

A Tutorial on Face Recognition

Dr Sébastien Marcel

marcel@idiap.ch

*IDIAP Research Institute
Martigny, Switzerland*

<http://www.idiap.ch>



Before We Start

MI2 not IM2



Click on the picture to play from the file

[Click here to play/download from the Internet]

Before We Start

MI2 not IM2



→ How to do this is not the purpose of this tutorial sorry !

Introduction

- Automatic face recognition is still a challenging problem with many applications,
- Two modes for face recognition:
 - identification: establish the identity of a given person out of a pool of N people (1-to- N matching),
 - verification (or authentication): confirm or deny the identity claimed by a person (1-to-1 matching),
- Distinct applications:
 - identification: video surveillance (public places, restricted areas), information retrieval (police databases, multimedia data management) or human computer interaction (video games, personal settings identification),
 - verification: access control, such as computer or mobile device log-in, building gate control, digital multimedia data access.

Outline

- × Introduction
- ▷ **Applications**
 - Face Detection (in short)
 - Face Recognition
 - Conclusion
 - Credits

Application Examples

- Biometrics
- Content-based Image/Video Indexing and Retrieval (CBIR)

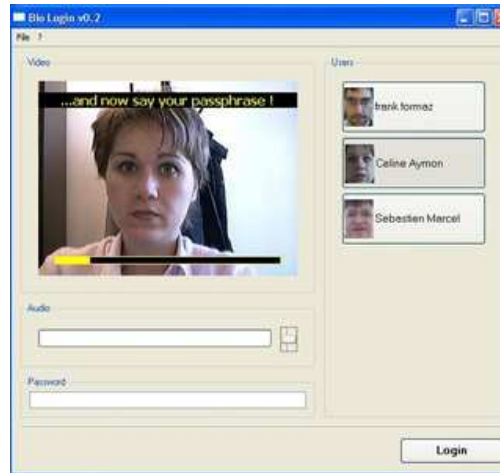
Application Examples

▷ Biometrics:

- secure transactions and secure access to online services
 - * micro payment services,
 - * phone card reloading,
 - * remote purchase,
 - * telephone banking, ...
- embedded applications:
 - * PIN code replacement,
 - * lock/unlock device,
 - * personal data protection (agenda, address book), ...
- Content-based Image/Video Indexing and Retrieval (CBIR)

Biometrics: Targeted Applications

- PC or laptop:

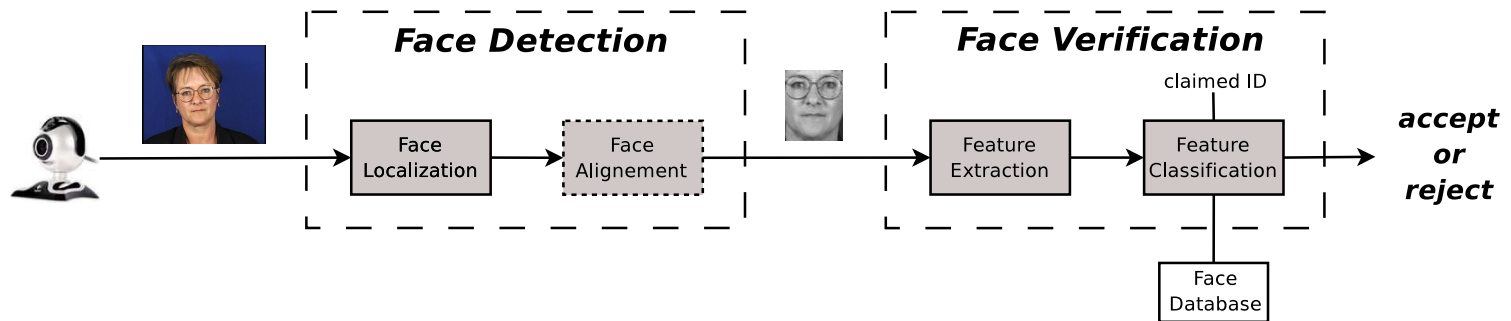


- Mobile/PDA: built-in microphone, video camera and also fingerprint scanner (Fujitsu DoCoMo F902i, Pantech PG-6200, Radio Co. WX310J, Omron, ...)



Biometrics: Example

- Automatic face verification system:



- Why verification ?:
 - verification is more simple to design,
 - verification scales well to identification,
 - a lot of databases as well as strict experiment protocols exist,
 - identification and verification often share the same feature extraction and classification algorithms,
 - face detection/localisation is also included.
- Demos:
 - **BioLogin**.
 - **BananaScreen** (www.bananasecurity.ch).

Application Examples

- × Biometrics
- ▷ Content-based Image/Video Indexing and Retrieval (CBIR):
 - automatic annotation of personal photos and home videos,
 - multimedia data management and organisation,
 - automatic indexing by image/video content, by (Exif) meta-data (time, location via GPS tags, ...)
 - tools to facilitate search in large (photo) collections (QBE, QBT, ...).

CBIR: Targeted Applications

- Photo Sharing and Automatic Annotation of Photo Albums:
 - PolarRose (<http://www.polarrose.com>): face detection and recognition,
 - Riya (www.riya.com): face detection, face recognition and text recognition.
- Visual shopping: Like (www.like.com)
- Photo Correction: SilverWire (www.silverwire.com)
- Others:
 - myHeritage (www.myheritage.com): face detection in family photos,
 - Nevenvision: face detection and recognition (acquired by Google in 2006),
 - Google (www.google.com) with Picasa (picasa.google.com).

CBIR: Examples

- Image indexing:
 - from meta-data: Locate these pictures ?



- from content: [Google Portrait](#).
- Video indexing: [I know Kung Fu](#)
 - *[Play from the file]*
 - *[Play/Download from the Internet]*

Outline

- × Introduction
- × Applications
- ▷ Face Detection (in short)
 - Face Recognition
 - Conclusion
 - Credits

Outline

- × Introduction
- × Biometrics
- ▷ **Face Detection (in short)**
 - Goal and Challenges
 - Feature-based vs Appearance-based Approaches
 - Frontal Face Detection
 - Multi-View Face Detection
 - Experiment Results
- Face Recognition
- Conclusion
- Credits

Goal and Challenges

- Goal:
 - determine if there are any face in an image (or video),
 - provide location, size and eventually orientation for all faces.
 - Challenges
 - high variability of the face in shape, colour and texture,
 - lighting conditions, pose, background, occlusions.
- Face detection is the first step to any face processing systems

Feature-based vs Appearance-based Approaches

- Feature-based approaches (without Machine Learning “a priori”)
- Appearance-based approaches (with Machine Learning inside !!)

Feature-based vs Appearance-based Approaches

- Feature-based approaches:
 - make explicit use of face knowledge:
 - * local features of the face (nose, mouth, eyes),
 - * structural relationship between these facial features.
 - are generally used for face localisation (one face),
 - require good quality images,
 - are robust to illumination conditions, occlusions and viewpoint,
 - but may also be computationally expensive.
- Appearance-based approaches

Feature-based vs Appearance-based Approaches

- Feature-based approaches:
 - make explicit use of face knowledge:
 - * local features of the face (nose, mouth, eyes),
 - * structural relationship between these facial features.
 - . . .

- Appearance-based approaches:
 - consider face detection as a two-class pattern recognition problem,
 - rely on statistical learning methods to build a face/non-face classifier from training samples,
 - are used for face detection in images with low resolution,
 - have received considerable attention,
 - have proven to be more successful and robust than feature-based approaches.

Feature-based vs Appearance-based Approaches

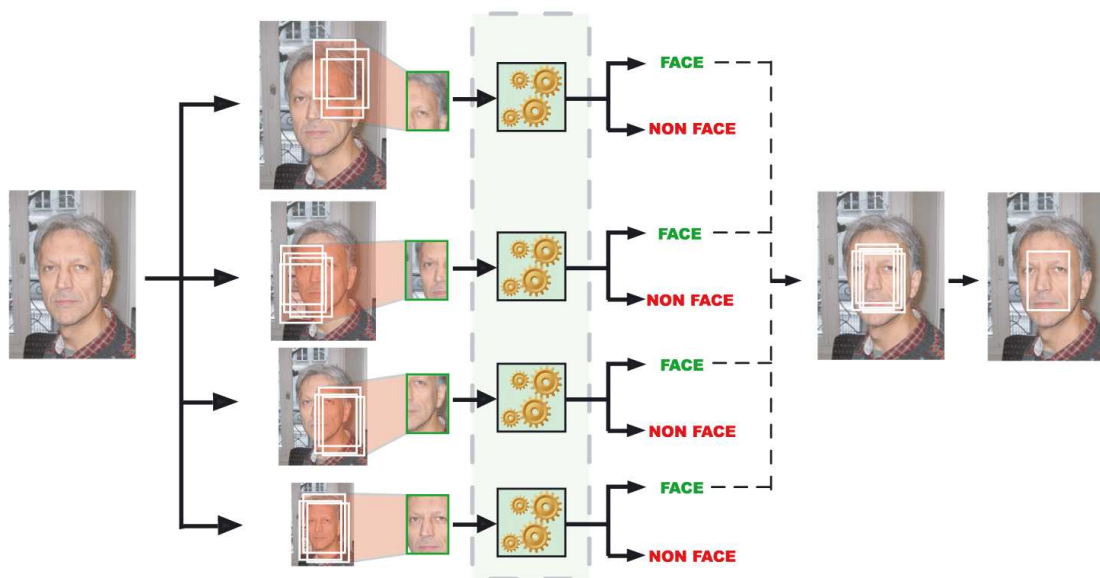
- Feature-based approaches:
 - make explicit use of face knowledge:
 - * local features of the face (nose, mouth, eyes),
 - * structural relationship between these facial features.
 - . . .

the detection of local features is often done using appearance-based approaches !

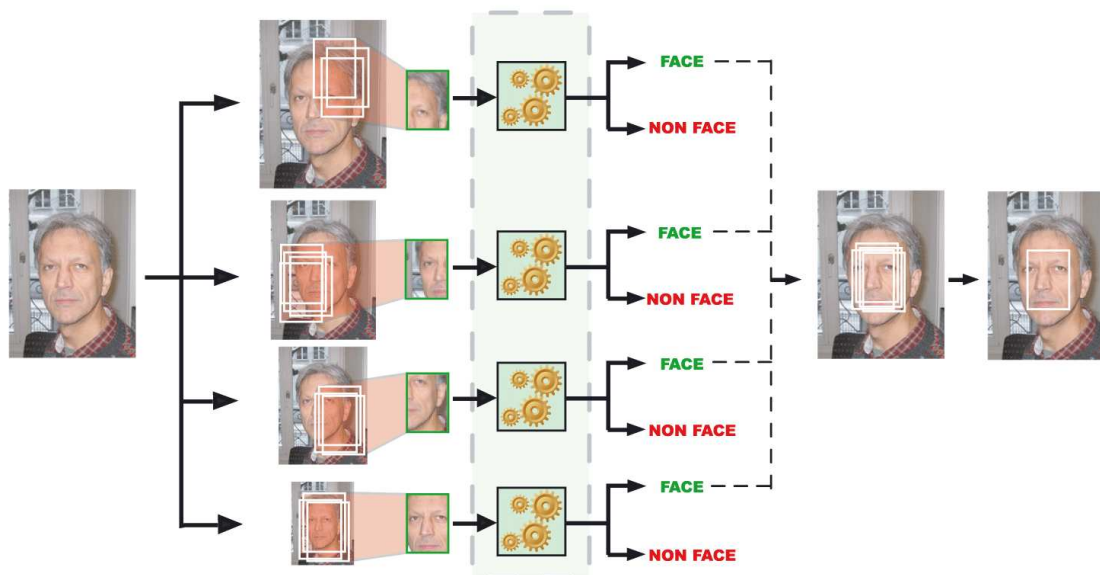
- Appearance-based approaches:
 - consider face detection as a two-class pattern recognition problem,
 - rely on statistical learning methods to build a face/non-face classifier from training samples,
 - are used for face detection in images with low resolution,
 - have received considerable attention,
 - have proven to be more successful and robust than feature-based approaches.

Main concepts of Appearance-based Approaches

1. *scanning window*: this is the root idea of appearance-based methods: a sliding window scans the input image at different locations and scales.
2. *face/non-face classifier*: Each sub-window is then given to a classifier whose goal is to classify the sub-window as either a *face* or a *non-face*.
3. *merging overlapped detections*: Multiple detections at different locations and scales may occur around a face in the image.



Main concepts of Appearance-based Approaches



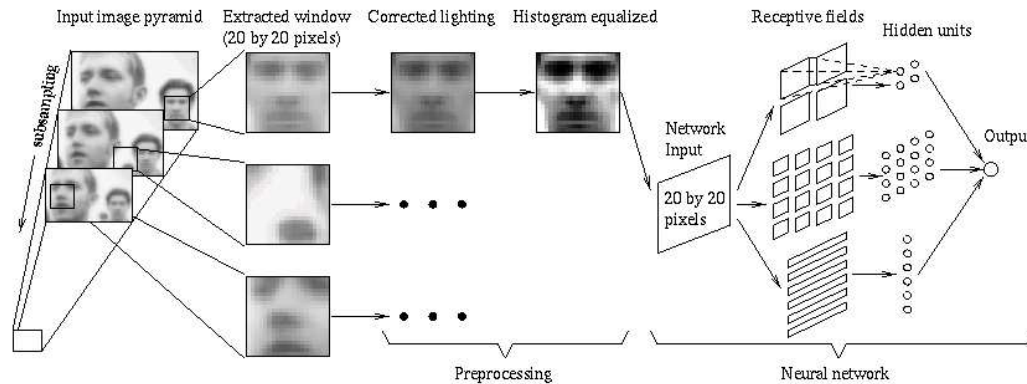
- Appearance-based methods mainly differ in the choice of the classifier: Support Vector Machines, Neural Networks, Bayesian classifiers or Hidden Markov Models.
- We will report here the most significant appearance-based approaches both for frontal face detection and multi-view face detection.

