Face Processing

from face detection to face recognition

Prof Sébastien Marcel
Senior researcher
www.idiap.ch/∼marcel

Idiap Research Institute
Martigny, Switzerland
www.idiap.ch

January 25, 2010

Overview

Outline

Outline

1. Face Recognition
   - Face Recognition
   - Challenges
   - Normalization
   - Feature Extraction
   - Classification
   - Statistical Model based Approaches
   - Evaluation
   - Research Directions

Face Recognition

Applications

- 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.

Outline

1. Face Recognition
   - Face Recognition
   - Challenges
   - Normalization
   - Feature Extraction
   - Classification
   - Statistical Model based Approaches
   - Evaluation
   - Research Directions

Challenges

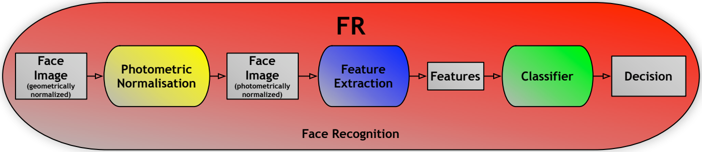
Challenges

- 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 and Non-Perfect Detection,
  - Occlusion and 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, due to different expression, lighting, background, head pose, hair cut, etc.

Methods for Face Recognition

General FR system

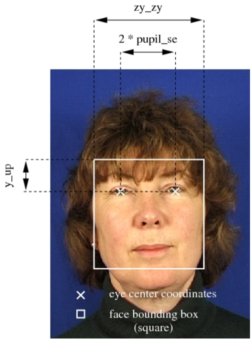
1. geometric normalization (see part on face detection/alignment),
2. photometric normalization (or illumination normalization),
3. feature extraction,
4. classification.



Outline

1. Face Recognition
   - Face Recognition
   - Challenges
   - Normalization
   - Feature Extraction
   - Classification
   - Statistical Model based Approaches
   - Evaluation
   - Research Directions

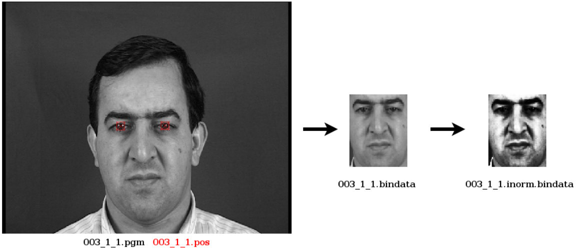
Geometric Normalization



Goal

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

Photometric Normalization



Techniques

1. Histogram Equalization (HEQ),
2. Multi-scale Retinex (Illuminance-reflectance model),
3. Self-Quotient Image,
4. Diffusion (Gross and Brajovic),
5. Local Binary Patterns.

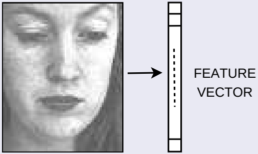
Outline

1. Face Recognition
   - Face Recognition
   - Challenges
   - Normalization
   - Feature Extraction
   - Classification
   - Statistical Model based Approaches
   - Evaluation
   - Research Directions

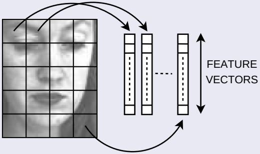
Feature Extraction

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

- 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.

Feature Extraction

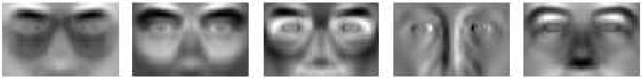
Local

- 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 analyzed 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 Binary Patterns: Ahonen [2005] proposed to use LBP histograms computed in face regions

Principal Component Analysis (PCA)

Eigenfaces

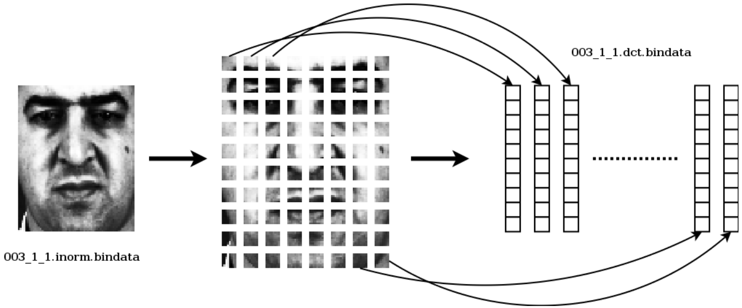
- $\mathbf{x}_i \in \mathbb{R}^n$ is a face image which $n = w \times h$,
- $\mathbf{W} = [\mathbf{e}_1 ... \mathbf{e}_m]^T$:



