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

Mykola Volovyk, 19.01.2012

INSTITUTE FOR ANTHROPOMATICS, FACIAL IMAGE PROCESSING AND ANALYSIS YIG

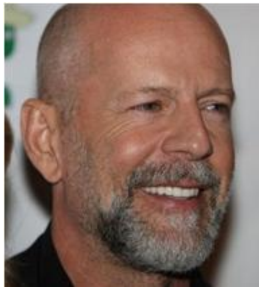
Outline

- 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

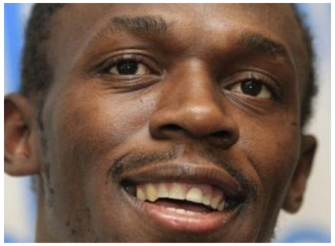
Motivation

Different pose

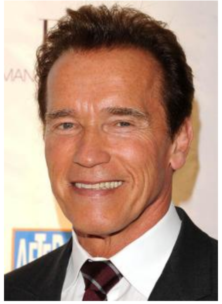



Different illumination


Different expression

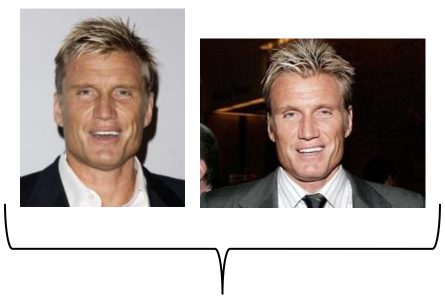



Simply different




Motivation: the same / different settings

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



