Benchmarks for Facial Image Analysis Technologies (BeFIT 2012)

On the importance of benchmarking in facial image analysis

Dr Sébastien Marcel
Senior research scientist
www.idiap.ch/~marcel

Idiap Research Institute
Martigny, Switzerland
www.idiap.ch

October 13, 2012
Outline

1. Introduction
   - Introduction
   - Facial image analysis
   - Applications

2. So what?

3. Facial image analysis

4. The Scientific Method

5. Conclusion
Introduction

Dr S. Marcel – Idiap 2012

On the importance of benchmarking in facial image analysis
OMG

Face recognition is solved!
Facial image analysis

- Face detection
- Facial feature localization
- Age estimation
- Gender recognition
- Face recognition
- Expression recognition
- Face tracking
Facial image analysis

Applications

- Security applications (face biometrics)
- Multimedia information management (face indexing & retrieval)
- Human computer interface (video games, personal settings)
Face biometrics

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.
Face indexing and retrieval

Examples

- Nevenvision: face detection and recognition (acquired by Google in 2006),
- Riya and Like: face detection, face recognition and text recognition, object recognition (acquired by Google),
- PittPatt: face detection and recognition (acquired by Google),
- PolarRose (www.polarrose.com): face detection and recognition (acquired by Apple),
- Google (www.google.com) with Picasa (picasa.google.com),

Demonstration

Google Portrait
Outline

1. Introduction
2. So what?
3. Facial image analysis
4. The Scientific Method
5. Conclusion
Commercial

Face Recognition

- Is the face recognition problem solved?
- Can I do better (or faster)?
Face detection

- Is the face detection problem solved?
- Can I do better (or faster)?
• Is facial feature localization accurate enough?
• Can I do better (or faster)?
Face recognition

- Is the face recognition problem solved?
- Can I do better (or faster)?
Peer reviews

- “results should be compared to state-of-the-art”
- “authors forgot to cite important prior work from ... and to compare with their proposed approach”

Rebuttal

- What is the state-of-the-art?
- How should we do this comparison? Often it turns out that the prior work is not comparable (no open results or not the same database)
The whole point is about evaluation!

Data and Protocol

- Am I using adequate data for my problem?
- Is this open data? (replicable research)
- Is it provided with a detailed protocol? (fair and unbiased evaluations)
The whole point is about evaluation!

**Data and Protocol**
- Am I using adequate data for my problem?
- Is this open data? (replicable research)
- Is it provided with a detailed protocol? (fair and unbiased evaluations)

**Reference**
- Is there prior work?
- Is it using the same data? (comparable experiments)
- Is there open results? (compare on same plots)
- Is it open source? (replicable research)
The whole point is about evaluation!

Data and Protocol
- Am I using adequate data for my problem?
- Is this open data? (replicable research)
- Is it provided with a detailed protocol? (fair and unbiased evaluations)

Reference
- Is there prior work?
- Is it using the same data? (comparable experiments)
- Is there open results? (compare on same plots)
- Is it open source? (replicable research)

Metric
- Is there any existing metric already?
- Is there any tools available?
Outline

1. Introduction

2. So what?

3. Facial image analysis
   - Overview
   - Face detection
   - Facial feature localization
   - Face recognition

4. The Scientific Method

5. Conclusion
Overview

Focus on 3 core tasks

1. face detection
2. facial feature localization (face alignment)
3. face recognition
Face detection

Example
Face detection

Goal of face detection

- any number
- any location and scale
- any pose appearance (rotated, profile)
Face detection (FD)

Examples

Classification problem
- two-class: face vs background
- multi-class: face with multiple pose vs background
- reference methods: MLP, SVM, boosting ...
Standard FD approach

3 elements
Standard FD approach

Grid Spacing
Scanning Sub-window
Face/Non-face Classification

3 elements

1 scanning an image
Standard FD approach

- Grid Spacing
- Scanning Sub-window
- Scale (Pyramid scan)
- Face / Non-face Classification

3 elements

1. scanning at multiple scales
Standard FD approach

3 elements

1. scanning (image pyramid, integral image)
2. classification (MLP, SVM, Boosted classifier)
Standard FD approach

1. scanning (image pyramid, integral image)
2. classification (MLP, SVM, Boosted classifier)
3. merging (non-maxima suppression, heuristics)
Standard FD metrics

**Detection Rate (DR)**

% of face correctly detected
Synonyms: hit rate, true positive rate (TPR)

**False Alarms (FA)**

Number (not %) of false detections
Synonyms: false positive (FP), false accept (FA)
Standard FD metrics

What is a correct detection?
**Standard FD metrics**

**Jesorsky’s relative error measure**

\[
d_{\text{eye}} = \frac{\max(d(C_l, \tilde{C}_l), d(C_r, \tilde{C}_r))}{d(C_l, C_r)}
\]

- \( C_l \) and \( C_r \) represent the true eye positions,
- \( \tilde{C}_l \) and \( \tilde{C}_r \) represent the detected eye positions,
- \( C_0 \) (resp. \( \tilde{C}_0 \)) is the middle of the segment \([C_l C_r]\) (resp. \([\tilde{C}_l \tilde{C}_r]\)).