Outline

- × Introduction
- × Biometrics
- ▷ Face Detection (in short)
 - × Goal and Challenges
 - × Feature-based vs Appearance-based Approaches
 - ▷ Frontal Face Detection
 - ▷ State-of-the-art
 - * Boosting-based Methods
 - Multi-View Face Detection
 - Experiment Results
- Face Recognition
- Conclusion
- Credits

State-of-the-art

- Rowley et al. [1998]: ensemble of Neural Networks which works on pixel intensities,

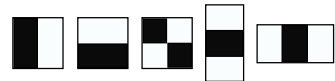


- Discussion:
 - Advantages: provide accurate detection with few false alarms.
 - Drawbacks: need several seconds at best to process an image (sub-windows need to be photo-metrically normalised before classification).

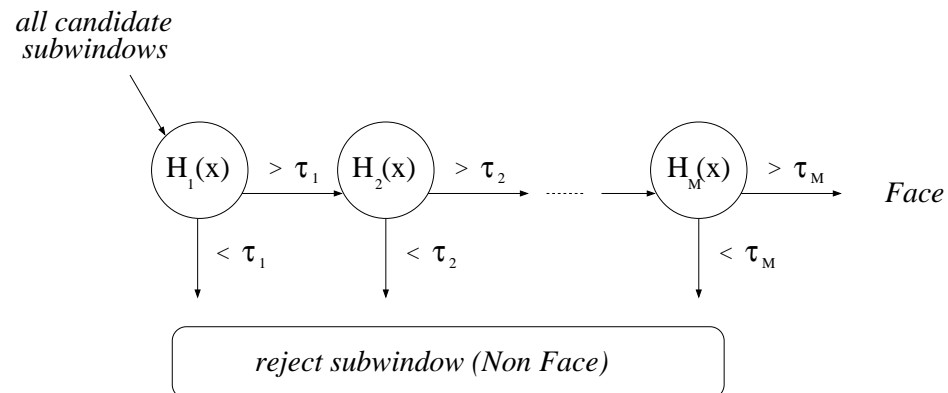
This limitation is restrictive for real-life applications that need real-time face detection (> 15 frames per second)

State-of-the-art

- Viola and Jones [2001]: the first real-time frontal face detection system based on boosting learning (Adaboost)
 - simple image features (Haar-Like) can be computed at any position and scale in constant time using the integral image representation,



- weak-classifiers are assembled into strong classifiers using boosting,
- a cascade of strong classifiers with increasing complexity is built.

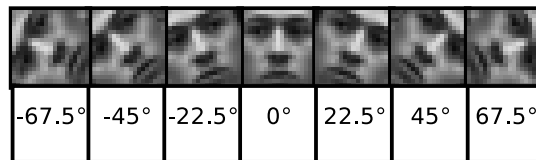


Outline

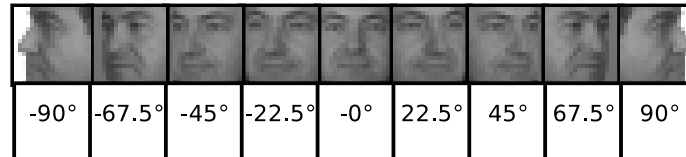
- × Introduction
- × Biometrics
- ▷ **Face Detection (in short)**
 - × Goal and Challenges
 - × Feature-based vs Appearance-based Approaches
 - × Frontal Face Detection
 - ▷ **Multi-View Face Detection**
 - Experiment Results
- Face Recognition
- Conclusion
- Credits

Multi-View Face Detection

- Multi-view = multiple views (or pose) of the face
- In-plane rotation (roll) [In-plane views]:



- Frontal to profile out-of-plane rotation (pan) [Out-plane views]:

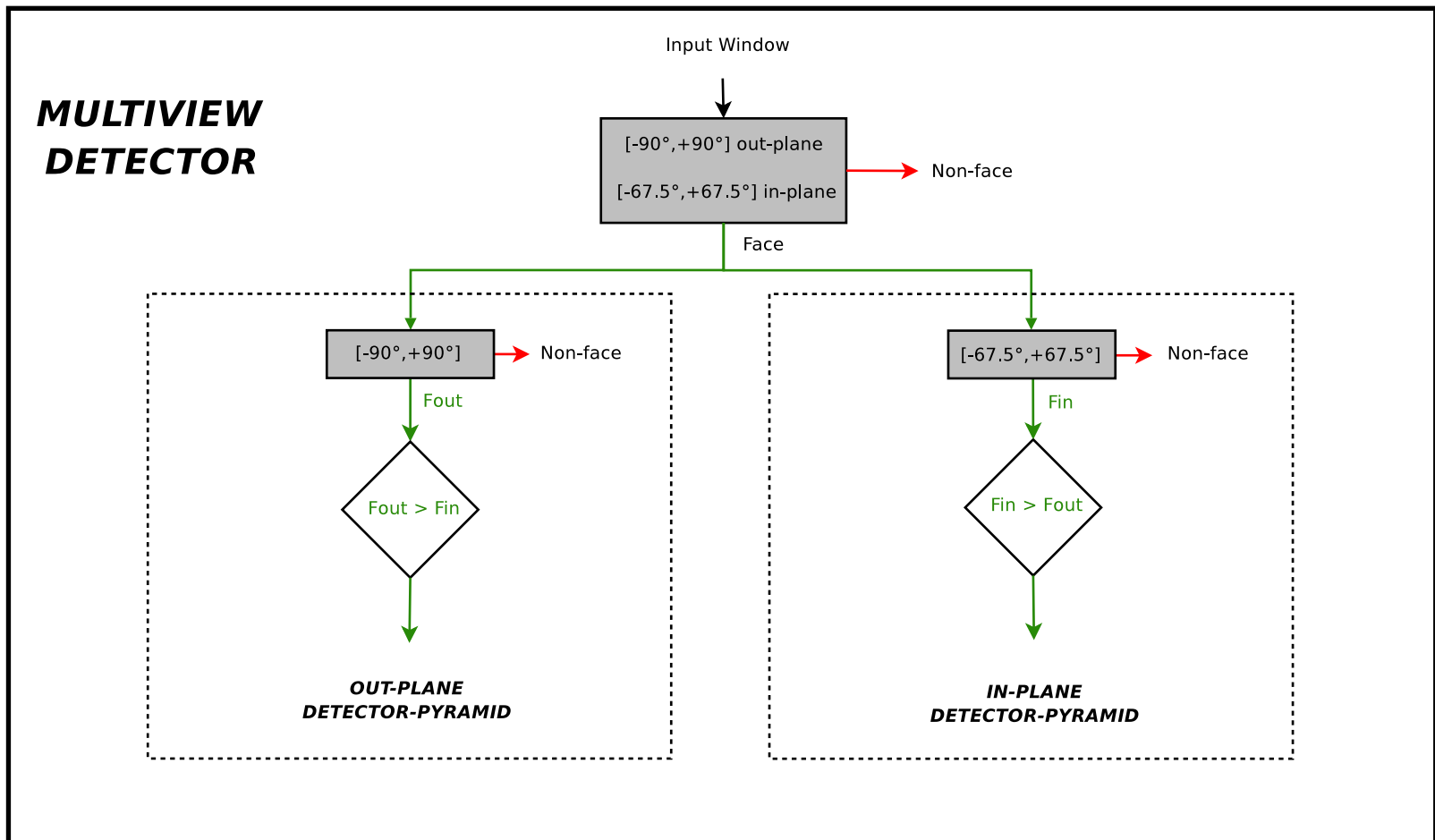


- Up-down nodding rotation (tilt),
- Combinations are also possible.

The different viewpoints largely increase the variety of face appearance and make the detection of multi-view faces much more difficult than the detection of frontal faces.

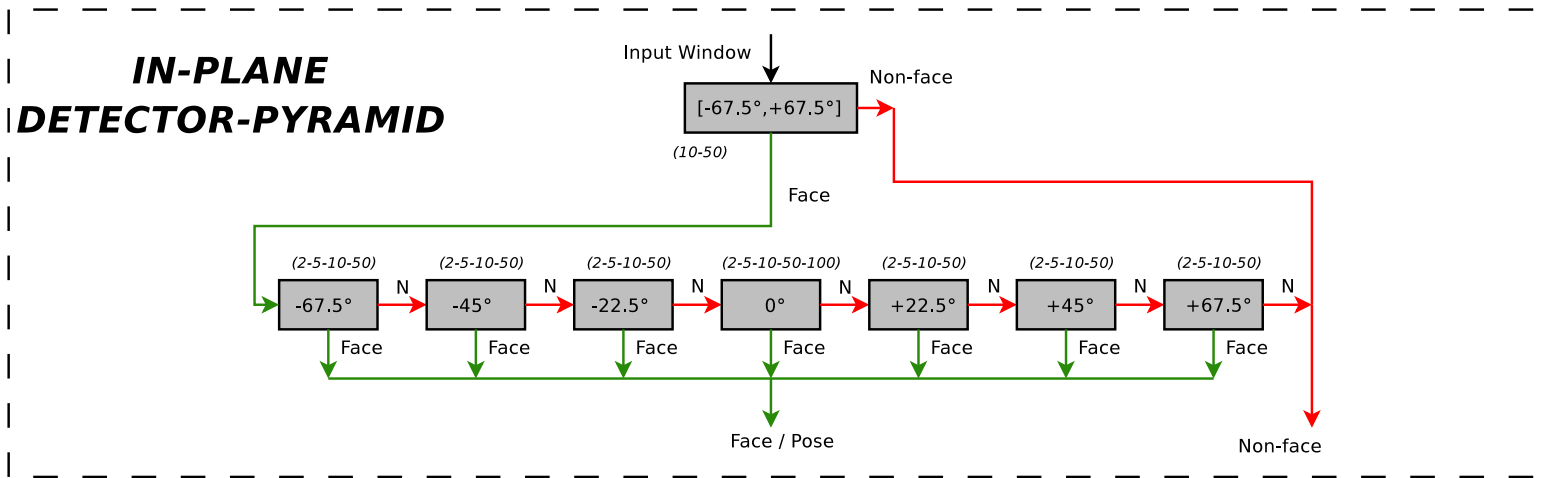
Pyramid-Cascade Architecture

- A multi-view face detector based on a pyramid-cascade architecture is composed of a top-level cascade, an in-plane pyramid and an out-plane pyramid:

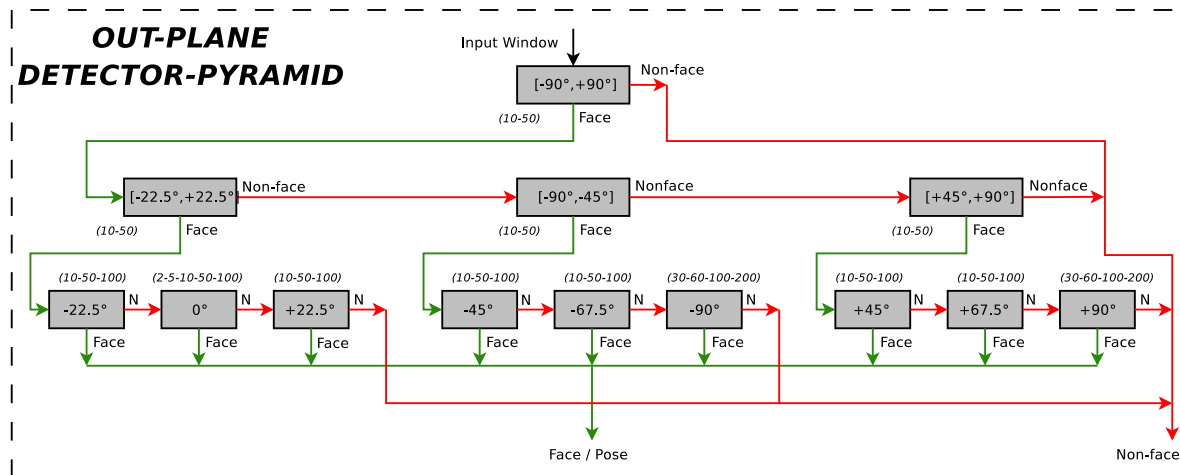


Pyramid-Cascade Architecture

- In-plane pyramid-cascade:



- Out-plane pyramid:



Pyramid-Cascade Architecture

- Advantages:
 - modular (easy to add novel views),
 - fast.
- Drawbacks:
 - overall training can be long,
 - architecture design is painful (# of view partitions, # of stages, # of classifiers).