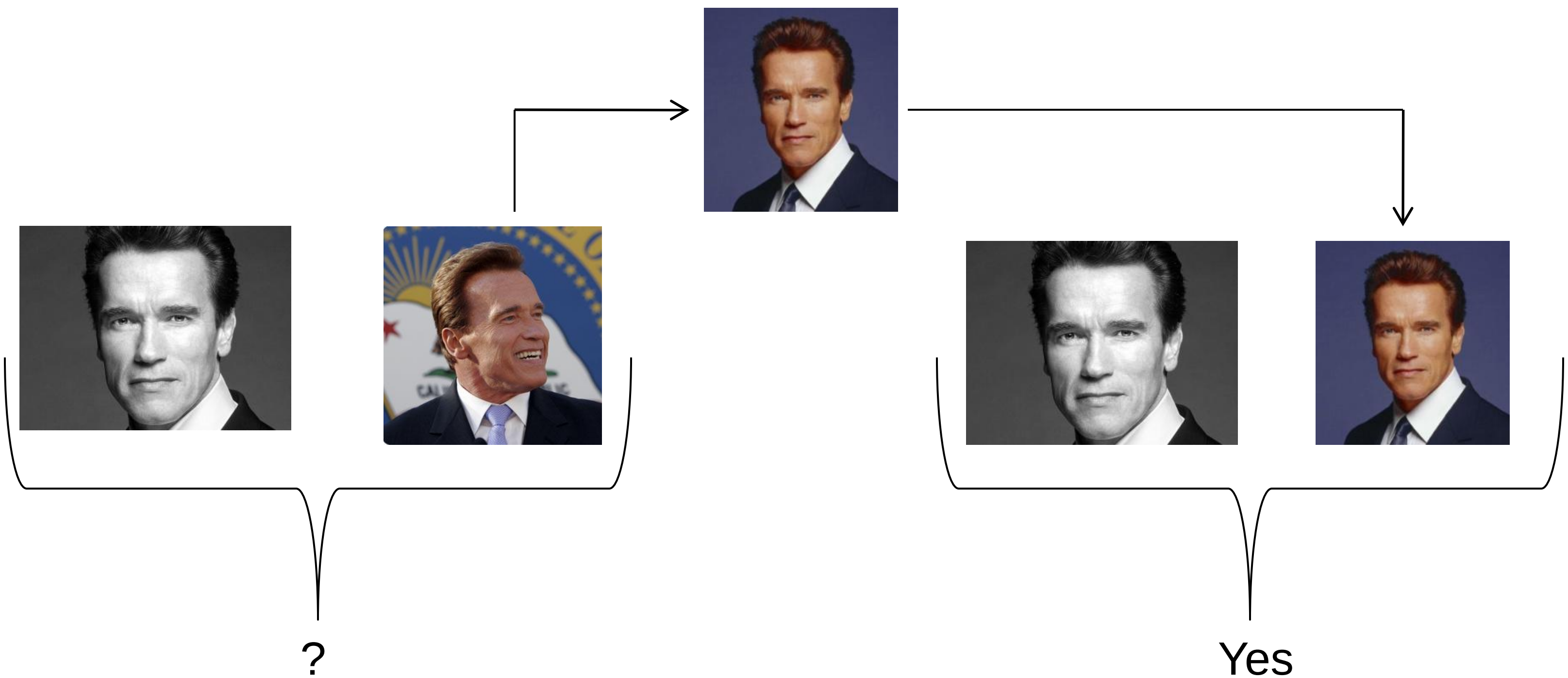
Yes

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

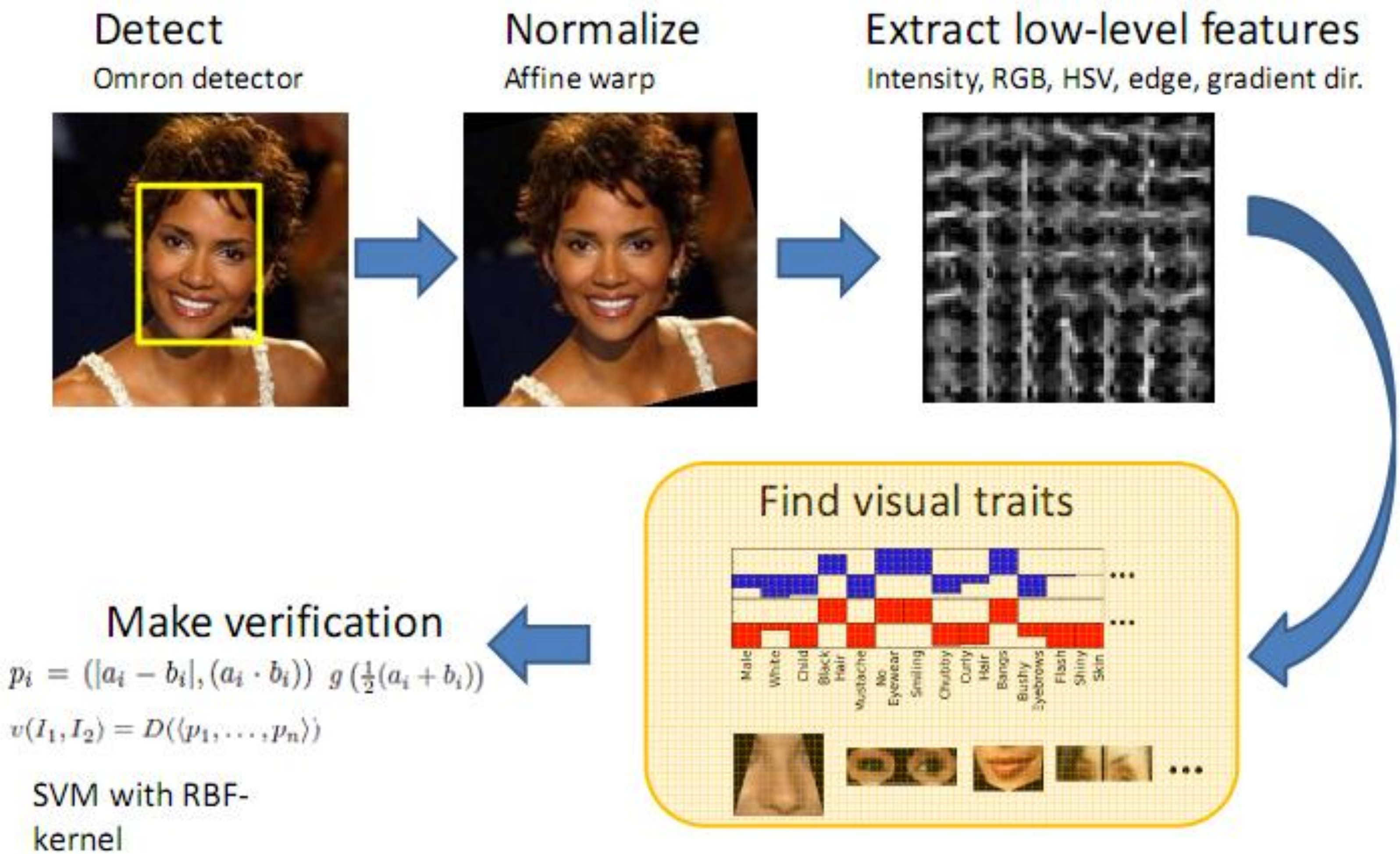
Motivation: different settings

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



Related works

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



- Introduction
- Basic ideas
- Approach
- Results

Memory

- People's memory == Machine's gallery

Goal

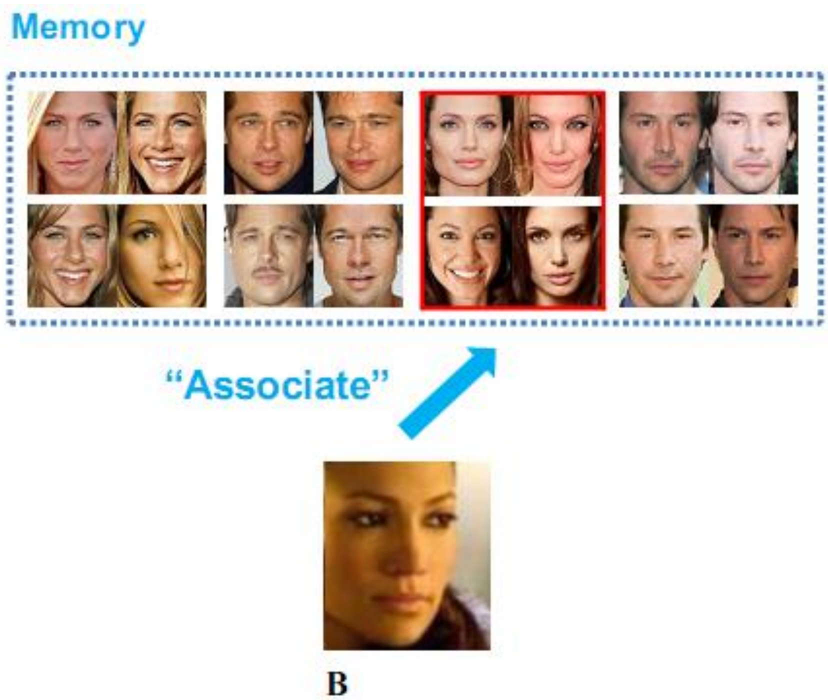
- Main goal of our approach: **to deal with intra-personal variation**
- Basic idea:
  - By different settings