**Correct face detection**

if: \( d_{\text{eye}} \leq 0.25 \)
Standard FD metrics

**Jesorsky’s relative error measure**

- $I_d(\epsilon_j = 0.25)$
- $I_d(\epsilon_j = 0.10)$
- $I_g$

- $r_d(\epsilon_j = 0.25)$
- $r_d(\epsilon_j = 0.10)$
- $r_g$

- 10 px
- 25 px

- $D = 100$ px

**Limitation**

- Not adapted to out-of-plane rotations of the face
Standard FD metrics

Jaccard distance

similarity between two sets $A$ and $B$:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

<table>
<thead>
<tr>
<th>J(A, B)</th>
<th>Example Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>![Image 1]</td>
</tr>
<tr>
<td>0.20</td>
<td>![Image 2]</td>
</tr>
<tr>
<td>0.50</td>
<td>![Image 3]</td>
</tr>
<tr>
<td>0.80</td>
<td>![Image 4]</td>
</tr>
<tr>
<td>1.00</td>
<td>![Image 5]</td>
</tr>
</tbody>
</table>

Advantage

Adequate for multi-pose face detection
Standard FD plot

Receiver Operating Characteristic (ROC)

Plots the True Positive Rate (TPR) vs False Positive Rate (FPR):
Real ROC for FD (DR vs FA)

MIT+CMU

EPT-BOOT
VAR-BOOT
Viola
FCBoost
Facial feature localization

Goal of facial feature localization

- locates key-points on the face (eyes, nose-tip, mouth, ...)
- any location and scale
- any pose appearance (rotated, profile)
Facial feature localization (FFL)

Examples

Classification or Regression problem

- multi-class: one-hot-encoding, one-versus-all
- regression: predicts location
- reference methods: ASM, AAM, CLM, BoRMaN
## Standard FFL approach

### 3 elements

1. **observation model** (local for each point, global)
2. **shape model** (statistical)
3. **fitting algorithm** (iterative search until convergence)
Standard FFL metrics

Average point-to-point distance

- let $F$ be a set of facial features
- let $G_i$ be the ground truth locations
- let $D_i$ be the predicted locations for each point
- let $d(.,.)$ the Euclidean distance between two-dimensional points
- let $\Delta = d(G_{\text{leye}}, G_{\text{reye}})$ the distance between the eye centers

$$E(D, G) = \frac{1}{\Delta} \sum_{1 \leq i \leq F} d(D_i, G_i)$$ (2)
Cumulated error distribution

Plots the cumulated distribution of $E$. 
Face recognition (FR)

Standard FR approach

1. geometric normalization (face detection/alignment),
2. photometric normalization (or illumination normalization),
3. feature extraction,
4. classification (training, enrollment and matching).
# Face recognition

## Training (optional)
Compute a prior model $\theta_I$ (PCA, UBM)

## Enrollment
Compute a models $\theta_i$ (also called template) from image(s) $X_i^e$ of each individual $i$

## Matching
Compute a matching score $\Lambda_i(X^m|\theta_i, \theta_I)$ of image(s) $X^m$ for a given identity $i$

## Reference methods
- PCA×LDA+distance ($L^2$, cosine, ...), EBGM+distance (Gabor phase similarity), LBPHS+distance ($\chi^2$), DCT-GMM+LLR
## Classification

Consists of attributing a label to the input data and differs according to the specific task (closed or open set identification, verification).

## Task-specific label

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*
Face recognition

**Classification**

- **verification:** given a threshold $\tau$, 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

\[
i^* = \arg\max_i \Lambda_i(X)
\]  

(3)

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

\[
i^* = \begin{cases} 
    \text{unknown} & \text{if } \Lambda_i(X) < \tau \ \forall \ i \\
    \arg\max_i \Lambda_i(X) & \text{otherwise}
\end{cases}
\]  

(4)
Standard FR metrics

**False Reject Rate (FRR)**

% of identities incorrectly rejected
Synonyms: False Negative Rate (FNR), 1 minus Correct Acceptance Rate (CAR), 1 minus True Positive Rate (TPR)

**False Accept Rate (FAR)**

% of impostors incorrectly accepted
Synonyms: False Positive Rate (FPR)
Detection Error Trade-off (DET)

Plots the False Reject Rate (FRR) and False Accept Rate (FAR)
Real DET for FR (FRR vs FAR)

Dr S. Marcel – Idiap 2012

On the importance of benchmarking in facial image analysis
Discussion

Let’s recap

- Facial image analysis technologies are complex,
- multiple algorithms,
- various databases,
- various metrics.

