda_C)$ is the likelihood of the claim coming from the true claimant
    - $P(X|\lambda_{\overline{C}})$ is the likelihood of the claim coming from an impostor.
- The generic face model (also called *world model* or *Universal Background Model*) is trained with data from many people.
- The decision is then reached as follows: given a threshold $\tau$, the claim is accepted when $\Lambda(X) \geq \tau$ and rejected when $\Lambda(X) < \tau$.

# Gaussian Mixture Model

## GMM

- The likelihood of a set of feature vectors is given by

$$P(X|\lambda) = \prod_{t=1}^{T} P(\vec{x}_t|\lambda) \tag{3}$$

where

$$\begin{aligned} P(\vec{x}|\lambda) &= \sum_{k=1}^{N_G} m_k \mathcal{N}(\vec{x}|\vec{\mu}_k, \Sigma_k) && (4) \\ \lambda &= \{m_k, \vec{\mu}_k, \Sigma_k\}_{k=1}^{N_G} && (5) \end{aligned}$$

# Gaussian Mixture Model

## GMM

- The likelihood of a set of feature vectors is given by

$$P(X|\lambda) = \prod_{t=1}^{T} \sum_{k=1}^{N_G} m_k \mathcal{N}(\vec{x}|\vec{\mu}_k, \Sigma_k) \tag{6}$$

- $\mathcal{N}(\vec{x}|\vec{\mu}, \Sigma)$ is a $D$-dimensional Gaussian density function with mean $\vec{\mu}$ and diagonal covariance matrix $\Sigma$.
- $N_G$ is the number of gaussians and $m_k$ is the weight for gaussian $k$ (with constraints $\sum_{k=1}^{N_G} m_k = 1$ and $\forall\ k : m_k \geq 0$).

# Gaussian Mixture Model

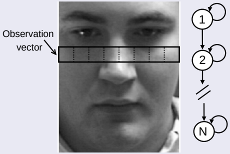
## GMM

- Generally, each feature vector $X$ describes a different part of the face (a local approach).
- We note that the spatial relations between face parts are lost (the position of each part does not matter in the likelihood estimation).
  - Advantage: this lead to a robustness to imperfect localization of the face,
  - Drawback: discriminatory information carried by spatial relations is lost. Fortunately, there is a simple way to restore a degree of spatial relations.

# 1D-Hidden Markov Model

## 1D HMM

- The face is represented as a sequence of overlapping *rectangular* blocks from top to bottom of the face:




- The model is characterized by the following:
  - $N$, the number of states in the model,
  - The state transition matrix $A = \{a_{ij}\}$,
  - The state probability distribution $B = \{b_j(\vec{x}_t)\}$.

# 1D-Hidden Markov Model

## $N$ the number of states in the model

each state corresponds to a region of the face;
$S = \{S_1, S_2, \ldots, S_N\}$ is the set of states. The state of the model at row $t$ is given by $q_t \in S$, $1 \leq t \leq T$, where $T$ is the length of the observation sequence (number of rectangular blocks).

## The state transition matrix $A = \{a_{ij}\}$

The topology of the 1D-HMM allows only self transitions or transitions to the next state:

$$a_{ij} = \begin{cases} P(q_t = S_j | q_{t-1} = S_i) & \text{for } j = i,\ j = i+1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

## The state probability distribution $B = \{b_j(\vec{x}_t)\}$

$$b_j(\vec{x}_t) = P(\vec{x}_t | q_t = S_j) \tag{8}$$

# 1D Hidden Markov Model

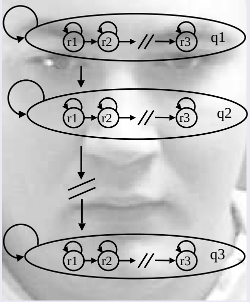
## 1D HMM

- Compared to the GMM approach the spatial constraints are much more strict, mainly due to the rigid preservation of horizontal spatial relations (e.g. distance between the eyes).
- The vertical constraints are more relaxed, though they still enforce the top-to-bottom segmentation (e.g. the eyes have to be above the mouth).
- The relaxation of constraints allows for a degree of vertical translation and some vertical stretching (caused, for example, by an imperfect face localization).

# Pseudo-2D Hidden Markov Model

## 2D HMM

- Emission probabilities of the HMM (now referred to as the "main HMM") are estimated through a secondary HMM (referred to as an "embedded HMM"):




- The states of the embedded HMMs are in turn modeled by a mixture of gaussians.

# Pseudo-2D Hidden Markov Model

## 2D HMM

- The degree of spatial constraints present in the P2D-HMM approach can be thought of as being somewhere in between the GMM and the 1D-HMM approaches. While the GMM approach has no spatial constraints and the 1D-HMM has rigid horizontal constraints, the P2D-HMM approach has relaxed constraints in both directions.
- However, the constraints still enforce the left-to-right segmentation of the embedded HMMs (e.g. the left eye has to be before the right eye), and top-to-bottom segmentation (e.g. like in the 1D-HMM approach, the eyes have to be above the mouth). The relaxed constraints allow for a degree of both vertical and horizontal translations, as well as some vertical and horizontal stretching of the face.