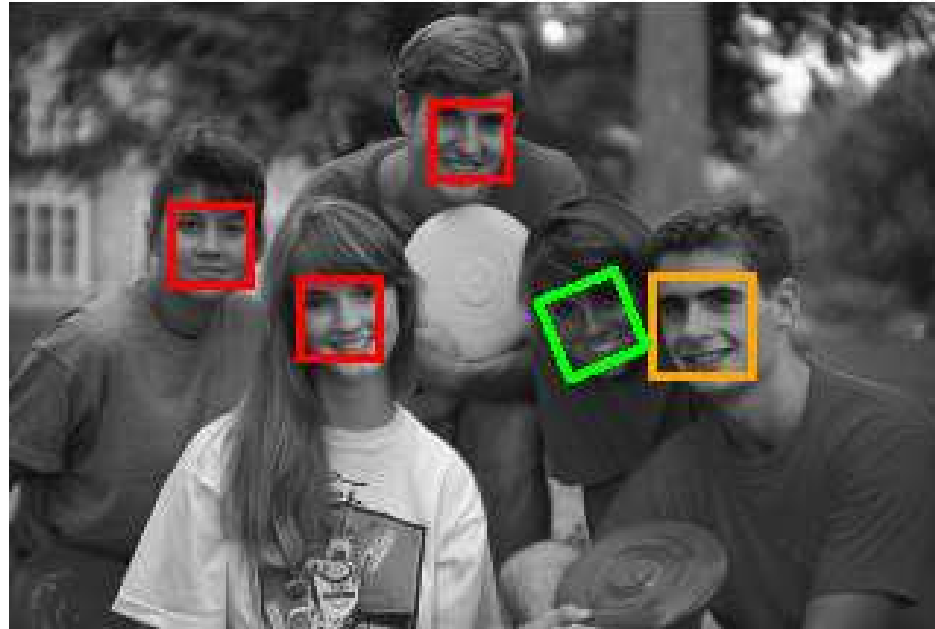
There is a need for machine learning algorithms to build automatically the structure and to train the modules.

Outline

- × Introduction
- × Biometrics
- ▷ **Face Detection (in short)**
 - × Goal and Challenges
 - × Feature-based vs Appearance-based Approaches
 - × Frontal Face Detection
 - × Multi-View Face Detection
 - ▷ **Experiment Results**
- Face Recognition
- Conclusion
- Credits

Experiment Results: Multi-View

- Inplane:

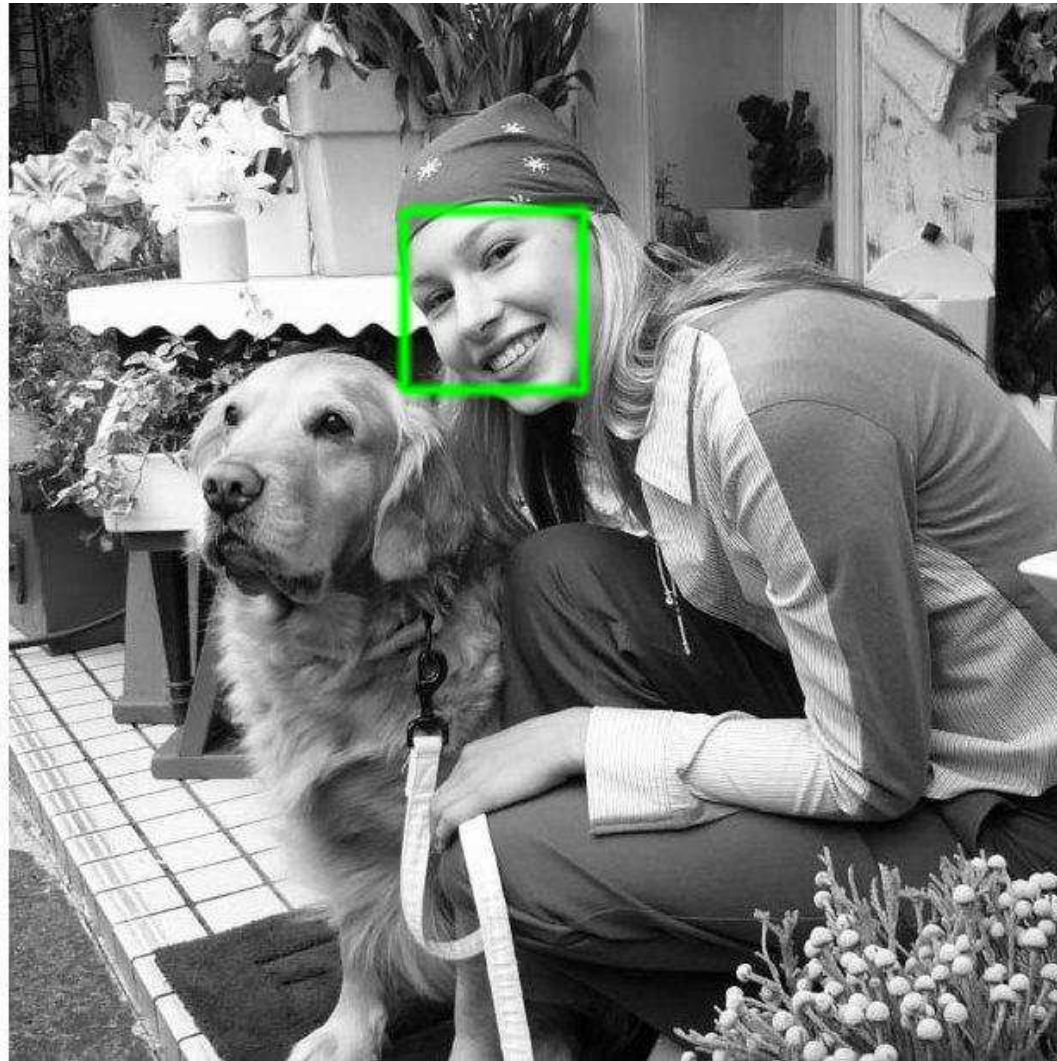


- Out-of-plane:



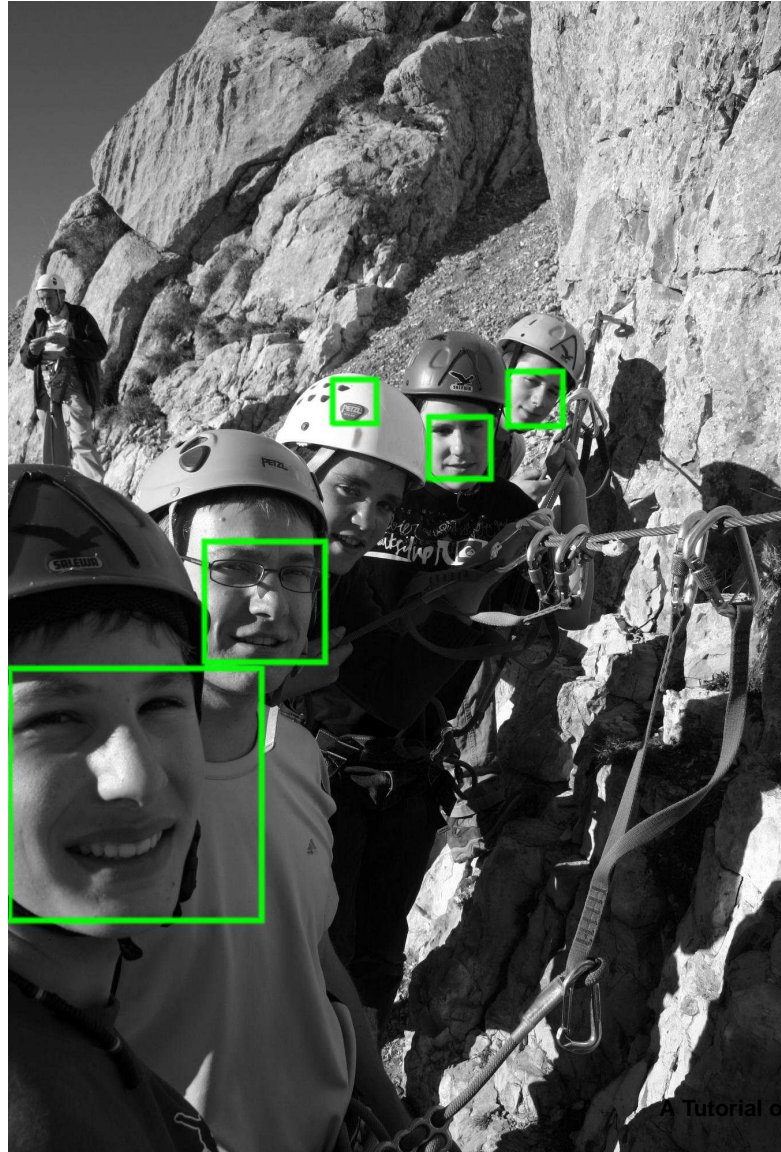
Experiment Results: Multi-View

- Tolerance to combinations of inplane and outplane rotations:



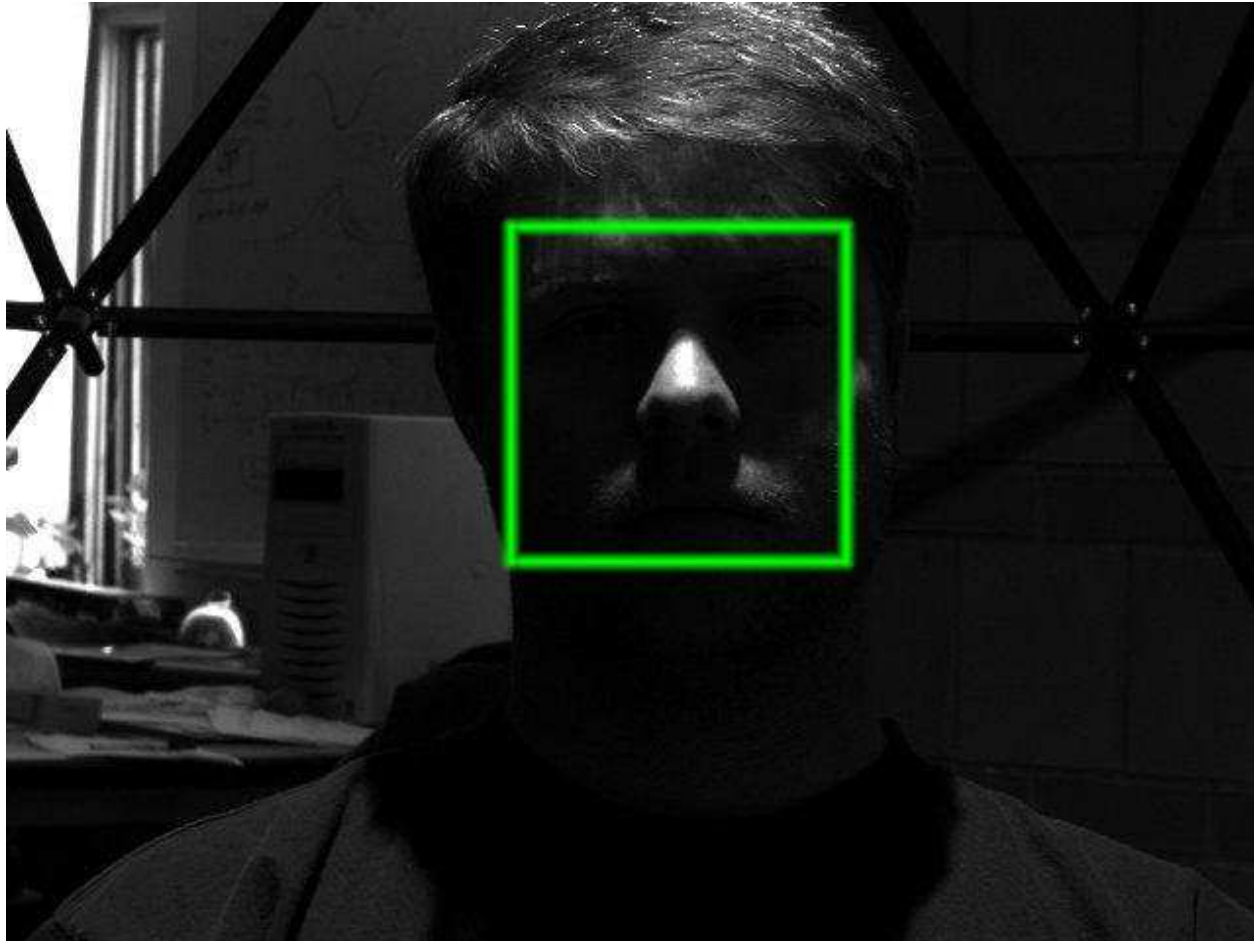
Experiment Results: Multi-View

- High-res images (1300×2000) and difficult illumination conditions:



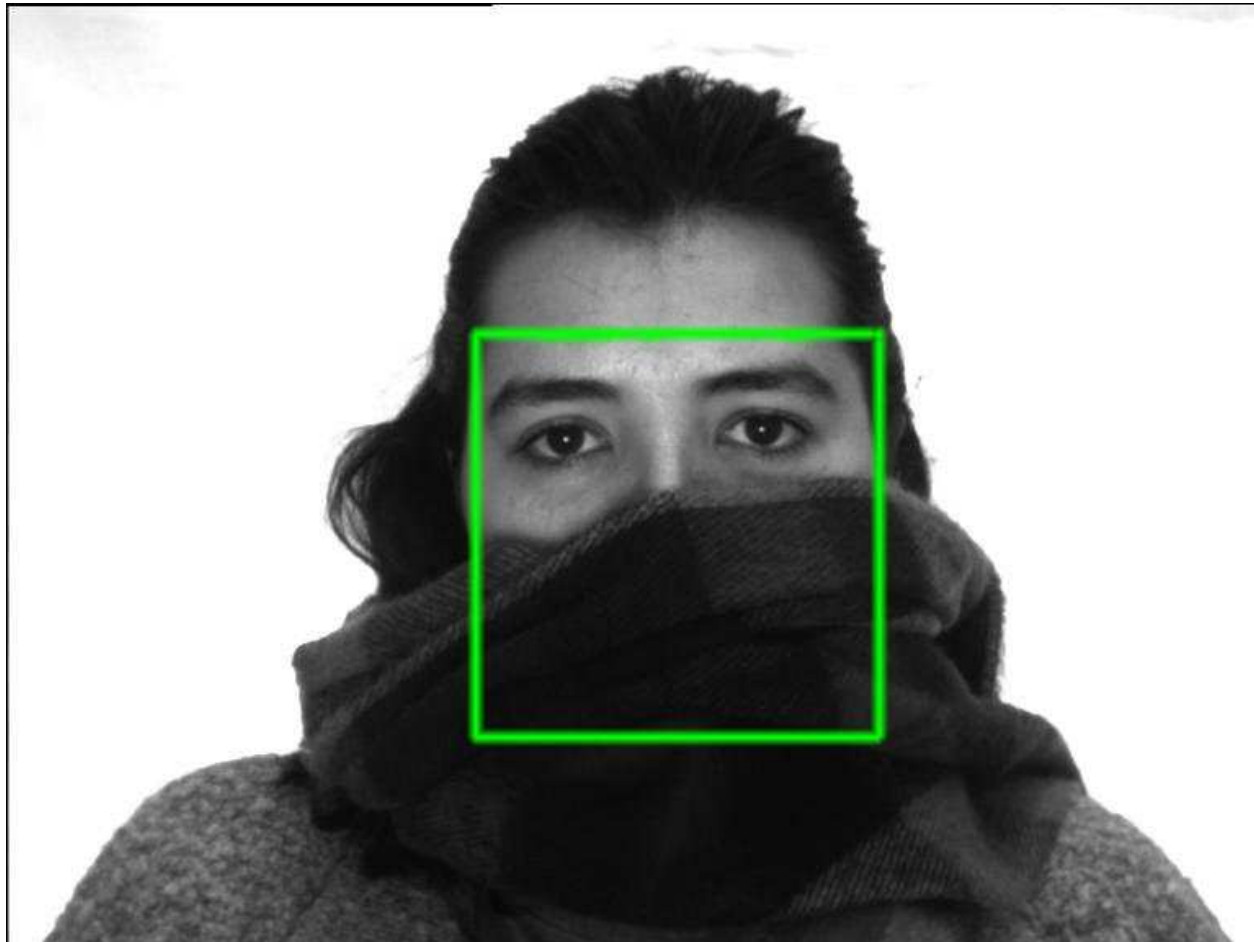
Experiment Results: Multi-View

- Robust to illumination:



Experiment Results: Multi-View

- Robust to occlusion:



Outline

- × Introduction
- × Applications
- × Face Detection (in short)
- ▷ Face Recognition
 - Conclusion
 - Credits

Outline

- × Introduction
- × Biometrics
- × Face Detection (in short)
- ▷ **Face Recognition**
 - ▷ **Introduction**
 - Challenges
 - Feature Extraction: Holistic vs Local
 - Classification
 - Database and Protocols
 - Experiment Results
 - Discussion
- Conclusion
- Credits

Introduction

- In spite of the expanding research in the field of face recognition, a lot of problems are still unsolved,
- Today, several systems that achieve high recognition rates have been developed, however:
 - such systems work in controlled environments,
 - for most of them, face images must be frontal or profile,
 - background must be uniform,
 - lighting must be constant.
- Furthermore, lot of published systems are evaluated using manually located faces,
- and the ones which have been evaluated using a fully automatic system showed a big degradation in performance.

Challenges

- In most real life applications, the environment is not known *a-priori* and the system should be fully automatic. A Face Recognition system has to deal with:
 - Lighting Variation,
 - Head Pose changes,
 - Non-Perfect Detection,
 - Occlusion,
 - Aging.
- The main problem is the large variability between face images:
 - *extra-personal* variabilities: variations in appearance between different identities,
 - *intra-personal* variabilities: variations in appearance of the same identity, due to different expression, lighting, background, head pose, hair cut, etc.

Outline

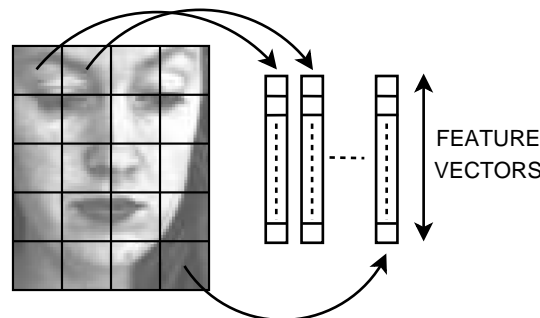
- × Introduction
- × Biometrics
- × Face Detection (in short)
- ▷ **Face Recognition**
 - × Introduction
 - × Challenges
 - ▷ **Feature Extraction: Holistic vs Local**
 - Classification
 - Database and Protocols
 - Experiment Results
 - Discussion
- Conclusion
- Credits

Feature Extraction: Holistic vs Local

- The goal of feature extraction is to find a specific representation of the data that can highlight relevant information.
- An image is represented by:
 - a high dimensional vector containing pixel values (holistic representation),



- a set of vectors where each vector contains gray levels of a sub-image (local representation).



Feature Extraction: Holistic vs Local

- Vectors are projected into a new space (the feature space) where the least relevant features can be removed to reduce the dimensionality according to a criterion (such as lowest amount of variance):
 - Holistic representations (representations found using the statistics of image data)
 - Local representations (researchers have argued that local filters are more robust than global representation)

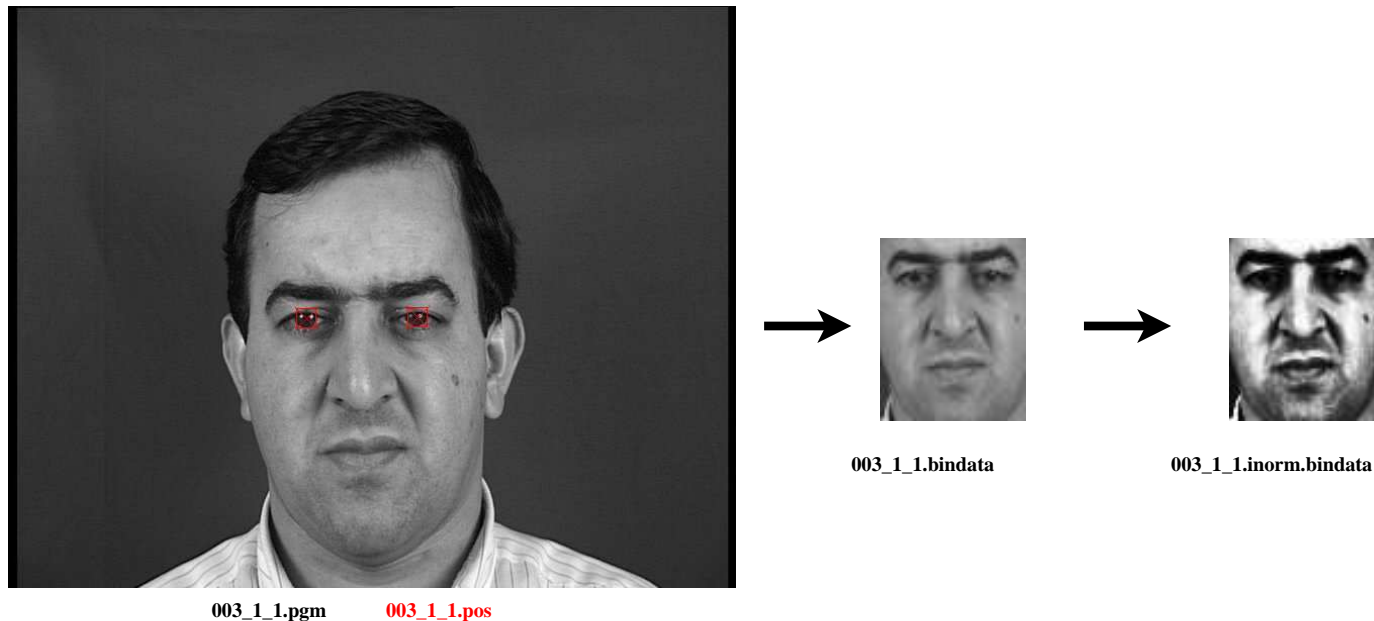
Feature Extraction: Holistic vs Local

- Vectors are projected into a new space (the feature space):
 - Holistic representations:
 - * Turk and Pentland [1991]: *Principal Component Analysis* (PCA)
 - * Zhao and al. [1999], Li and al. [2000]: *Linear Discriminant Analysis* (LDA, also known as Fisher Discriminant Analysis)

For face recognition, LDA should outperform PCA because it inherently deals with class discrimination. However, Martinez and Kak [2001] have shown that PCA might outperform LDA when the number of samples per class is small.
 - Local representations

Holistic Representation: Pre-Processing

- Geometric normalisation: the face is geometrically normalised to a 64 width \times 80 height image (5120 dimensions)
 1. translation (eyes centre location),
 2. rotation (compensates for in-plane rotations),
 3. scale.
- Photometric normalisation: Retinex, histogram equalisation, ...











Holistic Representation: PCA

- Let:
 - \mathbf{x} be a face image $\mathbf{x} \in \mathbb{R}^n$ with $n=5120$,
 - $\mathbf{X} = \{\mathbf{x}_1 \dots \mathbf{x}_P\}$ be the set of faces with P the number of images,
 - $\mu = \frac{1}{P} \sum_{k=1}^P \mathbf{x}_k$ the mean face image,
 - $\Sigma = \frac{1}{P} \sum_{k=1}^P (\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^T$ the covariance matrix of faces.
- Computes the m eigenvectors $\mathbf{e}_1 \dots \mathbf{e}_m$ corresponding to the m largest non-zero eigenvalues of the covariance matrix Σ of data \mathbf{X} by solving $(\Sigma - \alpha_i \mathbf{I})\mathbf{e}_i = 0, i = 1..m$:
- The coordinate system of eigenvectors forms the **Eigenface** space:



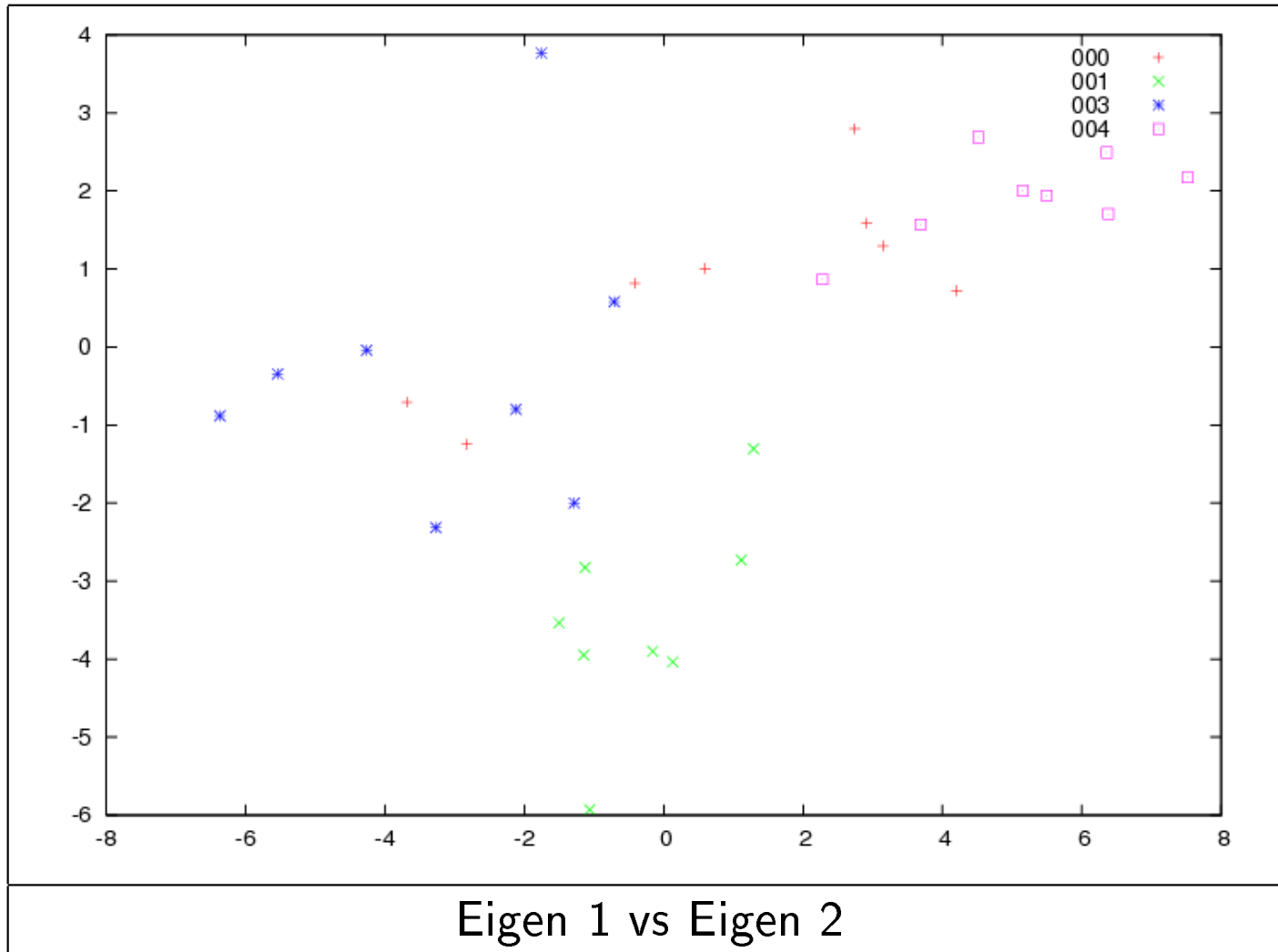
Holistic Representation: PCA

- Projects an image \mathbf{x} into the **Eigenface** space: $\mathbf{u} = \mathbf{W}\mathbf{x}$ where $\mathbf{u} \in \mathbb{R}^m$
- This achieves:
 - information compression,
 - de-correlation,
 - and dimensionality reduction to facilitate decision making.
- Illustration of information compression:

