An Associate-Predict Model for Face Recognition
FIPA Seminar WS 2011/2012

Mykola Volovyk, 19.01.2012
Outline

- Introduction
  - Motivation
  - Related works

- Basic ideas
  - Approach scheme
  - Identity data set
  - Face components features
  - Settings estimation

- Approach
  - Appearance-prediction
  - Likelihood-prediction
  - Switching mechanism

- Results
- Introduction
- Basic ideas
- Approach
- Results
Motivation

Different pose

Different illumination

Different expression

Simply different
Motivation: the same / different settings

- Jeff Hawkins‘ „On intelligence“ brain study
- Two types of face matching
- 1) Similar settings
  - *Direct* matching (just measure component distances and compare them)

```
Yes
```

```
No
```

- 2) Different settings
  - Distances not informative → direct matching inefficient
  - Life full of faces → our memory == big face image gallery
  - Use memory as bridge between two images
  - *Associate-predict* matching
Motivation: different settings

Different settings

1) associate in memory database similar faces
2) predict from memory similar faces under searched settings
3) direct matching

?  

Yes
Attribute and Simile Classifiers by Kumar et. al [ICCV 2009]

**Detected**
Omron detector

**Normalize**
Affine warp

**Extract low-level features**
Intensity, RGB, HSV, edge, gradient dir.

**Make verification**

\[
p_i = (|a_i - b_i|, (a_i \cdot b_i)) \quad g(\frac{1}{2}(a_i + b_i))
\]

\[
v(I_1, I_2) = D(p_1, \ldots, p_n)
\]

SVM with RBF-kernel

**Find visual traits**

- Male
- White
- Child
- Bald
- Hair
- Mustache
- Bald
- Beard
- Body
- Eyes
- Face
- Skin
- Introduction
- Basic ideas
- Approach
- Results
Memory

People’s memory == Machine’s gallery
Goal

- **Main goal of our approach:** *to deal with intra-personal variation*
- **Basic idea:**
  - By different settings
  
  ![Image A](image1.png) ![Image B](image2.png)

- Find in the gallery **suitable bridge** between two compared images
- Two steps
Association step

First step: Associate face B with the most alike group from memory
Prediction step

- Second step: Find the image with searched settings
- That will be our predict

Similar settings:
- frontal
- illumination
- neutral expression

Details about settings estimation – in further slides
Big picture

Memory

“Associate”  “Predict”

A  B  B’

Appearance-based matching  Prediction-based matching
Identity data set

- 200 ids (persons) from Multi-PIE (CMU Face Database)
- For each person: 7 poses, 4 illuminations, 1 expression

- 7 poses:
  - -60%
  - -40%
  - -20%
  - 0%
  - 20%
  - 40%
  - 60%

- 4 illuminations:
  - no flash
  - left flash
  - right flash
  - left-right flash
Feature extraction

Four landmarks automatically detected

Alignments for 12 components

- Forehead
- Brows
- Eyes
- Cheeks
- Nose
- Mouth

Component representation

- Left eye
- Nose
Descriptors

- **LBP**
  - extract intensity for each pixel and its neighboring
  - invariant to rotation and grayscale (intensity) changes

- **SIFT**
  - Differences of Gaussians (DoG) - invariant to rotation and image scale
  - 1) DoG \rightarrow scale-space extrema regions
  - 2) gradients \rightarrow keypoints description

- **LE**
  - extract local microstructures (e.g., edges, lines, spots, flat areas)
  - invariant to grayscale changes

- **Gabor**
  - robustness against varying brightness, varying contrast
  - certain amount of robustness against translation, distortion, rotation, and scaling
Setting estimation

- Measure distances between extracted feature vectors \((Input, Template)\)
- Take the settings of the nearest template

\[
\text{Template} = \text{Average-face across all left-oriented images in gallery}
\]
Introduction
Basic ideas
Approach
Results
Associate-Predict Model

- "Associate" the component

- Measure distances between extracted feature vectors (A, gallery images)
- Take the nearest id (person)
Appearance prediction

Prediction possibility a) Appearance-prediction

Settings:
- right-oriented, 
- right flash

Choose inside of associated id the image with settings of B:
- frontal
- left-right flash

Calculate distance
\[ d_A = |f_{A'} - f_B| \]

Settings:
- frontal,
- left-right-flash
**Appearance prediction**

\[ A \rightarrow A' \rightarrow B \quad \text{----------}\quad d_A = |f_{A'} - f_B| \]

\[ B \rightarrow B' \rightarrow A \quad \text{----------}\quad d_B = |f_{B'} - f_A| \]

**Final distance**

Simple average of both

\[ d_p = \frac{1}{2} (d_A + d_B) \]

With weights \( \alpha_A, \alpha_B \)

\[ d_p = \frac{1}{\alpha_A + \alpha_B} \star (\alpha_A \ast d_A + \alpha_B \ast d_B) \]

\[ \alpha_A = e^{\gamma|f_{A} - f_{A'}|} \quad \alpha_B = e^{\gamma|f_{B} - f_{B'}|} \]
Appearance prediction

Results

- Fusion of 12 predicted components ($A_1', A_2', \ldots, A_{12}'$) = appearance-prediction result

Component distances $\left(d_{p_1}, d_{p_2}, \ldots, d_{p_{12}}\right)$

Fused decision: the same / not-the-same
Likelihood prediction

Prediction possibility b) Likelihood-prediction
Likelihood prediction

- 20 ids = negative samples (20/200 = 10%)
- Select $K$ – number of „positive“ ids (nearest neighbors)

- By associate-step instead of 1 nearest neighbor, we select $K$ nearest neighbors ($K$ the most similar ids)