Consequence

- A lot of common resources (databases, annotations, source code, tools) are not systematically shared and very often created again
- Wild West: databases/protocols, references and metrics

Need for some methodology to keep focus

The Scientific Method
Outline

1. Introduction
2. So what?
3. Facial image analysis
4. The Scientific Method
   - The Scientific Method
   - The Scientific Method: application to face detection
   - Replicable research
5. Conclusion

Dr. S. Marcel – Idiap 2012

On the importance of benchmarking in facial image analysis
1. Observation

Observation of a phenomenon, formulation of a question, literature review (actual research for prior work)
The Scientific Method

1. Observation
Observation of a phenomenon, formulation of a question, literature review (actual research for prior work)

2. Hypothesis
Propose a solution to the problem
# The Scientific Method

1. **Observation**

Observation of a phenomenon, formulation of a question, literature review (actual research for prior work)

2. **Hypothesis**

Propose a solution to the problem

3. **Prediction**

Describe the possible outcome of the hypothesis
The Scientific Method

1. Observation
Observation of a phenomenon, formulation of a question, literature review (actual research for prior work)

2. Hypothesis
Propose a solution to the problem

3. Prediction
Describe the possible outcome of the hypothesis

4. Experimentation
Test the hypothesis
## The Scientific Method

1. **Observation**
   - Observation of a phenomenon, formulation of a question, literature review (actual research for prior work)

2. **Hypothesis**
   - Propose a solution to the problem

3. **Prediction**
   - Describe the possible outcome of the hypothesis

4. **Experimentation**
   - Test the hypothesis

5. **Analysis**
   - Draw conclusions and communicate results
The Scientific Method

Graphical view

Observe natural phenomena → Formulate Hypothesis → Test hypothesis via rigorous Experiment → Establish Theory based on repeated validation of results

Dr S. Marcel – Idiap 2012

On the importance of benchmarking in facial image analysis
The Scientific Method

**Graphical view**

**THE SCIENTIFIC METHOD**

1. Observe natural phenomena
2. Formulate Hypothesis
3. Test hypothesis via rigorous Experiment
4. Establish Theory based on repeated validation of results

**THE ACTUAL METHOD**

1. Make up Theory based on what Funding Agency Manager wants to be true
2. Design minimum experiments that will prove show? suggest Theory is true
3. Publish Paper: rename Theory a “Hypothesis” and pretend you used the Scientific Method
4. Defend Theory despite all evidence to the contrary
The Scientific Method

1. Observation
Observation of a phenomenon, formulation of a question, literature review (actual research for prior work)

2. Hypothesis
Propose a solution to the problem

3. Prediction
Describe the possible outcome of the hypothesis

4. Experimentation
Test the hypothesis

5. Analysis
Draw conclusions and communicate results

On the importance of benchmarking in facial image analysis
The Scientific Method: Analysis

Only 2 options:

- you CAN reject the hypothesis
  results don’t support the hypothesis

- you CANNOT reject the hypothesis
  results support the hypothesis BUT no definite conclusion since an error is always possible (bug, experiment flaw, ...)

Dr S. Marcel – Idiap 2012
On the importance of benchmarking in facial image analysis
The Scientific Method

1. Observation
Observation of a phenomenon, formulation of a question, literature review (actual research for prior work)

2. Hypothesis
Propose a solution to the problem

3. Prediction
Describe the possible outcome of the hypothesis

4. Experimentation
Test the hypothesis

5. Analysis
Draw conclusions and communicate results
The Scientific Method: Experimentation

Groups

- Experimental group: measuring the prediction by changing the independent variable
- Control group: measuring the prediction without changing the independent variable (baseline or reference)

Variable

- Independent variable: the variable you are changing on the experimental group
- Dependent variable: the variable you are measuring to assess your prediction
- Controlled variables: the variables that are the same in the 2 groups
The Scientific Method: application to face detection

Is it Mission Impossible?
The Scientific Method: application to face detection

1. Observation: formulate a question

How to detect more (frontal) faces in difficult image conditions?