A B

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

Association step

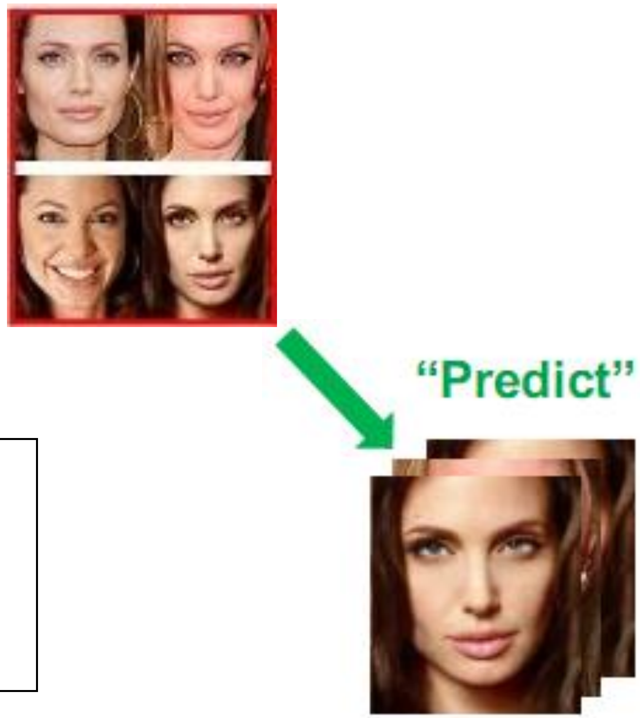
- First step: Associate face B with the most alike group from memory



Prediction step

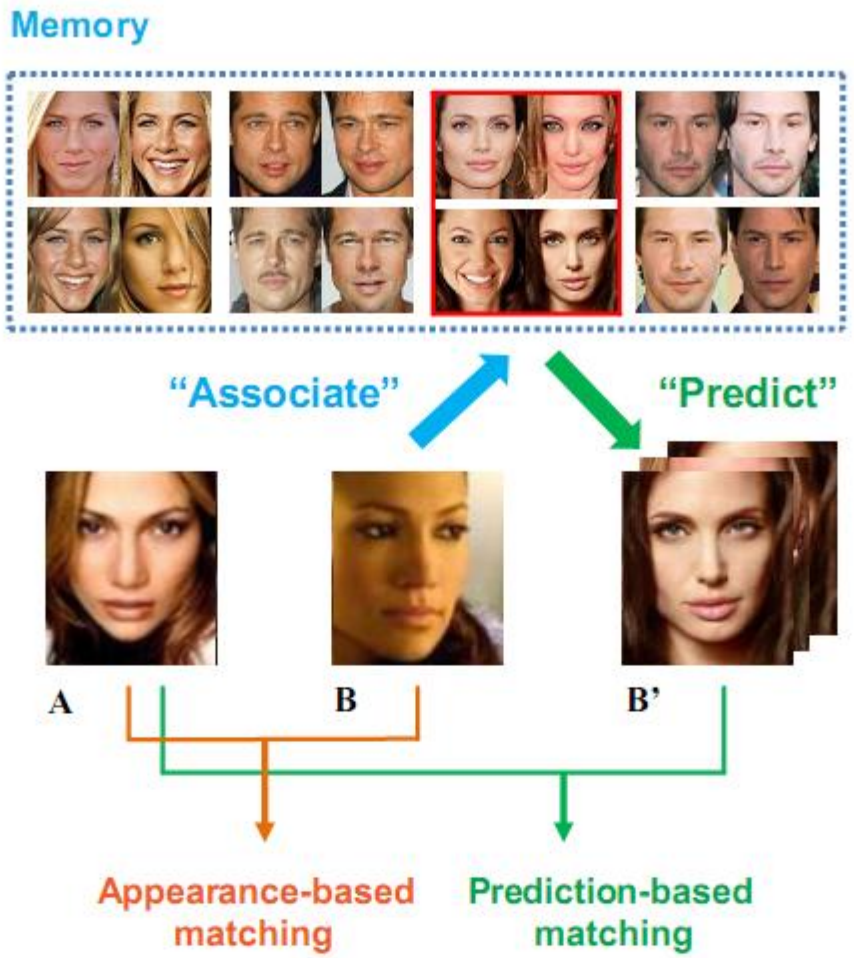
- Second step: Find the image with searched settings
- That will be our predict