Rec								
σ	96 %	90 %	80 %	70 %	60 %	50 %	30 %	10 %
m	751	243	94	45	25	14	5	2
RMSE	0.002	0.004	0.008	0.012	0.019	0.023	0.028	0.038

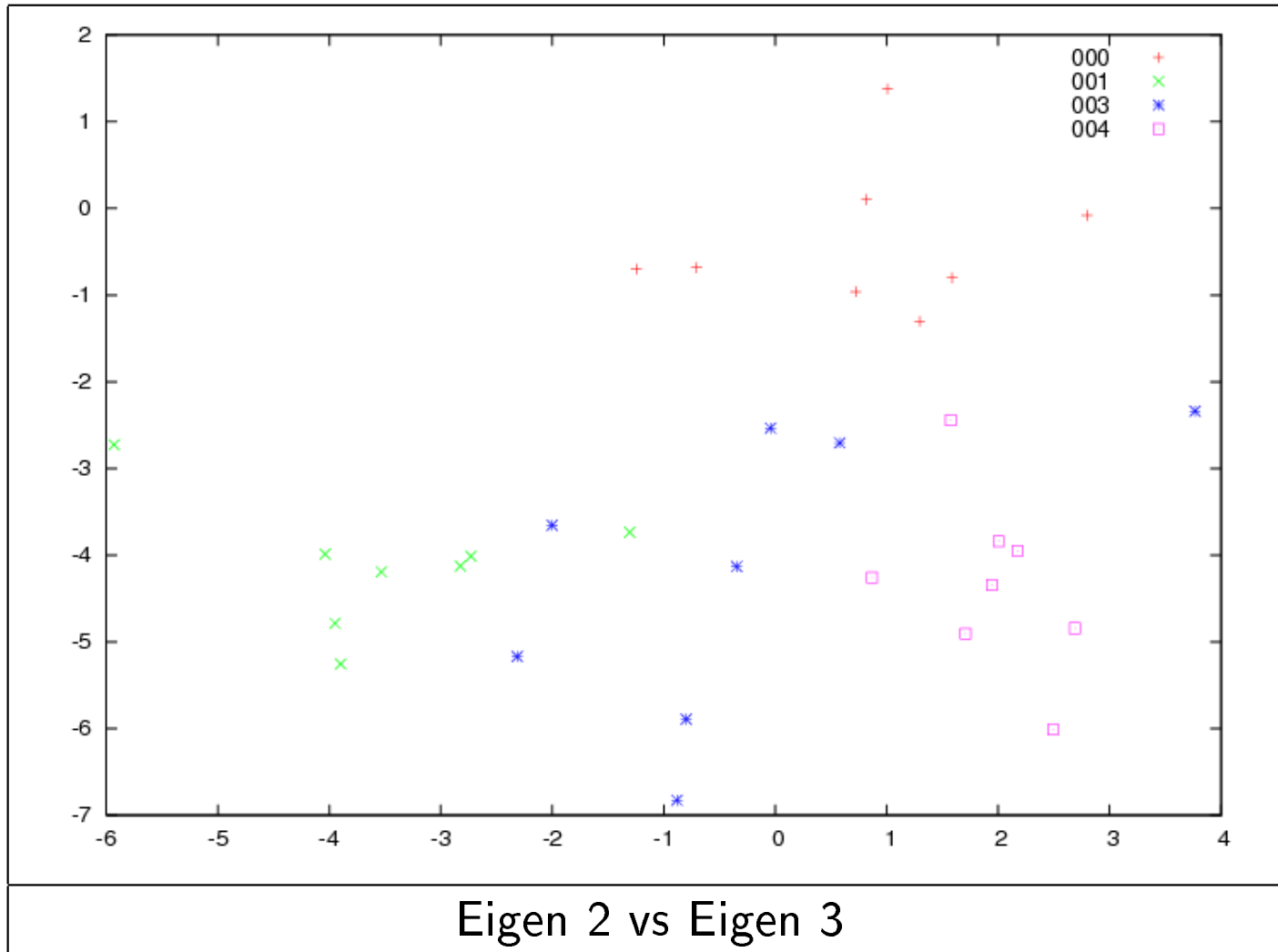
Holistic Representation: PCA

- Projections of different faces in 2D:



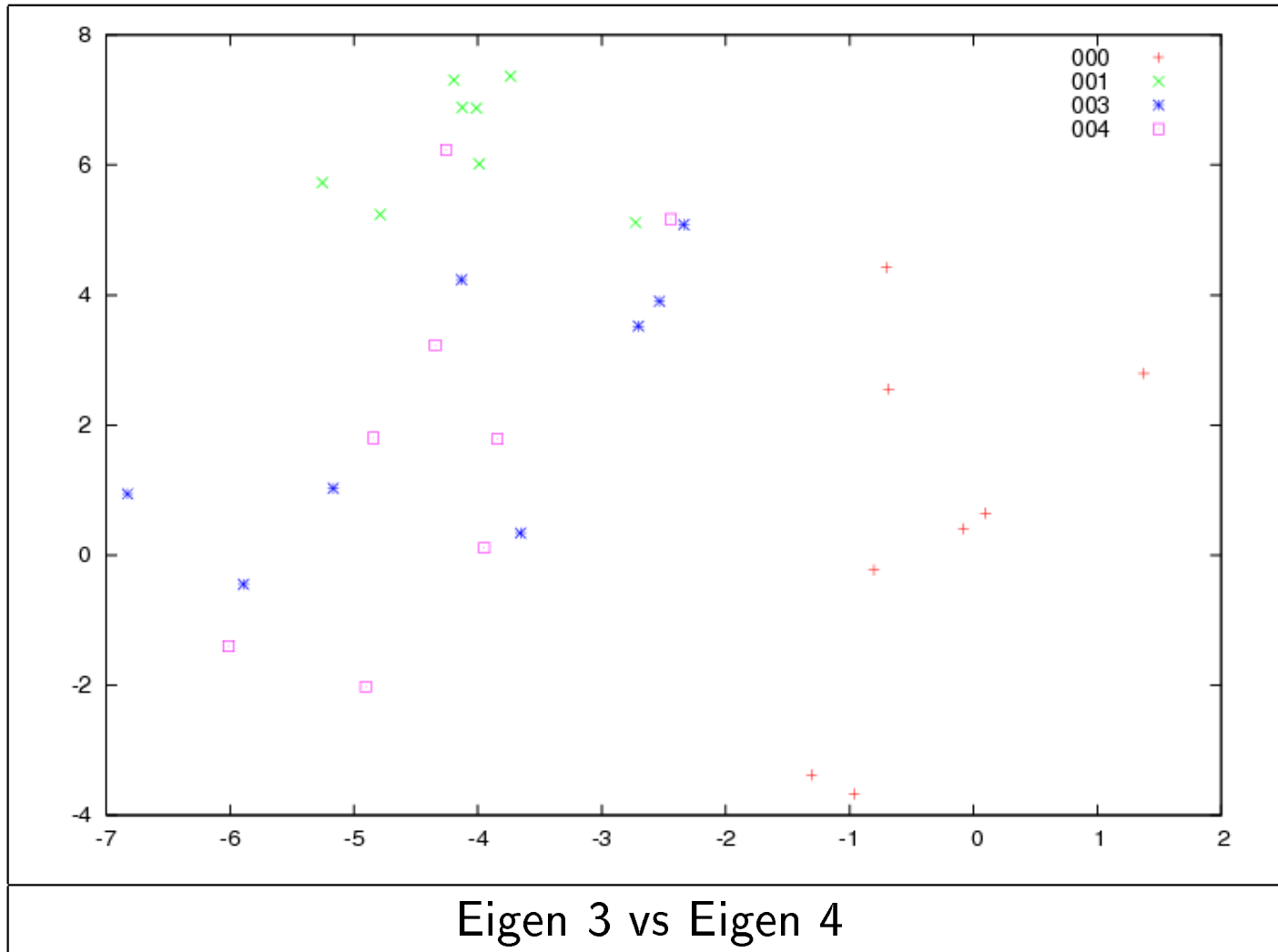
Holistic Representation: PCA

- Projections of different faces in 2D:



Holistic Representation: PCA

- Projections of different faces in 2D:



Holistic Representation: LDA using Fisher

- In Fisher LDA, Fisher criterion aims at maximising the ratio of between-class scatter \mathbf{S}_b to within-class scatter \mathbf{S}_w ,
- Let:
 - $\mathbf{X} = \{\mathbf{x}_1 \dots \mathbf{x}_P\}$ be the set of faces with P the number of images,
 - $\mathcal{C}_i, i = 1 \dots c$ be the set of classes where c is the number of classes,
 - l_i the number of images belonging to class i ,
 - $\mu_i = \frac{1}{l_i} \sum_{k \in \mathcal{C}_i} \mathbf{x}_k$ be the mean of each class,
 - $\mathbf{S}_w = \frac{1}{P} \sum_{i=1}^c \sum_{\mathbf{x}_k \in \mathcal{C}_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^T$ be the within-class scatter matrix,
 - $\mathbf{S}_b = \frac{1}{c} \sum_{i=1}^c (\mu_i - \mu)(\mu_i - \mu)^T$ be the between-class scatter matrix,
 - where μ is the grand mean, i.e the mean of the means μ_i .

Holistic Representation: LDA using Fisher

- Fisher's criterion can then be defined as maximising

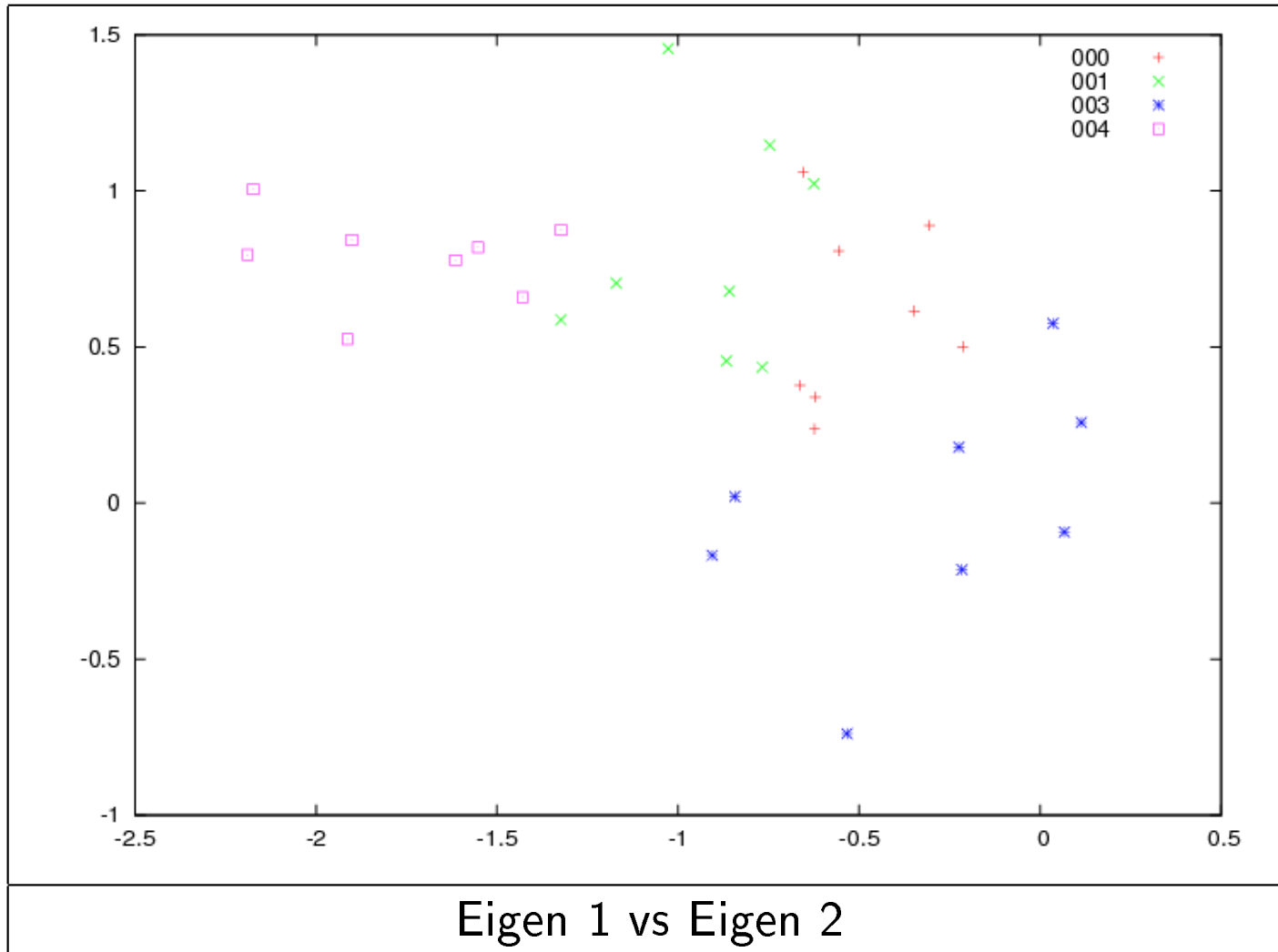
$$J(\mathbf{w}) = \frac{\mathbf{w}^t \mathbf{S}_b \mathbf{w}}{\mathbf{w}^t \mathbf{S}_w \mathbf{w}} . \quad (1)$$

