Face Detection and Bi-Modal Authentication

Sébastien Marcel

marcel@idiap.ch

IDIAP Research Institute
Martigny, Switzerland

http://www.idiap.ch

Outline

- Introduction and Background
- Face Detection + demo
- Bi-Modal Authentication + demo
- Discussions

Introduction and Background

- Technologies developed at IDIAP of potential interest:
  - Face detection and tracking,
  - Biometric authentication (Face $+$ Speech),
- Demonstrators:
  - Bi-modal authentication demo since 2003 (versions α and 1),
  - Face detector added in 2004 (version 2),
  - Software: based on Torch library (`www.torch.ch`)
  - Interface: QT
  - Platform: Linux
- Time-line:

A step forward ...

- A step forward: move IDIAP demonstrators (running Linux) to Win XP,
- Partnership with HEVs (Technical University of Valais),
- TheArk/IdeArk is funding a project between IDIAP and HEVs,
- The goal of the project is to develop two applications:
  - A face tracker (fast face detector)
  - A Bio-Login and its User Manager (BLUM)

Partners & Persons involved

- IDIAP (4 persons):
  - S. Marcel (Face detection and authentication),
  - J. Mariethoz (Speech authentication),
  - Y. Rodriguez (Face detection),
  - F. Cardinaux (Face authentication)
- TheArk: F. Bagnoud,
- IdeArk: F. Crittin,
- HEVs (4 persons):
  - D. Gabioud (Project Manager),
  - J-P. Gehrig (Face tracker and BLUM),
  - F. Dessimoz (Bio-Login and BLUM),
  - G. Maître (Quality control & Performance evaluation)

Outline

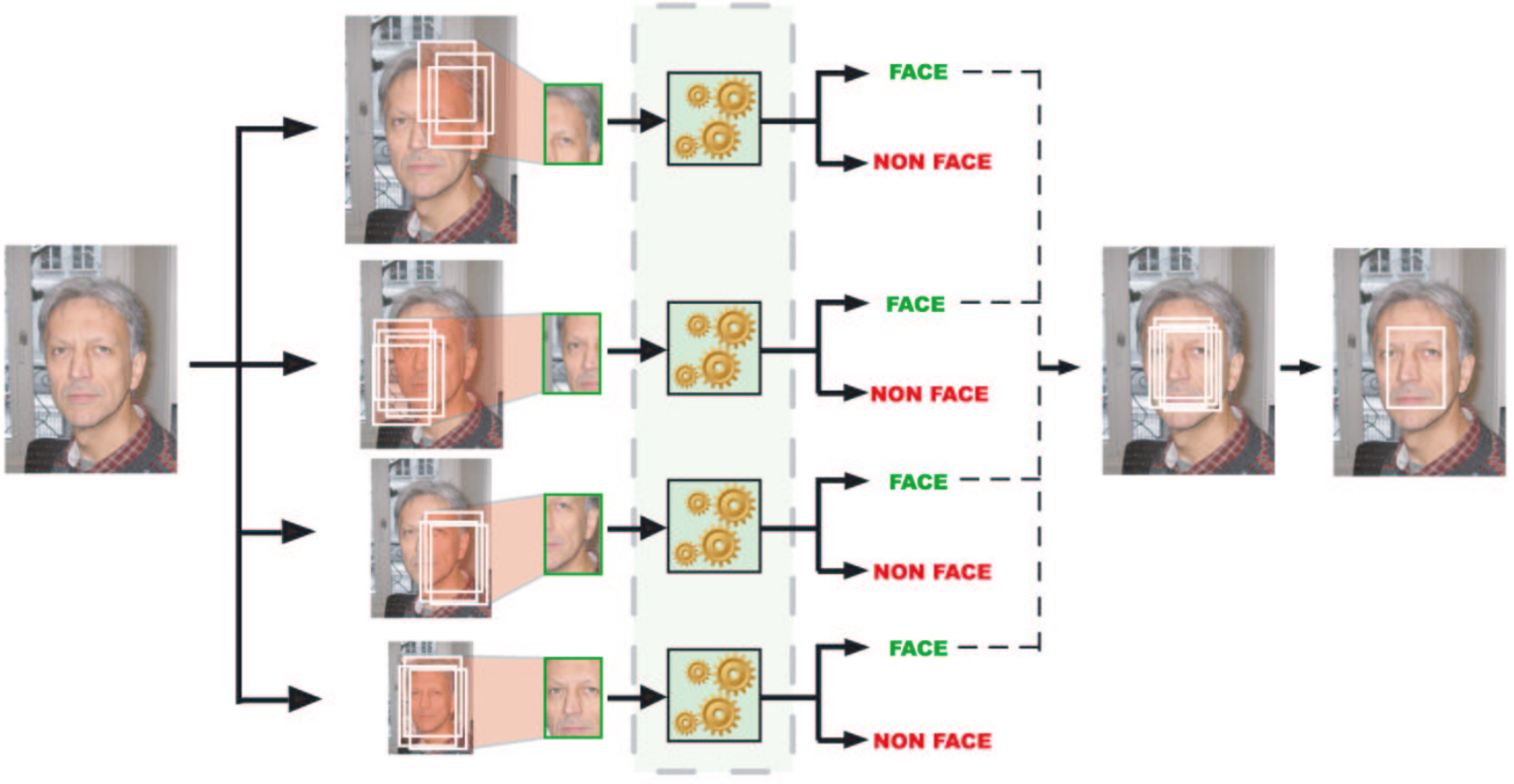
- Introduction and Background
- Face Detection + demo
- Bi-Modal Authentication + demo
- Discussions