Similar settings:
-frontal
-illumination
-neutral expression

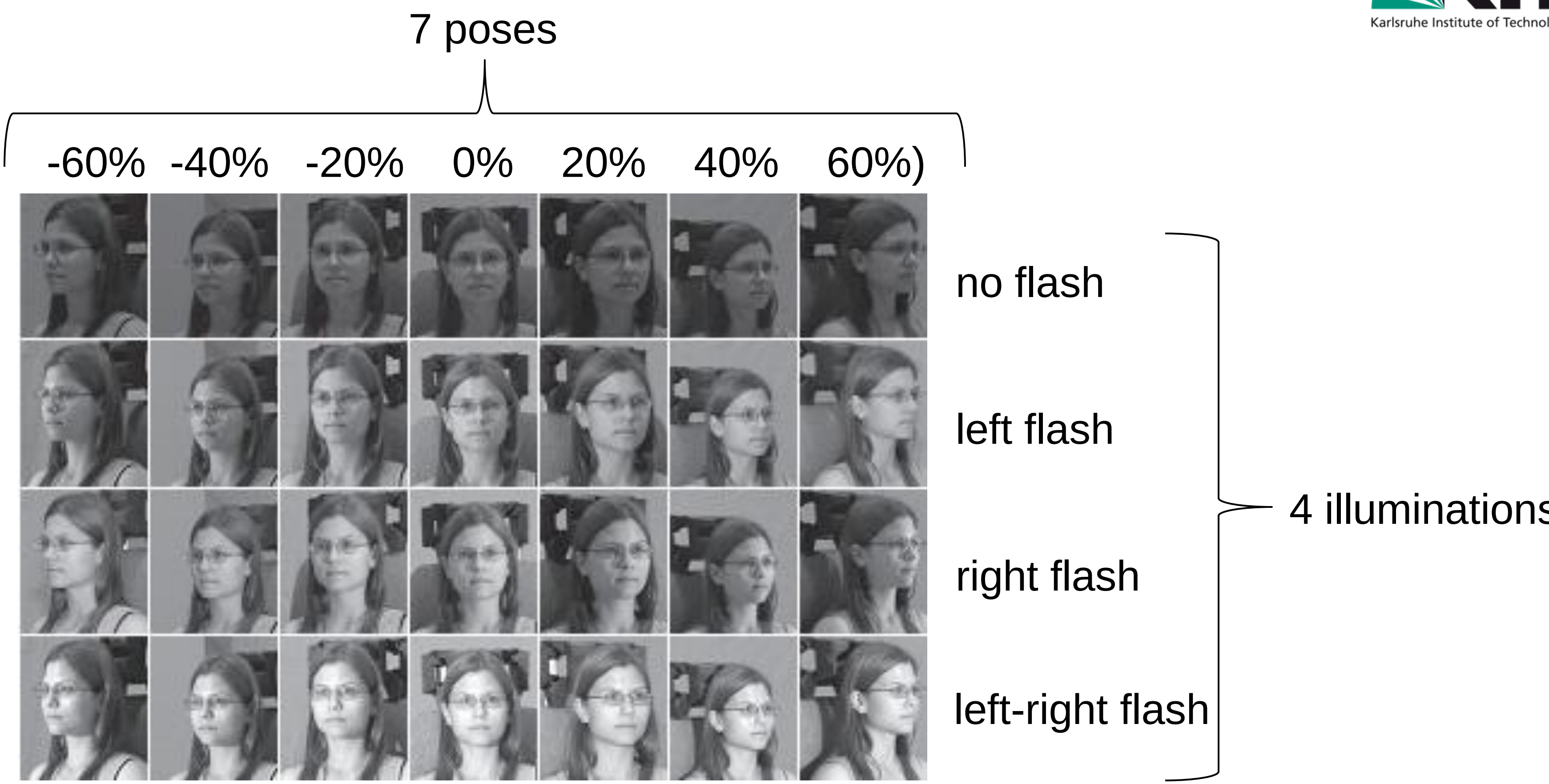


Details about settings estimation – in further slides

Big picture


Memory
"Associate"
"Predict"
A
B
B'
Appearance-based matching
Prediction-based matching

Identity data set

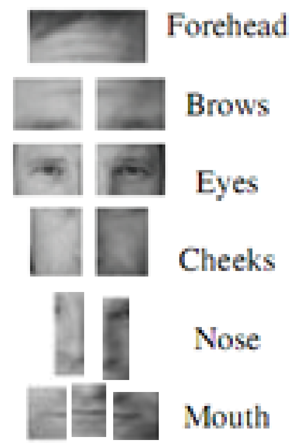


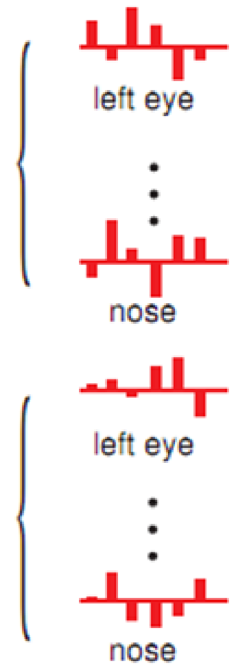
- 200 ids (persons) from Multi-PIE (CMU Face Database)
- For each person: 7 poses, 4 illuminations, 1 expression

Feature extraction

Four landmarks automatically detected

Alignments for 12 components





Component representaion

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 → scale-space extrema regions
  - 2) gradients → 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

Pose Templates

Illumination Templates

Input face

.
.
.

- Measure distances between extracted feature vectors *(Input, Template)*

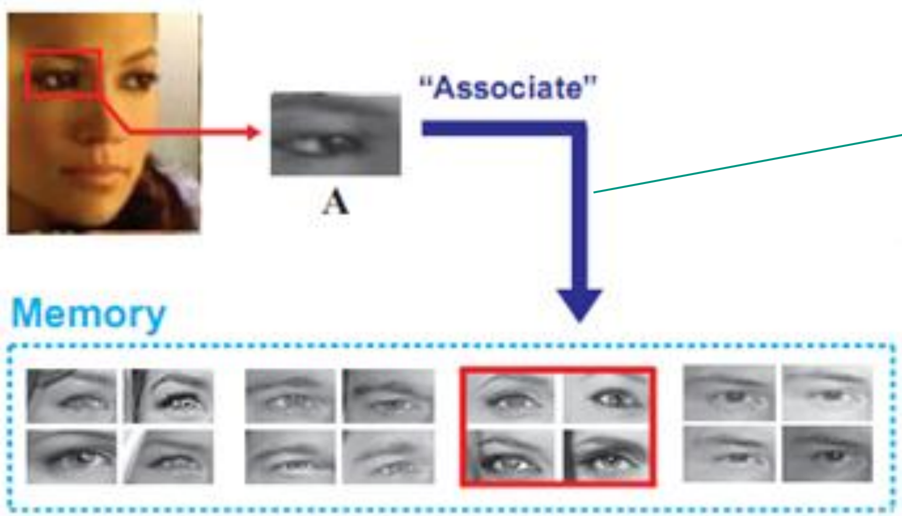
-Take the settings of the nearest template

Template = 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)