Discrete Cosine Transform (DCT)

Eigenfaces

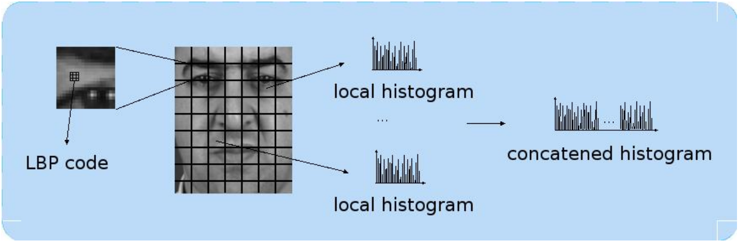
- the face image is decomposed into blocks,
- the DCT is computed for each block



Local Binary Patterns (LBP)

Histogram of LBP

- the face image is decomposed into blocks,
- a LBP histogram is computed for each block.



Outline

1. Face Recognition
   - Face Recognition
   - Challenges
   - Normalization
   - Feature Extraction
   - Classification
   - Statistical Model based Approaches
   - Evaluation
   - Research Directions

Classification

Definitions

- 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 $\Lambda_I(X)$ 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

Definitions

- 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 recognize identity I^* corresponding to the probe face pattern X as follows

$$I^* = \arg\max_I \Lambda_I(X) \tag{1}$$

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

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

Computing the score $\Lambda_I(X)$

Similarity measures

Euclidean, Mahalanobis [beveridge:2001], Normalized correlation [Kittler:2000], χ^2 [Ahonen:2004], ...

Feature based approaches

Elastic Graph Matching [Lades:1993] and *bunch graph* [Wiskott:1997] using Gabor filters and a labeled graph.

Statistical model based approaches

more robust than classical approaches but require a training:

- 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.

Outline

1. Face Recognition
   - Face Recognition
   - Challenges
   - Normalization
   - Feature Extraction
   - Classification
   - Statistical Model based Approaches
   - Evaluation
   - Research Directions

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.

Generative Models

Simple to complex models to compute $\Lambda_C(X) = P(X|\lambda_C)$

- 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].

Generative Models

Training

- 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.

Generative Models

Decision

- 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 $X = \{\vec{x}_t\}_{t=1}^{T}$ supporting the claim (extracted from the given face).
- We find an opinion on the claim using $\Lambda(X) = \log P(X|\lambda_C) - \log P(X|\lambda_{\overline{C}})$

Generative Models

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 τ, 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

$$P(\vec{x}|\lambda) = \sum_{k=1}^{N_G} m_k \mathcal{N}(\vec{x}|\vec{\mu}_k, \Sigma_k) \tag{4}$$

$$\lambda = \{m_k, \vec{\mu}_k, \Sigma_k\}_{k=1}^{N_G} \tag{5}$$

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 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

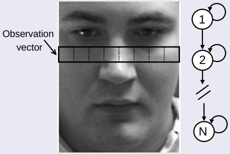
GMM

- 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

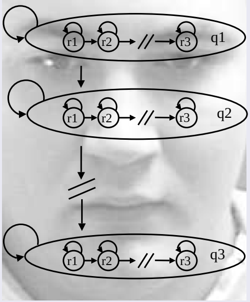
1D HMM

- 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.

Outline

1. Face Recognition
   - Face Recognition
   - Challenges
   - Normalization
   - Feature Extraction
   - Classification
   - Statistical Model based Approaches
   - Evaluation
   - Research Directions

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	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

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.

Outline

1. Face Recognition
   - Face Recognition
   - Challenges
   - Normalization
   - Feature Extraction
   - Classification
   - Statistical Model based Approaches
   - Evaluation
   - Research Directions

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),
- spoofing.

Directions

- Local Binary Patterns could be used to reduce the effect of illumination,
- Bayesian Networks provide elegant generative models able to fuse multiple cues,
- Joint Factor Analysis and UBM-GMM Super-Vectors techniques could be used for model adaptation.