- a solution can be found by computing the eigenvectors of:

$$\mathbf{w} = \mathbf{S}_w^{-1} \mathbf{S}_b . \quad (2)$$

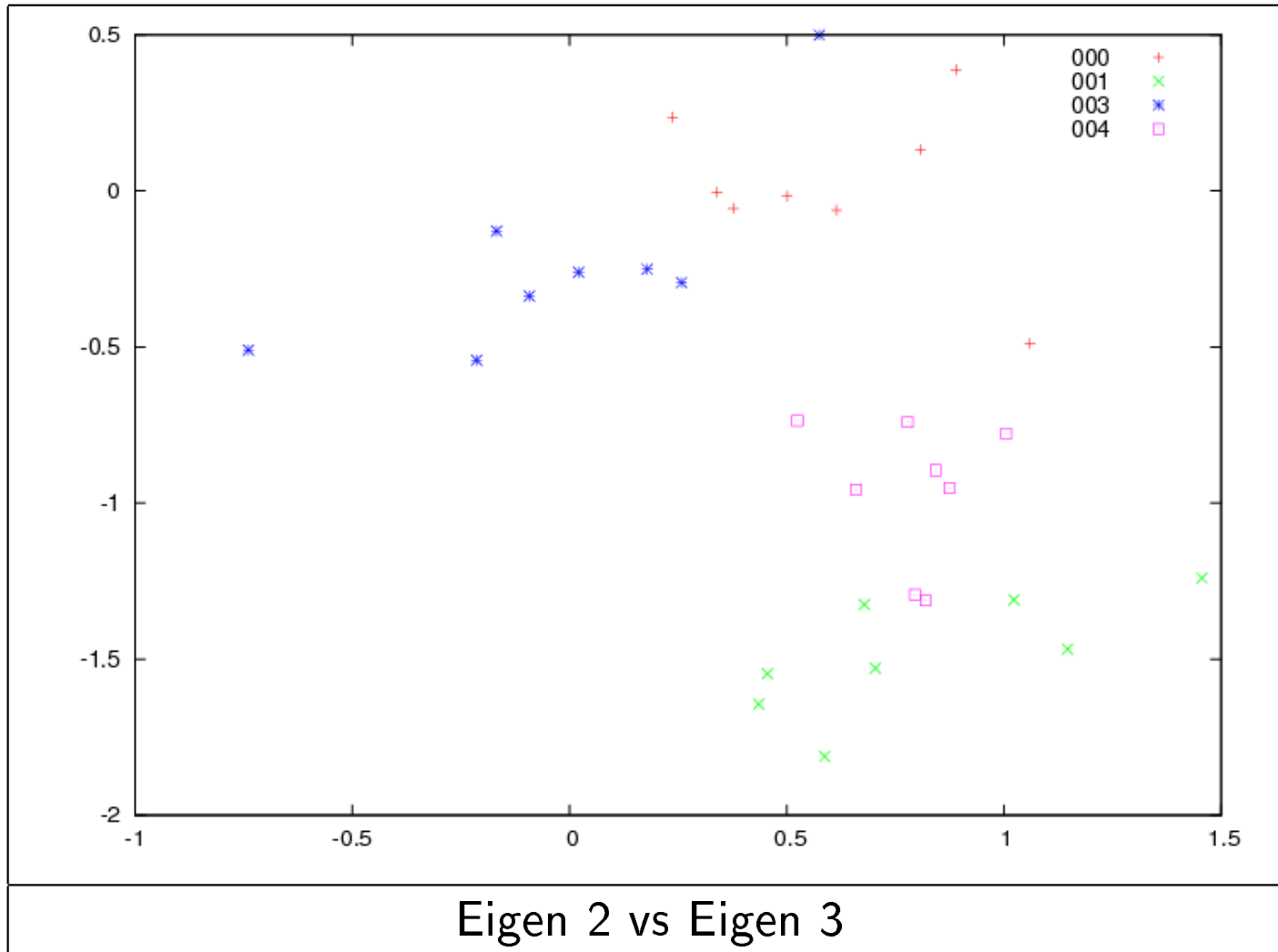
Holistic Representation: LDA

- Projections of different faces in 2D:



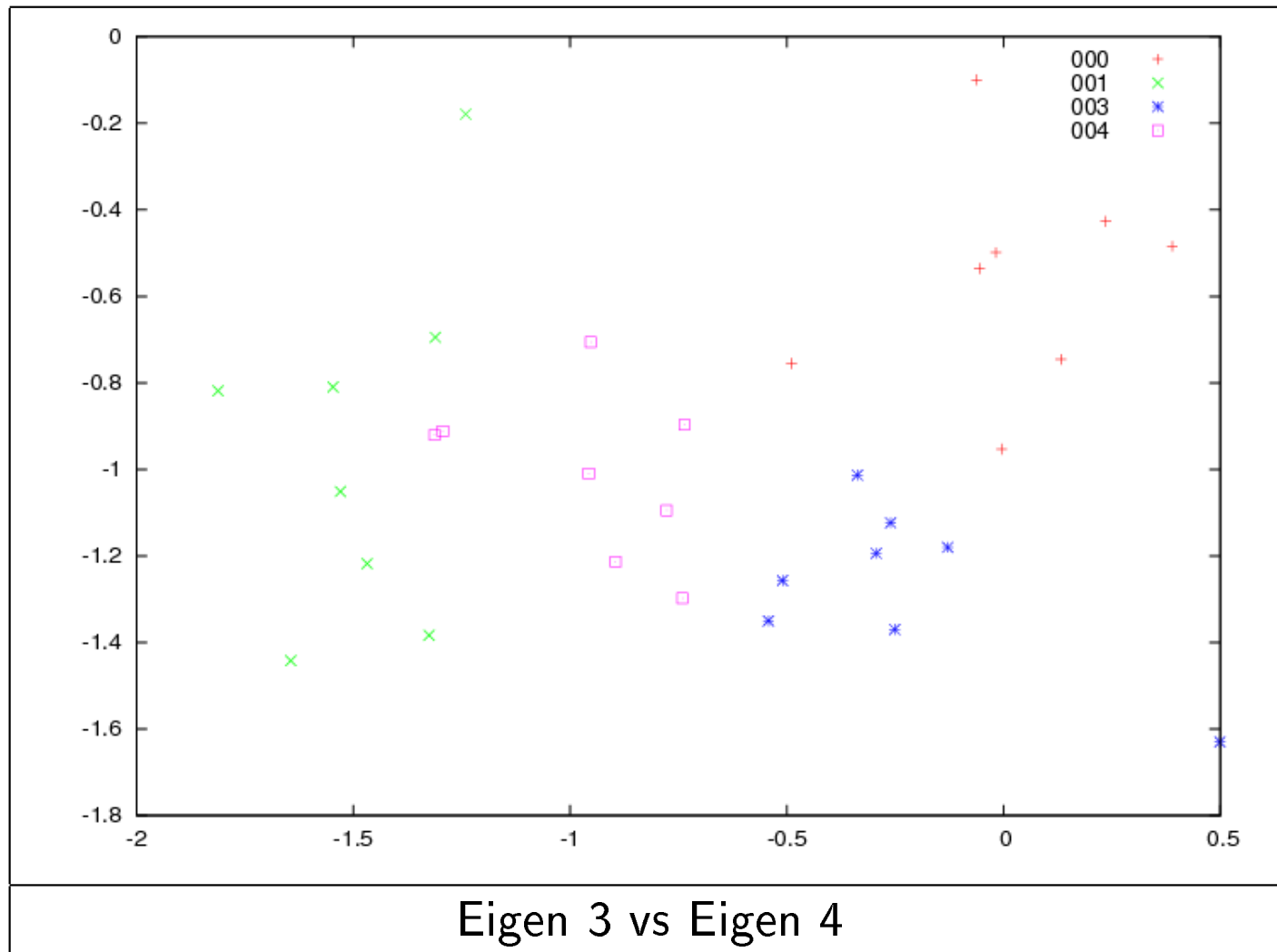
Holistic Representation: LDA

- Projections of different faces in 2D:



Holistic Representation: LDA

- Projections of different faces in 2D:

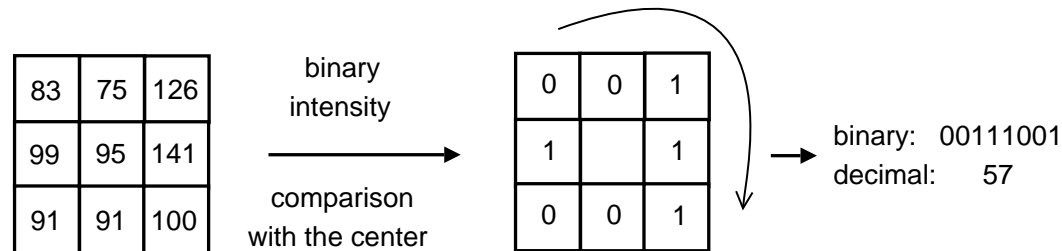


Holistic Representation: LBP

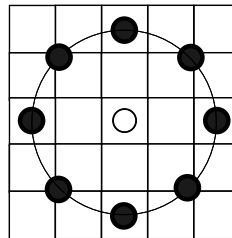
- What is LBP ?

Local Binary Patterns

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



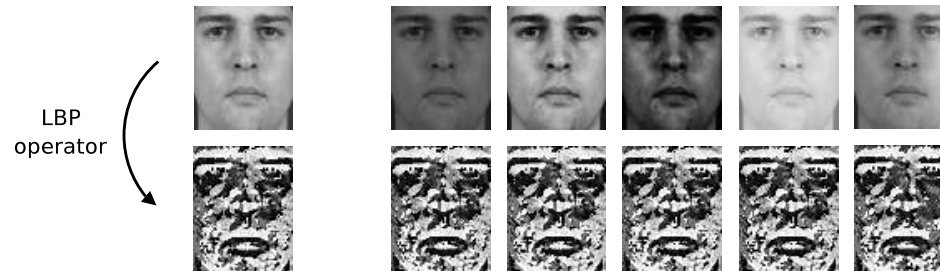
- $LBP_{P,R}$: P equally spaced pixels on a circle of radius R .