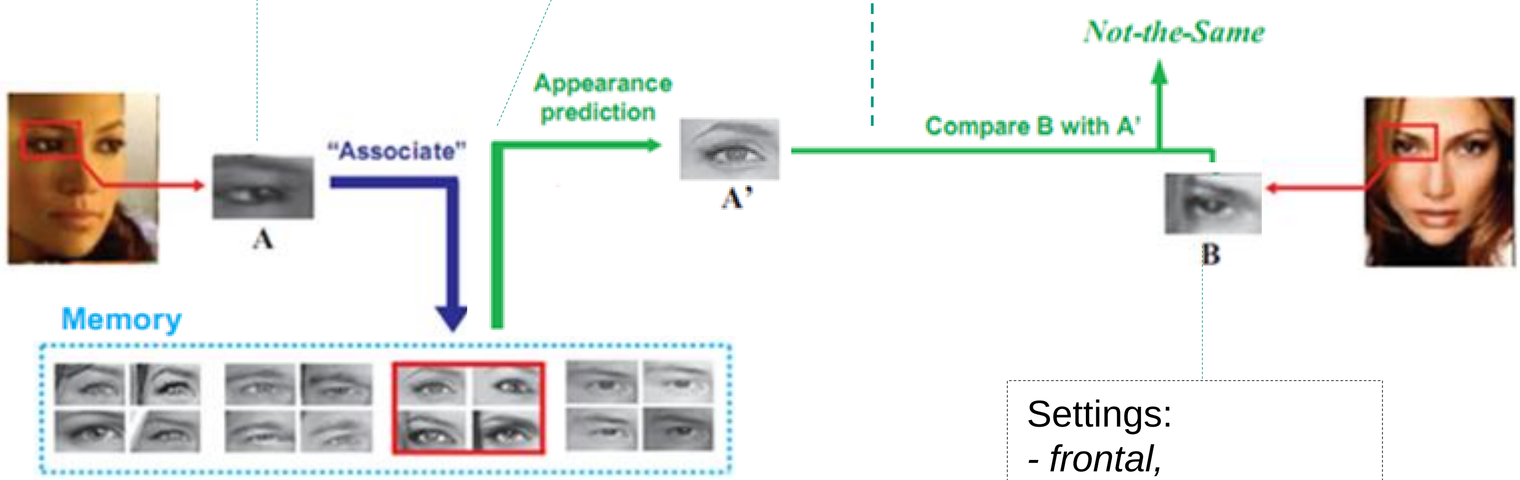
Appearance prediction

Prediction possibility **a)**
Appearance-prediction

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

Calculate distance
$d_A = |f_{A'} - f_B|$

Settings:
- *right-oriented,*
- *right flash*



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

Appearance prediction

A → A' → B ----------> $d_A = |f_{A'} - f_B|$

B → B' → A ----------> $d_B = |f_{B'} - f_A|$

Final distance

Simple average of both

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

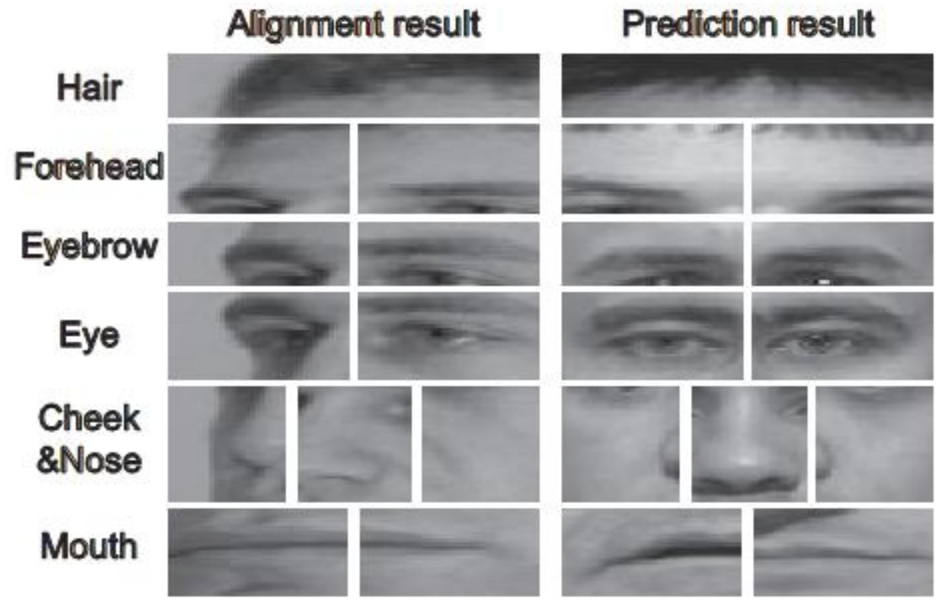
With weights α_A, α_B

$$d_p = \frac{1}{\alpha_A + \alpha_B} * (\alpha_A * d_A + \alpha_B * d_B)$$

$$\alpha_A = e^{\frac{1}{\gamma |f_A - f_{A'}|}} \qquad \alpha_B = e^{\frac{1}{\gamma |f_B - f_{B'}|}}$$

Appearance prediction

- Results
  - Fusion of 12 **predicted components** (A_1', A_2',..., A_{12}',) = appearance-prediction result