# Database

## BANCA

- BANCA (English) database with realistic conditions: *controlled*, *degraded* and *adverse*

# Database

## BANCA

- 12 recording sessions over several months, in different conditions and with different cameras,
- high variability in illumination, pose, resolution, background and quality of the camera.

# Protocols

## BANCA Protocols

- 7 distinct configurations that specify which images can be used for training and testing:

| Test Sessions | Train Sessions 1 | Train Sessions 5 | Train Sessions 9 | Train Sessions 1,5,9 |
|---|---|---|---|---|
| C: 2-4<br>I: 1-4 | Mc | | | |
| C: 6-8<br>I: 5-8 | Ud | Md | | |
| C: 10-12<br>I: 9-12 | Ua | | Ma | |
| C: 2-4,6-8,10-12<br>I: 1-12 | P | | | G |

Matched Controlled (Mc), Matched Degraded (Md), Matched Adverse (Ma), Unmatched Degraded (Ud), Unmatched Adverse (Ua), Pooled test (P) and Grand test (G).

# Performance Measure

## Verification errors

- A verification system makes two types of errors:
  - False Acceptance (FA) when the system accepts an impostor,
  - False Rejection (FR) when the system refuses a true claimant.
- The performance is measured in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR)
- FAR and FRR are related (decreasing one increases the other),
- To aid the interpretation of performance, FAR and FRR are often combined using the Half Total Error Rate (HTER):

$$\text{HTER} = \frac{\text{FAR} + \text{FRR}}{2} \tag{9}$$

# Experiment Results (manual)

| System | Protocol | | | |
|---|---|---|---|---|
| | Mc | Ud | Ua | P |
| PCA | 9.5 | 20.9 | 20.8 | 18.4 |
| LDA/NC | 4.9 | 16.0 | 20.2 | 14.8 |
| SVM | 5.4 | 25.4 | 30.1 | 20.3 |
| GMM *ML* | 12.9 | 28.9 | 26.0 | 22.9 |
| GMM *init* | 12.8 | 29.7 | 28.3 | 23.8 |
| GMM *MAP* | **8.9** | **17.3** | **20.9** | **17.0** |
| 1D-HMM *ML* | 9.1 | 17.8 | 17.1 | 15.9 |
| 1D-HMM *init* | 9.1 | **15.6** | 17.4 | **14.7** |
| 1D-HMM *MAP* | **6.9** | 16.3 | **17.0** | **14.7** |
| P2D-HMM *ML* | 9.0 | 19.0 | 18.0 | 17.5 |
| P2D-HMM *init* | 8.6 | 16.5 | 19.2 | 17.0 |
| P2D-HMM *MAP* | $*$ **4.6** | $*$ **15.3** | $*$ **13.1** | $*$ **13.5** |

# Experiment Results (auto)

| System | Protocol | | | |
|---|---|---|---|---|
| | Mc | Ud | Ua | P |
| PCA | 22.4 | 29.7 | 33.7 | 29.0 |
| LDA/NC | 22.6 | 25.4 | 27.1 | 25.2 |
| SVM | 19.7 | 30.4 | 33.2 | 27.8 |
| GMM *ML* | 16.7 | 33.3 | 33.3 | 27.7 |
| GMM *init* | 19.8 | 35.0 | 35.1 | 29.7 |
| GMM *MAP* | **9.5** | **21.0** | **24.8** | **19.5** |
| 1D-HMM *ML* | 21.0 | 28.8 | 29.5 | 27.0 |
| 1D-HMM *init* | 21.3 | 30.1 | 31.4 | 28.1 |
| 1D-HMM *MAP* | **13.8** | **25.9** | **23.4** | **21.7** |
| P2D-HMM *ML* | 12.1 | 25.2 | 26.9 | 22.3 |
| P2D-HMM *init* | 13.5 | 24.6 | 26.5 | 22.5 |
| P2D-HMM *MAP* | $*$ **6.5** | $*$ **15.9** | $*$ **14.7** | $*$ **14.7** |

# Discussion

## Discussion

- Maximum *a Posteriori* (MAP) training circumvents the lack of data problem,
- Systems that utilize rigid spatial constraints between face parts (such as PCA and 1D-HMM based systems) are easily affected by face localization errors,
- Systems which have relaxed constraints (such as GMM and P2D-HMM based), are quite robust.

# Research Directions

## Challenges

- illumination normalization is still an issue,
- dealing with faces with multiple poses is still a problem as well,
- exploiting multiple faces in videos is problematic (scalability),
- model adaptation (or template update).

## Directions

- Local Binary Patterns could be used to reduce the effect of illumination,
- Bayesian Networks provide elegant generative models able to fuse multiple cues,
- UBM-GMM Super-Vectors techniques, such as Inter Session Variability (ISV) modelling and Joint Factor Analysis (JFA), have shown recently to outperform UBM-GMM.

# The End

Thanks for listening
Questions ?