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

Local Binary Patterns

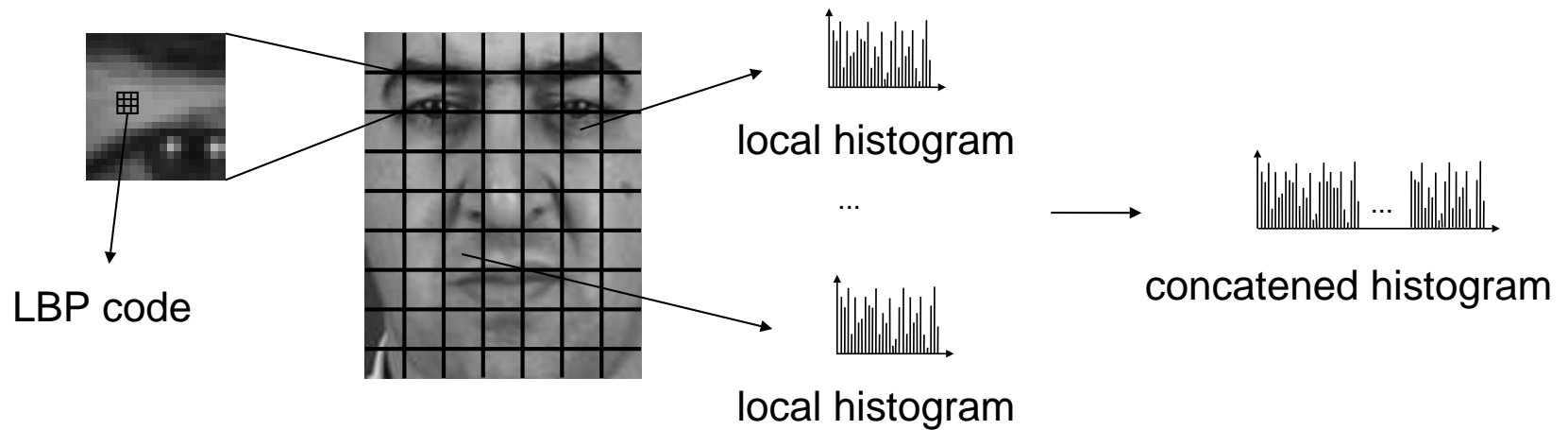
- Properties:
 - Very low computational cost
 - Powerful texture descriptor
 - Invariant to monotonic gray-scale transformation



- LBP can be computed also very quickly at any scale and position in constant time with the *integral image*.
- Potential Applications:
 - Texture classification
 - Face detection/recognition, image retrieval, motion detection, medical image analysis, surface inspection, ..

Holistic Representation: LBP

- Computes local LBP histograms,
- Concatenates local histograms into a single vector.



Holistic Representation: Drawbacks

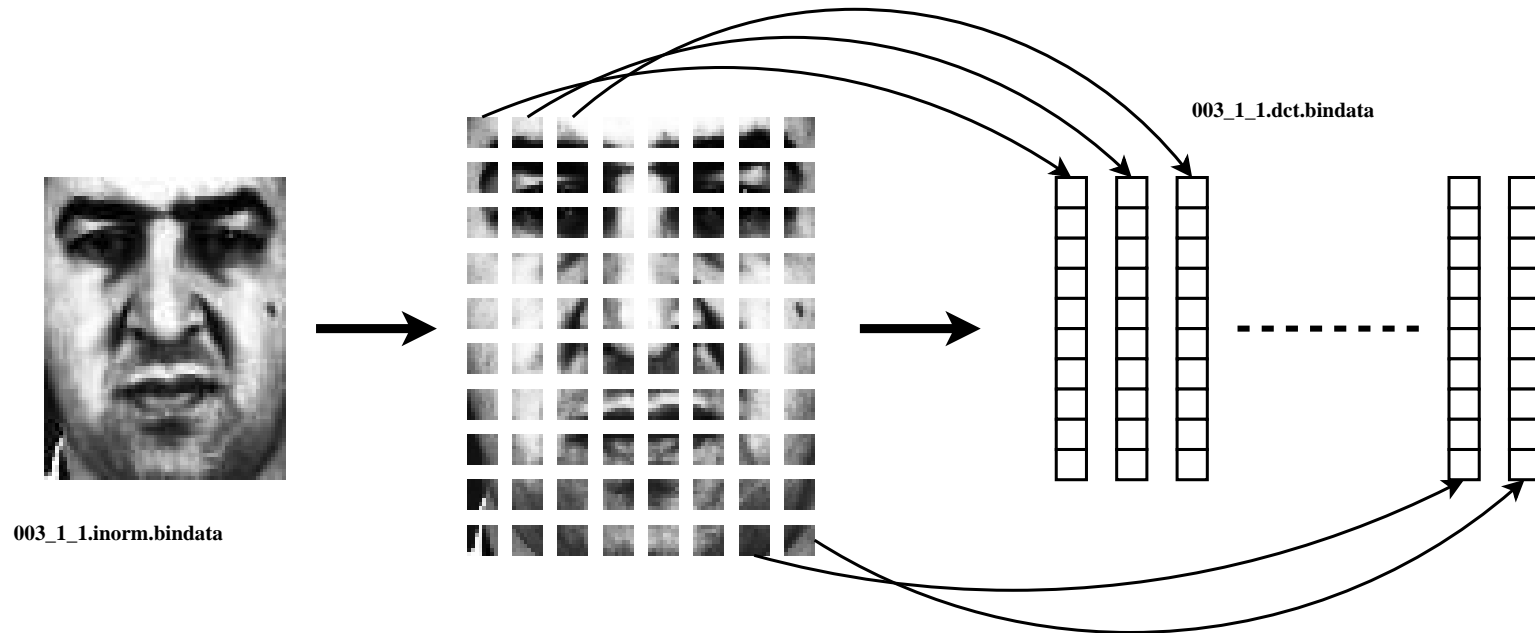
- PCA:
 - computing eigenvectors can be long (even using SVD),
 - not robust to errors in shift/scale/rotation.
- LDA: idem as PCA but also
 - works well **ONLY** in a close-set scenario,
 - number of dimensions after LDA is $< c - 1$,
 - requires enough training samples per class (Small Sample Size problem) otherwise \mathbf{S}_w is singular and thus non-invertible !
 - tricks exist to deal with the SSS problem such as the QZ algorithm.

Feature Extraction: Holistic vs Local

- Vectors are projected into a new space (the feature space):
 - Holistic representations:
 - * Turk and Pentland [1991]: *PCA*
 - * Zhao and al. [1999], Li and al. [2000]: *LDA*
 - Local representations:
 - * Local PCA, Padgett and Cottrell [1997]
 - * 2D Gabor Wavelet, Daugman [1985], Lades [1993]: Gabor filters are known as good feature detectors and such filters remove most of the variability in images that is due to variations in lighting.
 - * 2D Discrete Cosine Transform: Face images are analysed on a block by block basis. Each block is decomposed in terms of 2D Discrete Cosine Transform (DCT) basis functions. A feature vector for each block is then constructed with the DCT coefficients.
 - * Modification of the 2D DCT: Sanderson [2002] proposed the DCTmod2, where the first three DCT coefficients are replaced by their respective horizontal and vertical deltas in order to reduce the effects of illumination direction changes.

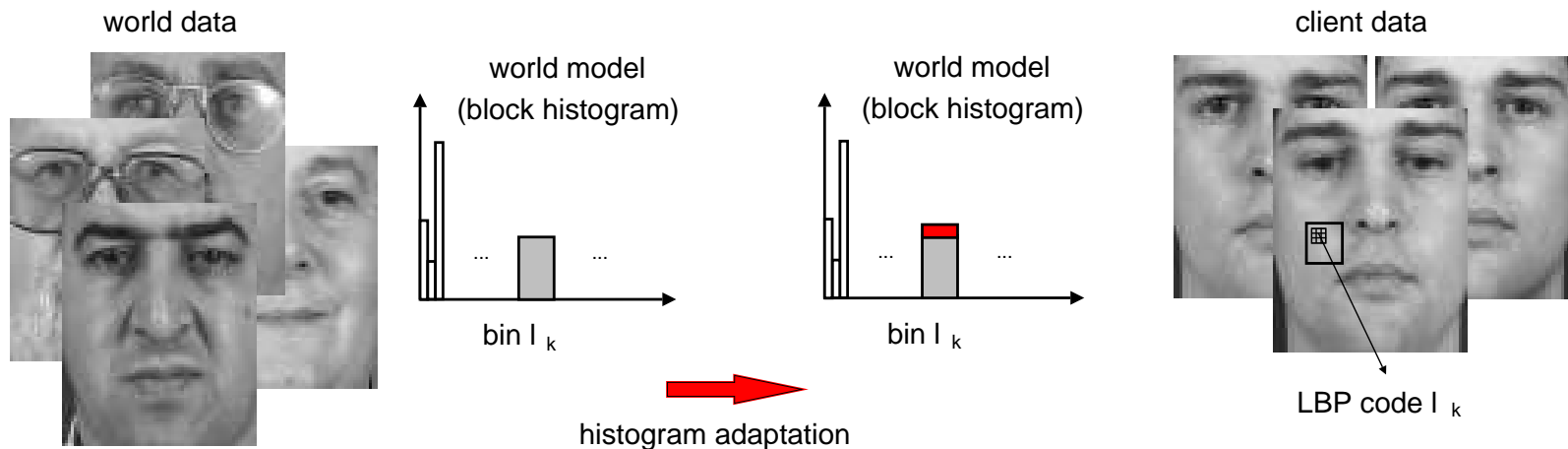
Local Representation: 2D DCT

- Divides the image into blocks (8×8 for instance),
- Computes 2D DCT in each block,
- Creates a set of T feature vectors $X = \{\mathbf{x}_t\}_{t=1}^T$ for each image.



Local Representation: LBP

- Computes local LBP histograms from world model data,
- Adapts local LBP histogram using training data.



Outline

- × Introduction
- × Biometrics
- × Face Detection (in short)
- ▷ **Face Recognition**
 - × Introduction
 - × Challenges
 - × Feature Extraction: Holistic vs Local
 - ▷ **Classification**
 - Database and Protocols
 - Experiment Results
 - Discussion
- Conclusion
- Credits

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 systems provide a score $\Lambda_I(X)$ corresponding to an opinion on the probe face pattern 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

- verification: given a threshold τ , the claim is accepted when $\Lambda_I(X) \geq \tau$ and rejected when $\Lambda_I(X) < \tau$
- closed set identification: we can recognise identity I^* corresponding to the probe face pattern X as follows

$$I^* = \arg \max_I \Lambda_I(X) \quad (3)$$

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

$$I^* = \begin{cases} \text{unknown} & \text{if } \Lambda_I(X) < \tau \ \forall I \\ \arg \max_I \Lambda_I(X) & \text{otherwise} \end{cases} \quad (4)$$

Computing the score $\Lambda_I(X)$

- Similarity measure: Euclidean, Mahalanobis [beveridge:2001], Normalised correlation [Kittler:2000], ...
- Feature based approaches: *Elastic Graph Matching* [Lades:1993] and *bunch graph* [Wiskott:1997] using Gabor filters and a labelled graph.
- Statistical model based approaches: more robust than classical approaches however they require a training process
 - a model is trained from a set of reference images for each identity,
 - and the score is then computed given a probe image and the parameters of the model corresponding to an identity.

Statistical Model based Approaches

- **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.

Generative Models

- Simple to complex models [Eickeler:2000], [Nefian:1999], [Sanderson:2003] to compute $\Lambda_C(X) = P(X|\lambda_C)$
 - Gaussian Mixture Models (GMM),
 - 1D Hidden Markov Models (1D-HMM),
 - Pseudo-2D Hidden Markov Models (P2D-HMM).
- Training:
 - using the Maximum Likelihood (ML) criterion via the Expectation Maximisation (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.

Generative Models

- Let us denote the parameter set for client C as λ_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 $X = \{\mathbf{x}_t\}_{t=1}^T$ supporting the claim (extracted from the given face).
- 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 τ , the claim is accepted when $\Lambda(X) \geq \tau$ and rejected when $\Lambda(X) < \tau$.

Gaussian Mixture Model

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

$$P(X|\lambda) = \prod_{t=1}^T P(\mathbf{x}_t|\lambda) \quad (5)$$

where

$$P(\mathbf{x}|\lambda) = \sum_{k=1}^{N_G} m_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k) \quad (6)$$

$$\lambda = \{m_k, \mu_k, \Sigma_k\}_{k=1}^{N_G} \quad (7)$$

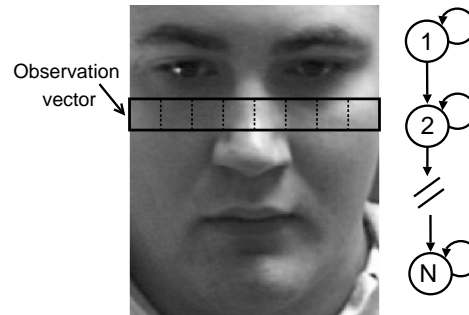
- $\mathcal{N}(\mathbf{x}|\mu, \Sigma)$ is a D -dimensional gaussian density function with mean μ and diagonal covariance matrix Σ .
- N_G is the number of gaussians and m_k is the weight for gaussian k (with constraints $\sum_{k=1}^{N_G} m_k = 1$ and $\forall k : m_k \geq 0$).

Gaussian Mixture Model

- 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 localisation 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

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



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

1D-Hidden Markov Model

- N , the number of states in the model; each state corresponds to a region of the face; $S = \{S_1, S_2, \dots, S_N\}$ 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:

$$a_{ij} = \begin{cases} P(q_t = S_j | q_{t-1} = S_i) & \text{for } j = i, j = i + 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

- The state probability distribution $B = \{b_j(\mathbf{x}_t)\}$, where

$$b_j(\mathbf{x}_t) = P(\mathbf{x}_t | q_t = S_j) \quad (9)$$

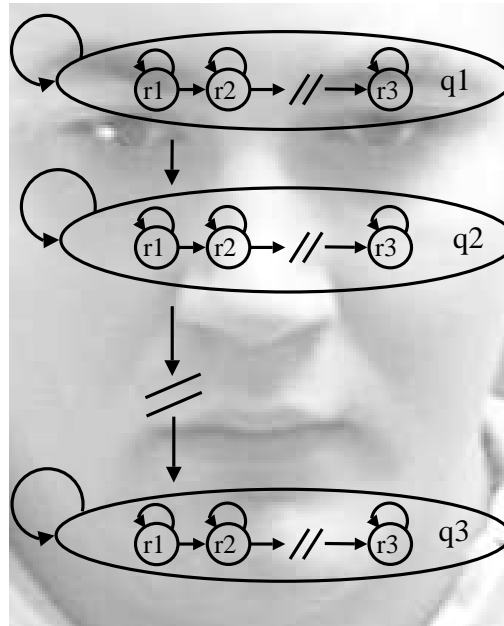
The features are expected to follow a continuous distribution and are modelled with mixtures of gaussians.