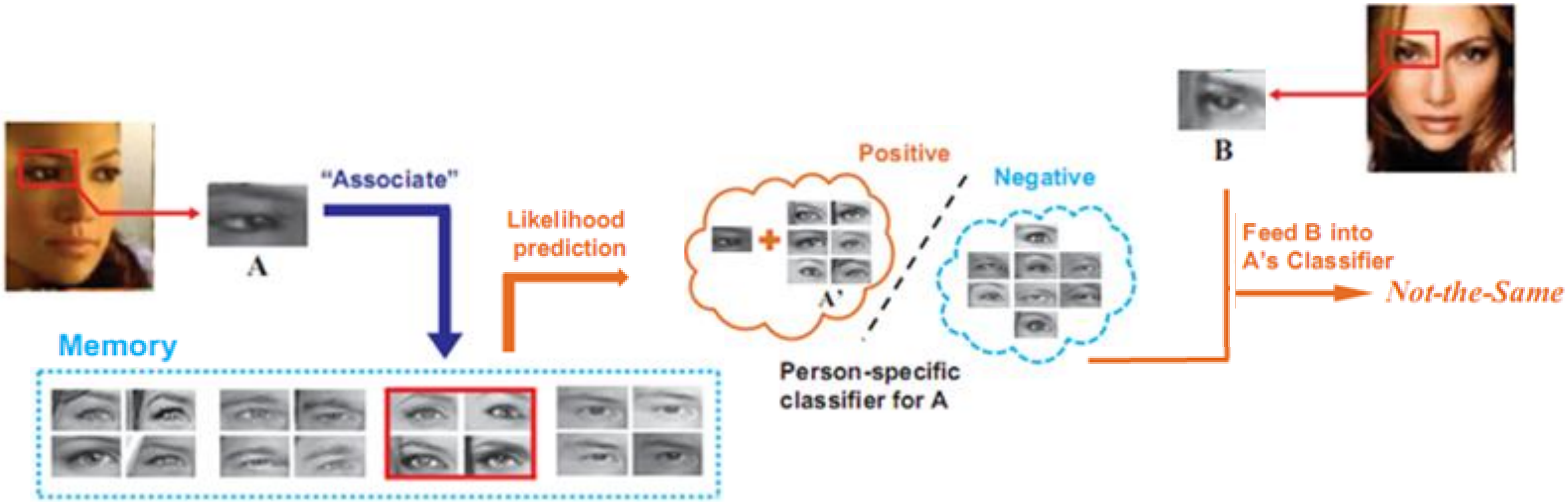


component distances $\xrightarrow{\text{linear SVM}}$ Fused decision: the same / not-the-same

$$(d_{p_1}, d_{p_2}, ..., d_{p_{12}})$$

Likelihood prediction

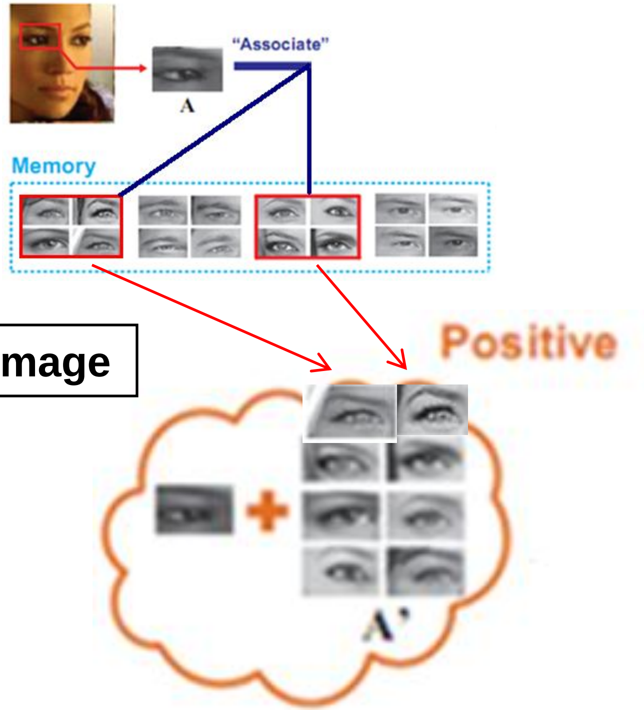
Prediction possibility **b)**
Likelihood-prediction



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)