2. Hypothesis: propose a solution

Does the use of Local Binary Patterns (LBP) increase the face detection rate (DR)?

3. Prediction: describe possible outcome

The use of LBP will increase the face DR compared to regular Haar-like features?
4. Experimentation: test the hypothesis

- Experimental group: LBP-based FD trained on database A and tested on database B
- Control group: Haar-based FD trained on database A and tested on database B
- Independent variable: LBP parameterization
- Dependent variable: DR
- Controlled variables: databases A and B, scanning-classifier-merging algorithms

Result

The DR of LBP-based FD > the DR of Haar-based FR
5. Analysis: draw a conclusion

results support the hypothesis hence we CANNOT reject the hypothesis formulated

Future work

might test also the hypothesis and draw a similar or a different conclusion, then revise the hypothesis, test it, draw a conclusion and so on ...
The Scientific Method: requirements for Experimentation

Data

to perform runs on experimental and control groups

Code

to control variables

Results

to perform runs or to use baselines (control group), dependent and controlled variables
The Scientific Method: requirements for Experimentation

**Data**
- to perform runs on experimental and control groups

**Code**
- to control variables

**Results**
- to perform runs or to use baselines (control group), dependent and controlled variables

**Replicable Research**
- results from a paper can be reproduced (easily) with code and data available online
### Replicable research (RR)

<table>
<thead>
<tr>
<th>Open data</th>
<th>public data available (online, HDD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open source</td>
<td>source code available, or executable eventually, for reference methods and for tools used to compute performance</td>
</tr>
<tr>
<td>Open results</td>
<td>detailed experiment setups (protocols, scripts, parameterizations)</td>
</tr>
</tbody>
</table>
Replicable research (RR)

Open data
public data available (online, HDD)

Open source
source code available, or executable eventually, for reference methods and for tools used to compute performance

Open results
detailed experiment setups (protocols, scripts, parameterizations)

RR is widely used
in Human Sciences and Physics

Much less used
in Computational Science!
Outline

1. Introduction
2. So what?
3. Facial image analysis
4. The Scientific Method
5. Conclusion
   - Discussion
   - Prior and current work
   - Future work
   - Thank you
Discussion

Why no more RR in FIA?

Incentives!
Discussion

Why no more RR in FIA?

Incentives!

Companies

$$$, protecting assets and knowledge. No incentive to disclose trade secrets (of course).
Discussion

Why no more RR in FIA?
Incentives!

Companies
$$$, protecting assets and knowledge. No incentive to disclose trade secrets (of course).

However
How/why to trust a product? A biometric product for instance?
Discussion

Why no more RR in FIA?
Incentives!

Academia
Impact (h-index) which is obtained by:
  - increasing the number of publications in conference/journals with high impact factor,
  - by focusing on innovations and trends,
  - by proposing a breakthrough method.
Hence no time for RR as priority is given to publications.
Discussion

Why no more RR in FIA?

Incentives!

Academia

Impact (h-index)

Incentives to RR!

Papers available online are cited 3 times more. Papers using online datasets have 69% more citations.
Discussion

Why no more RR in FIA?
Incentives!

Academia
Impact (h-index)

Incentives to RR!
Papers available online are cited 3 times more. Papers using online datasets have 69% more citations.

However
Is there another complementary/alternative direction?
Prior/ongoing work

Competitions
FRGC, XM2VTS, BANCA, MOBIO

Open source
OpenCV, VXL, RAVL, torch3vision, CSU, BOB
Current work

BOB

Free signal processing and machine learning toolbox Bob:
http://www.idiap.ch/software/bob

- Signal and image processing techniques (filtering, LBP, SIFT, optical flow etc.)
- Machine learning algorithms (PCA, LDA, MLP, SVM, JFA, GMM, clustering etc.)
- Many face databases (FRGC, LWF, XM2VTS, BANCA, MOBIO, MULTIPIE, ...) supported, documentation and examples

Source code of publications available as Bob's satellite package, for easy reproduction of results:
https://github.com/idiap/bob/wiki/Satellite-Packages
Future work

What if?

- upload an algorithm to an online platform
- setup an experiment remotely on any database
- collect metrics and plots

Dr. S. Marcel – Idiap 2012

On the importance of benchmarking in facial image analysis
Future work

BEAT

www.beat-euproject.org

- to build an independent platform for face recognition research, development and certification
- to make easy to compare results from distinct algorithms or/and parameterizations
Thank you for your attention

Sébastien Marcel

Idiap Research Institute
Martigny, Switzerland
marcel AT idiap DOT ch
www.idiap.ch/~marcel