1D Hidden Markov Model

- 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 localisation).

Pseudo-2D Hidden Markov Model

- 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 modelled by a mixture of gaussians.

Pseudo-2D Hidden Markov Model

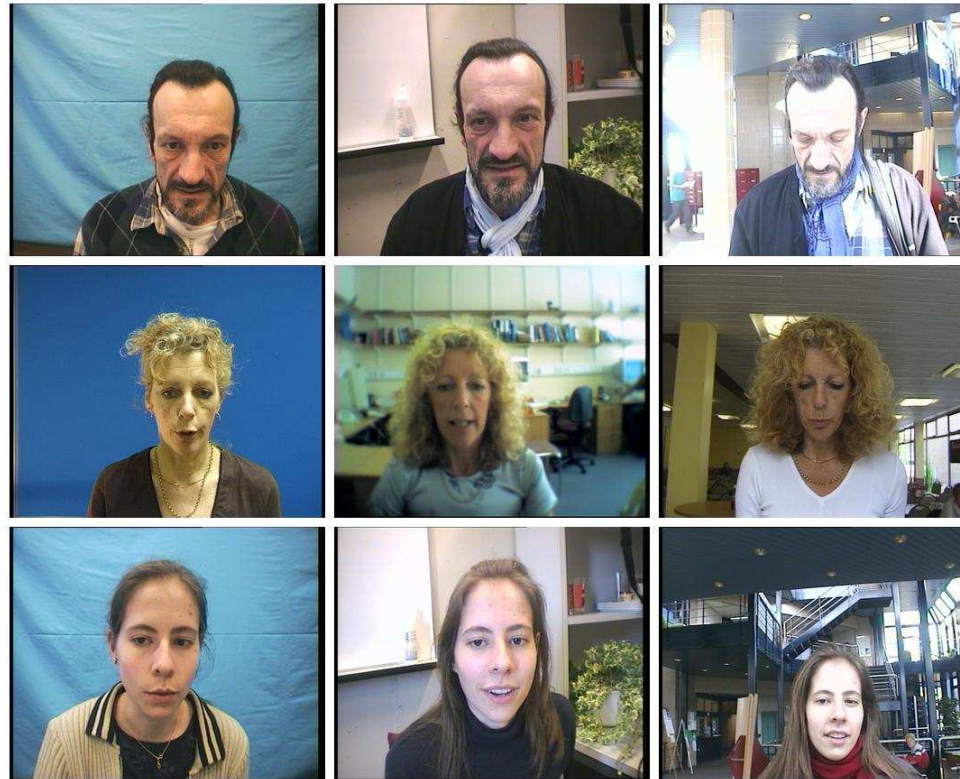
- The degree of spatial constraints present in the P2D-HMM approach can be thought of as being somewhere in between the GMM and the 1D-HMM approaches. While the GMM approach has no spatial constraints and the 1D-HMM has rigid horizontal constraints, the P2D-HMM approach has relaxed constraints in both 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.

Outline

- × Introduction
- × Biometrics
- × Face Detection (in short)
- ▷ **Face Recognition**
 - × Introduction
 - × Challenges
 - × Feature Extraction: Holistic vs Local
 - × Classification
 - ▷ **Database and Protocols**
 - Experiment Results
 - Discussion
- Conclusion
- Credits

Database

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



- 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

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

Test Sessions	Train Sessions			
	1	5	9	1,5,9
C: 2-4 I: 1-4	Mc			
C: 6-8 I: 5-8	Ud	Md		
C: 10-12 I: 9-12	Ua		Ma	
C: 2-4,6-8,10-12 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

- 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):

$$\text{FAR} = \frac{\text{number of FAs}}{\text{number of impostor accesses}} \quad (10)$$

$$\text{FRR} = \frac{\text{number of FRs}}{\text{number of true claimant accesses}} \quad (11)$$

- 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):

$$\text{HTER} = \frac{\text{FAR} + \text{FRR}}{2} \quad (12)$$

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 <i>ML</i>	12.9	28.9	26.0	22.9
GMM <i>init</i>	12.8	29.7	28.3	23.8
GMM <i>MAP</i>	8.9	17.3	20.9	17.0
1D-HMM <i>ML</i>	9.1	17.8	17.1	15.9
1D-HMM <i>init</i>	9.1	15.6	17.4	14.7
1D-HMM <i>MAP</i>	6.9	16.3	17.0	14.7
P2D-HMM <i>ML</i>	9.0	19.0	18.0	17.5
P2D-HMM <i>init</i>	8.6	16.5	19.2	17.0
P2D-HMM <i>MAP</i>	* 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 <i>ML</i>	16.7	33.3	33.3	27.7
GMM <i>init</i>	19.8	35.0	35.1	29.7
GMM <i>MAP</i>	9.5	21.0	24.8	19.5
1D-HMM <i>ML</i>	21.0	28.8	29.5	27.0
1D-HMM <i>init</i>	21.3	30.1	31.4	28.1
1D-HMM <i>MAP</i>	13.8	25.9	23.4	21.7
P2D-HMM <i>ML</i>	12.1	25.2	26.9	22.3
P2D-HMM <i>init</i>	13.5	24.6	26.5	22.5
P2D-HMM <i>MAP</i>	* 6.5	* 15.9	* 14.7	* 14.7

Discussion

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

Outline

- × Introduction
- × Applications
- × Face Detection (in short)
- × Face Recognition
- ▷ Conclusion
- Credits

Conclusion

- Face detection:
 - boosting-based methods provide an interesting trade-off between accuracy and speed,
 - LBP features are good candidates for face detection,
- Face recognition:
 - generative models outperform state-of-the-art systems in unconstrained conditions,
 - more general generative models (Bayesian Networks) is probably the next step.

Outline

- × Introduction
- × Applications
- × Face Detection (in short)
- × Face Recognition
- × Conclusion
- ▷ Credits

Credits

- PhD students: G. Heusch, A. Just, Y. Rodriguez, F. Cardinaux
- Torch and Torchvision <http://torch3vision.idiap.ch>



- pyVerif <http://pyverif.idiap.ch>

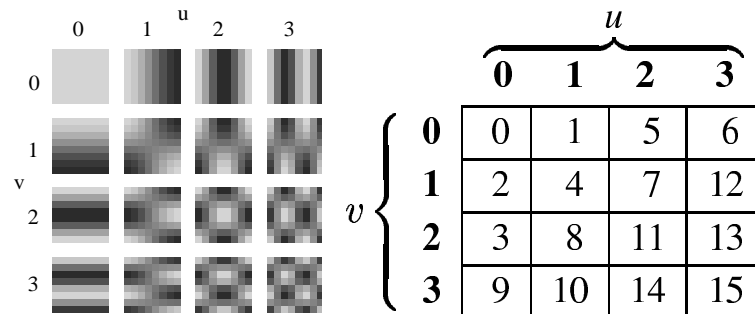


- More info available at <http://www.idiap.ch/~marcel>
- Demos available at <http://www.idiap.ch/~marcel/demos.php>

Additional Slides

2D Discrete Cosine Transform (1)

- Face image is analysed on a block by block basis
 - Each block is 8×8
 - 50% overlap
- Blocks decomposed in terms of 2D DCT basis functions (64)
- Coefficients are ordered according to a zig-zag pattern, reflecting the amount of information stored



2D Discrete Cosine Transform (2)

- For block located at (b, a) , the DCT feature vector is composed of:

$$\vec{x}^{(b,a)} = \left[c_0^{(b,a)} \quad c_1^{(b,a)} \quad \dots \quad c_{M-1}^{(b,a)} \right]^T$$

- Only need to retain 15 coefficients (24%)
- A face image of 56×64 (rows \times columns) is described by a set of 195 vectors
- Works well in non-challenging conditions
- Problem: illumination changes (intensity, direction)

DCT-mod2 (1)

- Most affected coefficients: c_0 c_1 c_2
- Throw them out ?
 - reduces performance
 - $\therefore c_0$ c_1 c_2 contain discriminative information
- Replace c_0 c_1 c_2 with their deltas
- In simplest form, deltas are differences between coefficients from neighbouring blocks

DCT-mod2 (2)

- Modify 2D DCT feature extraction by replacing the first 3 coefficients with their horizontal and vertical deltas:

$$\vec{x} = \left[\Delta^h c_0 \quad \Delta^v c_0 \quad \Delta^h c_1 \quad \Delta^v c_1 \quad \Delta^h c_2 \quad \Delta^v c_2 \quad c_3 \quad c_4 \quad \dots \quad c_{M-1} \right]^T$$

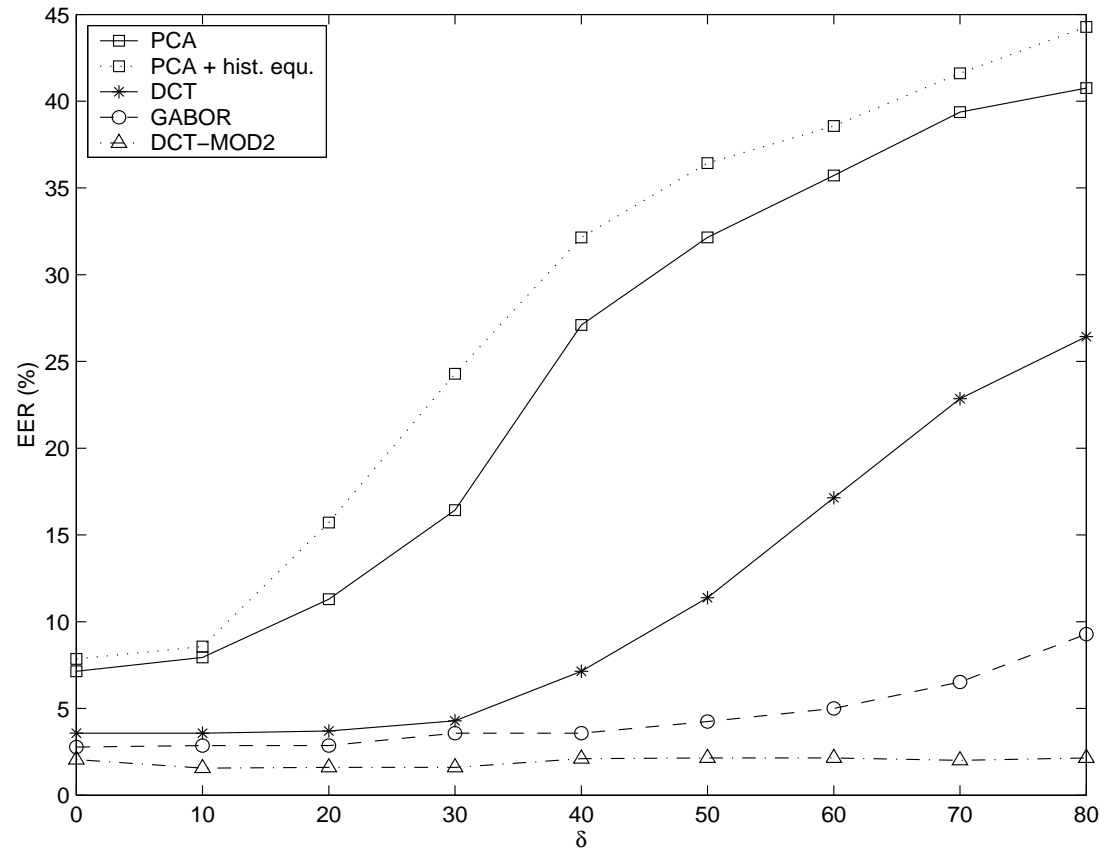
- Refer to this approach as *DCT-mod2*
- Compare *DCT-mod2* with DCT, PCA, PCA with histogram equalisation and 2D Gabor wavelet based features
- Use an artificial illumination direction change:



- left: $\delta = 0$ (no change); middle: $\delta = 40$; right: $\delta = 80$

DCT-mod2 (3)

- Results obtained using a GMM based classifier in a verification scenario:



Integral Image

- The *integral image* representation or *summed area table* was first introduced by [Crow:1984] for texture mapping,
- At a given location $(x; y)$ in an image, the value of the *integral image* $ii(x; y)$ is the sum of the pixels above and to the left of $(x; y)$:

$$ii(x; y) = \sum_{x' \leq x, y' \leq y} i(x'; y'),$$

where $i(x'; y')$ is the pixel value of the original image at location $(x'; y')$

- If $s(x; y)$ is the cumulative row sum, with $s(x; -1) = 0$ and $s(-1; y) = 0$, the *integral image* can be computed in one pass over the original image using the following pair of recurrences:

$$s(x; y) = s(x; y - 1) + i(x; y), \quad (13)$$

$$ii(x; y) = ii(x - 1; y) + s(x; y). \quad (14)$$