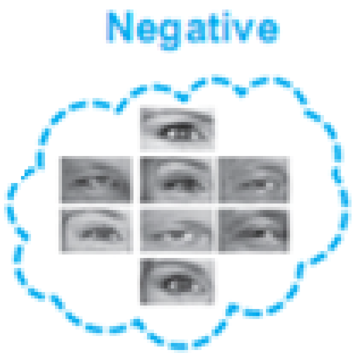


Positive sample set = **K * (# images per person) +1 input-image**

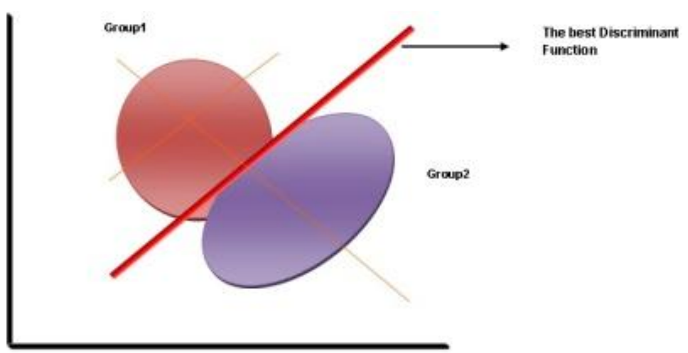
Or subset of this number

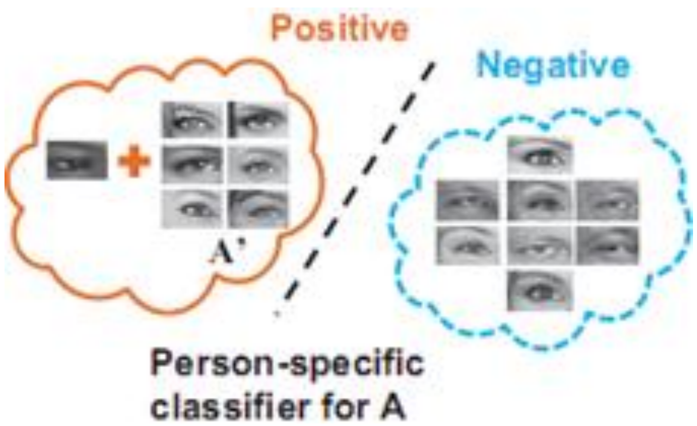
Likelihood prediction

- Rest: negative samples



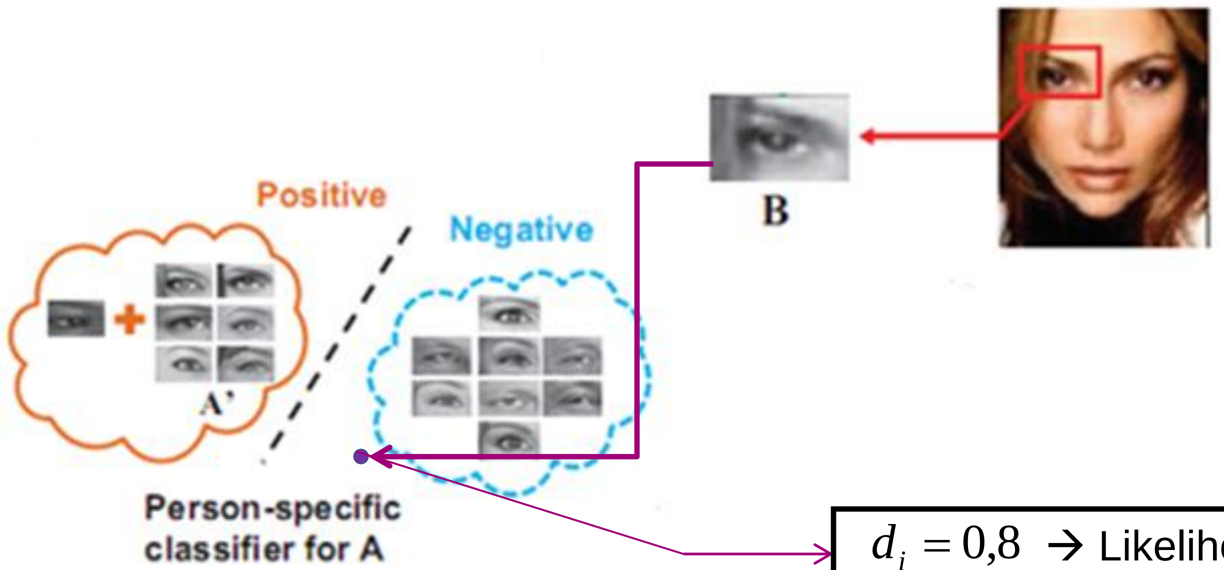
- We separate positive/negative with LDA:





Likelihood prediction

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



$d_i = 0{,}8$ → Likelihood distance high
→**(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: $$\boxed{d_p = \frac{1}{\alpha_A + \alpha_B} * (\alpha_A * d_A + \alpha_B * d_B)}$$
- d_p < Threshold → the positive sample

component distances $(d_{p_1}, d_{p_2}, \ldots, d_{p_{12}})$ $\xrightarrow{\text{linear SVM}}$ Fused decision: the same / not-the-same

Switching mechanism

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

Pose

Illumination

$|P_A - P_B| < 3$ <u>and</u> $|L_A - L_B| < 3$

- **Not comparable**
  - else

and

and

$|P_A - P_B| = 6$

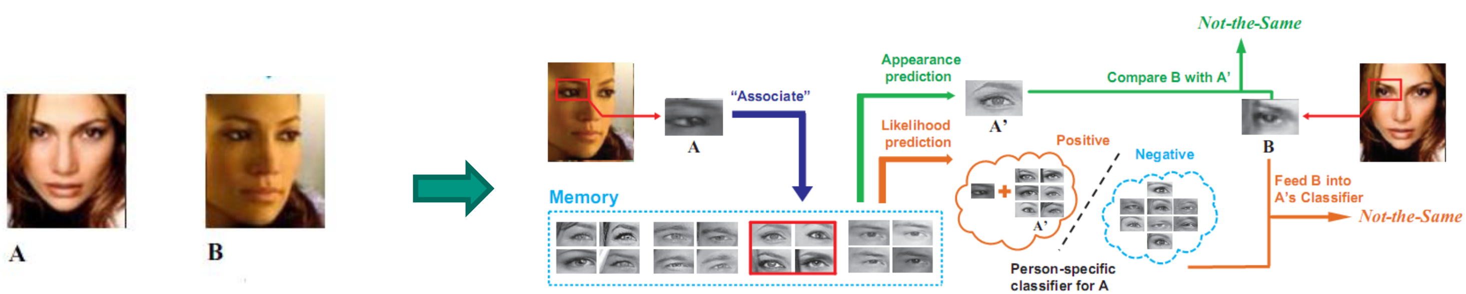
$|L_A - L_B| = 3$

Switching mechanism

- Final matching distance:

Yes

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



- Switching <u>reduces risk</u> 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