Face Processing: Definitions

- Face localization: locates human faces in images at different positions, scales, orientations and lighting conditions with the assumption that the image contains one and only one face,
- Face detection: generalization of face localization to multiple faces,
- Face tracking: keep trace of similar detected/localized faces within a video sequence.

Face Detection: Overview

- Frontal face detection:



it is so fast that it provides the illusion of tracking

Face Detection: Demonstration

- Details:
  - Programming: Torch, C++ and QT
  - Platform: Win XP
  - Device: a Logitech camera (QuickCam Zoom, 4000 Pro, Notebook Pro, Orbit Sphere)

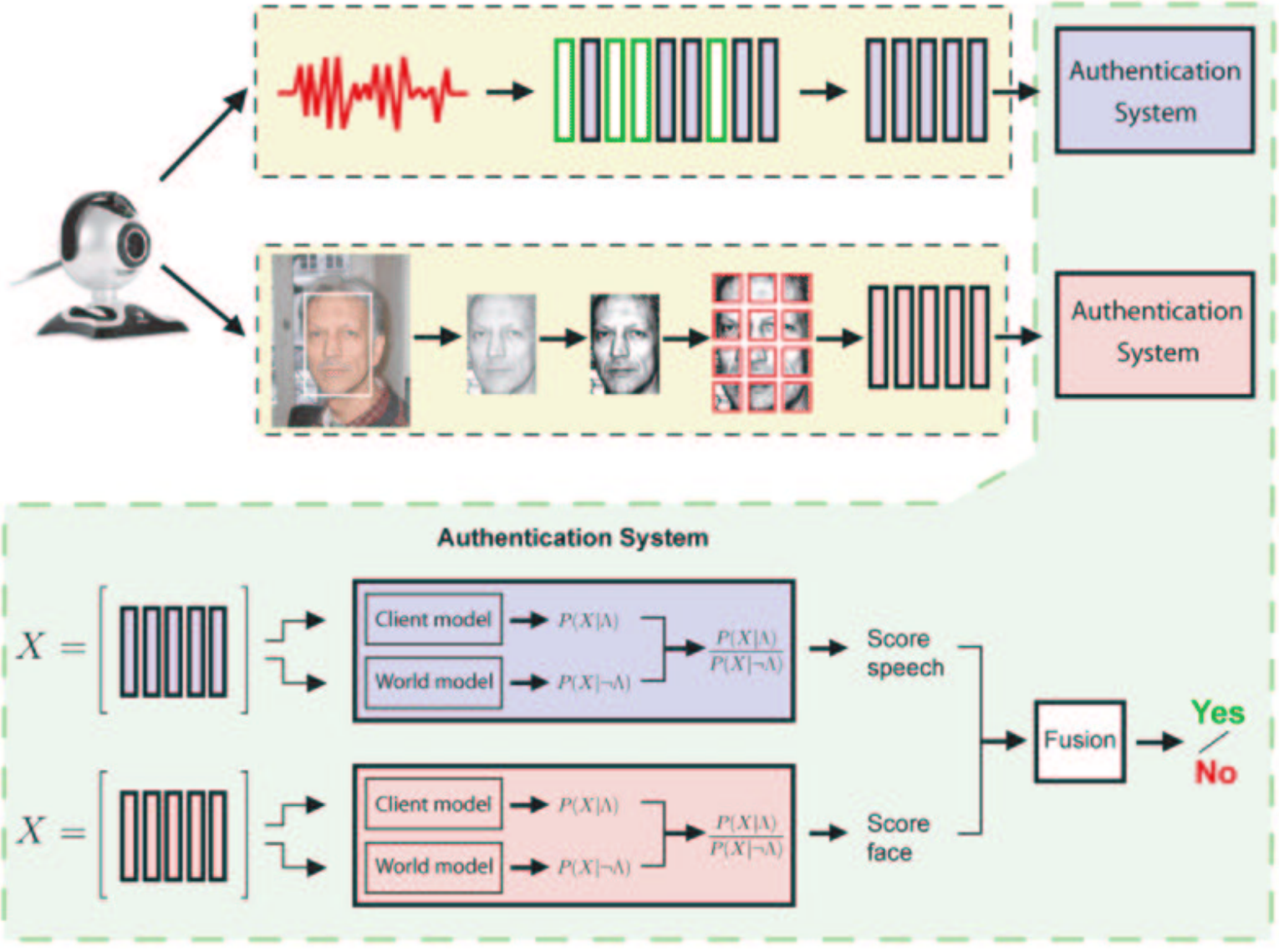
Outline

- Introduction and Background
- Face Detection + demo
- Bi-Modal Authentication + demo
- Discussions

Bi-Modal Authentication: Definitions

- Identification: identify a person based on the face/speech of a person. This face/speech has to be compared with all the registered persons (one-to-many matching),
- Authentication (or verification): is concerned with validating a claimed identity based on the face/speech , and either accepting or rejecting the identity claim (one-to-one matching),
- Recognition: general topic that includes both identification and authentication.

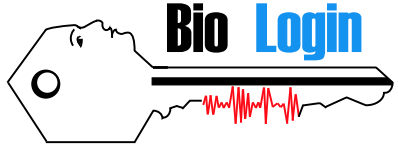
Bi-Modal Authentication: Overview



Bi-Modal Authentication: demonstration

- Details:
  - Programming: Torch, C++ and QT
  - Platform: Win XP
  - Device: a Logitech camera (QuickCam Zoom, 4000 Pro, Notebook Pro, Orbit Sphere)
- Two applications:
  - Bio-Login User Manager: Creates a new account (Enrollment)

  - Bio-Login: Login using your face and your voice (Test)



Discussions

- Questions ?
- More details ?