Positive sample set = $K \times (\# \text{ images per person}) + 1 \text{ input-image}$

⚠️ Or subset of this number
Likelihood prediction

- Rest: negative samples

- We separate positive/negative with LDA:
Likelihood prediction

- For each new sample B
  - LDA tells us: \( P(B \text{ belongs to the positive sample set}) = ? \)

\[ d_i = 0.8 \rightarrow \text{Likelihood distance high} \rightarrow \text{(Not-the-same)} \]
Likelihood prediction

- Build A-Classifier + feed new sample B → Likelihood distance $d_A$
- Build B-Classifier + feed new sample A → Likelihood distance $d_B$

- Average:
  \[
  d_p = \frac{1}{2} (d_A + d_B)
  \]

- With weights:
  \[
  d_p = \frac{1}{\alpha_A + \alpha_B} \alpha_A d_A + \alpha_B d_B
  \]

- $d_p < \text{Threshold} \rightarrow \text{the positive sample}$

Component distances $d_{p1}, d_{p2}, \ldots, d_{p12}$ → linear SVM → Fused decision: the same / not-the-same
Switching mechanism

- Pair A,B is **Comparable** if

  \[ |P_A - P_B| < 3 \quad \text{and} \quad |L_A - L_B| < 3 \]

- **Not comparable**
  - else

\[ |P_A - P_B| = 6 \quad \text{and} \quad |L_A - L_B| = 3 \]
Switching mechanism

- Final matching distance:

\[ d_{sw} = \begin{cases} 
  d_a, & \text{if comparable} \rightarrow \text{direct matching} \\
  d_p, & \text{if not comparable} \rightarrow \text{associate-predict model} 
\end{cases} \]

- Switching reduces risk of inaccurate association/prediction
- Introduction
- Basic ideas
- Approach
- Results
Experimental results

- **Training set**
  - **Multi-PIE:** 200 persons (from CMU, over all 337 persons, >750,000 images)

- **Test sets**
  - **Multi-PIE:** 49 persons mutually exclusive to training set
    - 10 folds cross-validation
    - Each fold has 300 intra-personal pairs, 300 extra-personal pairs
  - **LFW** (Labeled Faces in the Wild, over all 5749 people, >13,000 images)
    - **Restricted protocol** (fixed number of intra-personal and extra-personal pairs provided for training)
      - 10 folds cross-validation
      - Each fold has 300 intra-personal pairs, 300 extra-personal pairs
    - **Unrestricted protocol** (random number of training pairs can be generated based the given faces‘ labels)
Experimental results

**Holistic vs. Component** on Multi-PIE

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>LBP</th>
<th>SIFT</th>
<th>Gabor</th>
<th>LE</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct</td>
<td>80.55%</td>
<td>78.30%</td>
<td>81.00%</td>
<td>84.20%</td>
</tr>
<tr>
<td>appearance (H)</td>
<td>83.85%</td>
<td>82.40%</td>
<td>83.95%</td>
<td>87.55%</td>
</tr>
<tr>
<td>appearance (C)</td>
<td>86.75%</td>
<td>86.65%</td>
<td>86.80%</td>
<td>89.75%</td>
</tr>
<tr>
<td>likelihood (H)</td>
<td>85.65%</td>
<td>83.45%</td>
<td>87.05%</td>
<td>87.40%</td>
</tr>
<tr>
<td>likelihood (C)</td>
<td>89.05%</td>
<td>87.60%</td>
<td>89.85%</td>
<td>92.25%</td>
</tr>
</tbody>
</table>

- by appearance: component ~3% better
- by likelihood: component ~4% better

Direct matching always worse than other
Experimental results

- Effect of positive sample number for likelihood-prediction on Multi-PIE benchmark (LBP feature)

- 1 sample vs. 71 samples → improvement of ~10%
- Each id has 28 different images
- For K associated ids → max. 28*K +1 images
- For K=3 → from 59 to 78 positive images
Experimental results

- Improvement of Switching

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>LE on MPIE</th>
<th>LE on LFW</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct</td>
<td>84.20%</td>
<td>82.33%</td>
</tr>
<tr>
<td>appearance (no switching)</td>
<td>86.95%</td>
<td>82.95%</td>
</tr>
<tr>
<td>appearance (switching)</td>
<td>89.30%</td>
<td>88.16%</td>
</tr>
<tr>
<td>likelihood (no switching)</td>
<td>89.75%</td>
<td>84.30%</td>
</tr>
<tr>
<td>likelihood (switching)</td>
<td>92.25%</td>
<td>89.25%</td>
</tr>
</tbody>
</table>

- Improvement of ~5% on Multi-PIE dataset
- Improvement of ~2.5% on LFW dataset
Experimental results

Result on Multi-Pie benchmark

- Direct LE descriptor: 84.2%
- Fusion (the best experimental model): 94%

It means: 60% of errors were eliminated
Experimental results

- Result on LFW benchmark
  - Again clear improvement

Likelihood a little bit better than appearance

Fusion = appearance & likelihood fused by linear SVM
Experimental results (LFW benchmark)

- 90.57% the best experimental model
Final remarks

- Advantages of Associate-Predict model
  - Using universal identities as bridge between two images
  - Effective use of gallery with flexible switch model

- Achievements
  - Good handling of intra-personal variation (pose, illumination)
  - Best result under restricted protocol on LFW

- Improvement ideas
  - More prior knowledge → better results
The End

Thanks for your attention!
Questions?
References


Learning-based descriptor (LE)

“learning-based descriptor” pipeline

Preprocessed image → Sampling and normalization → Normalized low-level feature vectors → Learning-based encoding → Code image → Concatenated patch histogram → LE descriptor

PCA and normalization