More details

- Bi-Modal authentication

Bi-Modal Authentication: Framework

- Face (64x80 pixels): Discrete Cosine Transform (DCT) based features and Gaussian Mixture Models (GMMs)
  - features: sequence ($X_1^F = \{\mathbf{x}_1 ... \mathbf{x}_F\}$) of DCTmod2 frames $\mathbf{x}_f \in \mathbb{R}^{18}$ (15 DCT coeff. -3 first coeff. $+3\ \Delta_x\ +3\ \Delta_y$)
  - models: diagonal GMMs (λ^{face}) with 512 gaussians ($18'944$ param.)
  - score: $\Lambda_C^{face}(X_1^F) = \log P(X_1^F|\lambda_C^{face}) - \log P(X_1^F|\neg\lambda_C^{face})$
- Speech (8 KHz): Linear Freq Cepstral Coefficient (LFCC) features and GMMs
  - features: sequence ($Y_1^S = \{\mathbf{y}_1 ... \mathbf{y}_S\}$) of LFCC frames $\mathbf{y}_s \in \mathbb{R}^{33}$ (16 basis + ∂ + log energy ∂)
  - models: diagonal GMMs (λ^{speech}) with 200 gaussians ($13'400$ param.)
  - score: $\Lambda_C^{speech}(Y_1^S) = \log P(Y_1^S|\lambda_C^{speech}) - \log P(X_1^S|\neg\lambda_C^{speech})$
- Fusion: $P(X, Y|C) = w \cdot \Lambda_C^{face}(X) + (1-w) \cdot \Lambda_C^{speech}(Y)$

Bi-Modal Authentication: Enrollment

- Let us note:
  - the parameter set for client C as λ_C,
  - the parameter set describing a generic client as $\neg\lambda_C$,
  - $X = \{\mathbf{x}_t\}_{t=1}^{T}$ as the set of feature vectors $\mathbf{x}_t$ supporting the claim.

1 train a world model $\neg\lambda_C$ from a large dataset by Maximum Likelihood,

2 adapt a client model λ_C from $\neg\lambda_C$ using client data by Maximum A Posteriori,

- given a claim for client C's identity and a set of feature vectors X supporting the claim, we find an opinion on the claim using:

$$\Lambda_C(X) = \log P(X|\lambda_C) - \log P(X|\neg\lambda_C) \quad (1)$$

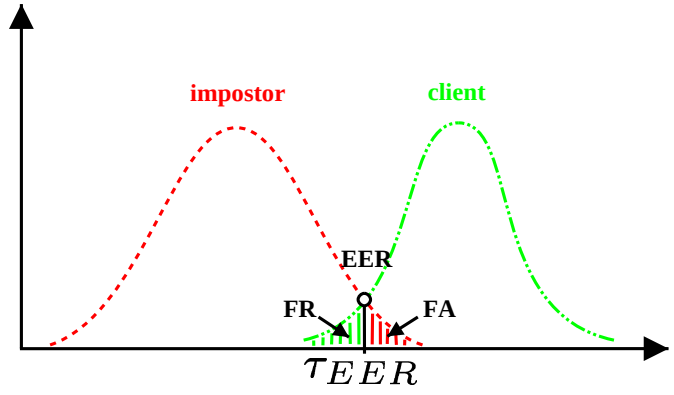
where $P(X|\lambda_C)$ is the likelihood of the claim coming from the true claimant and $P(X|\neg\lambda_C)$ is the likelihood of the claim coming from an impostor,

Fusion

- Goal: merge outputs of face and speech experts (2 or more) into a feature vector $[\Lambda^1(X), ..., \Lambda^n(X)]$ and try to classify it as a client or an impostor,
- Classifiers: Linear, MLP (Multi-Layer Perceptrons), SVM (Support Vector Machines),
- Fusion produces an opinion $\Lambda^*(X)$ that might be used for final decision.

Performance Evaluation

- the verification decision is then reached as follows: given a threshold τ, the claim is:
  - is accepted when $\Lambda(X) \geq \tau$,
  - is rejected when $\Lambda(X) < \tau$.
- Select a threshold to take the final decision:



  - False Rejection (FRR): when the system rejects a client,
  - False Acceptance Rate (FAR): when the system accepts an impostor,
  - Half Total Error Rate ($HTER = \frac{FRR+FAR}{2}$): unique measure,
  - the decision threshold τ chosen on a development data set.

Results

- results (in terms of HTER) on XM2VTS, BANCA and IDIAP databases:

	XM2VTS (LP1)	BANCA (Mc)	BANCA (P)	IDIAP
Face	1.67	5.77	18.96	7.61
Speech	1.14	4.32	12.29	3.15
Fusion	0.48	4.32	9.99	1.49

- XM2VTS can be found at `http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/`,
- BANCA can be found at `http://www.ee.surrey.ac.uk/banca/`.