Direct matching always worse than other

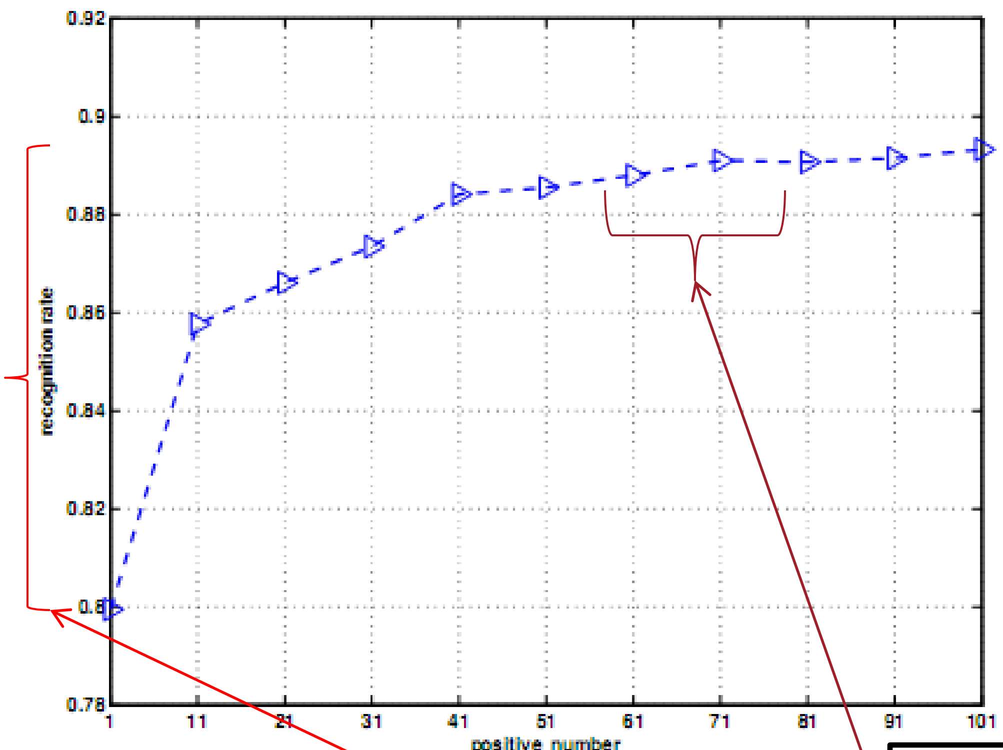
Descriptor	LBP	SIFT	Gabor	LE
direct	80.55%	78.30%	81.00%	84.20%
appearance (H)	83.85%	82.40%	83.95%	87.55%
appearance (C)	86.75%	86.65%	86.80%	89.75%
likelihood (H)	85.65%	83.45%	87.05%	87.40%
likelihood (C)	89.05%	87.60%	89.85%	92.25%

• by appearance: component ~3% better

• by likelihhood: component ~4% better

Experimental results

- **Effect of positive sample number for likelyhood-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**

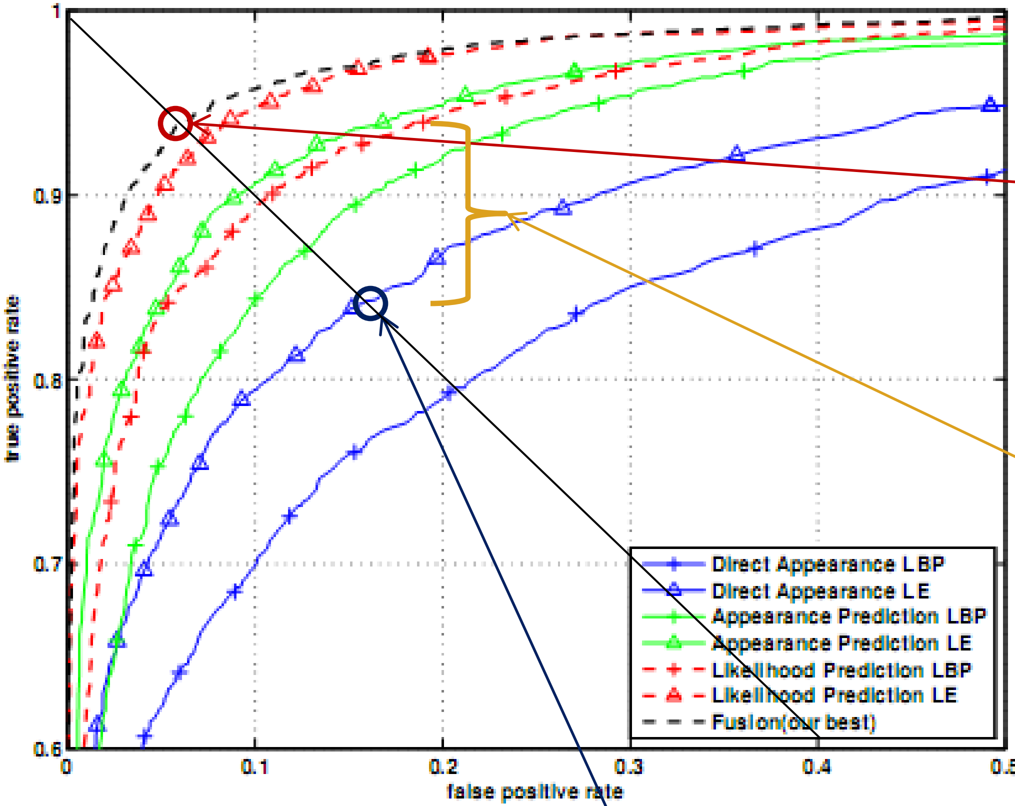
Descriptor	LE on MPIE	LE on LFW
direct	84.20%	82.33%
appearance (no switching)	86.95%	82.95%
appearance (switching)	89.30%	88.16%
likelihood (no switching)	89.75%	84.30%
likelihood (switching)	92.25%	89.25%

on Multi-PIE dataset
– improvement of ~2,5%

on LFW dataset
– improvement of ~5%

Experimental results

- **Result on Multi-Pie benchmark**



Fusion (the best experimental model): 94%

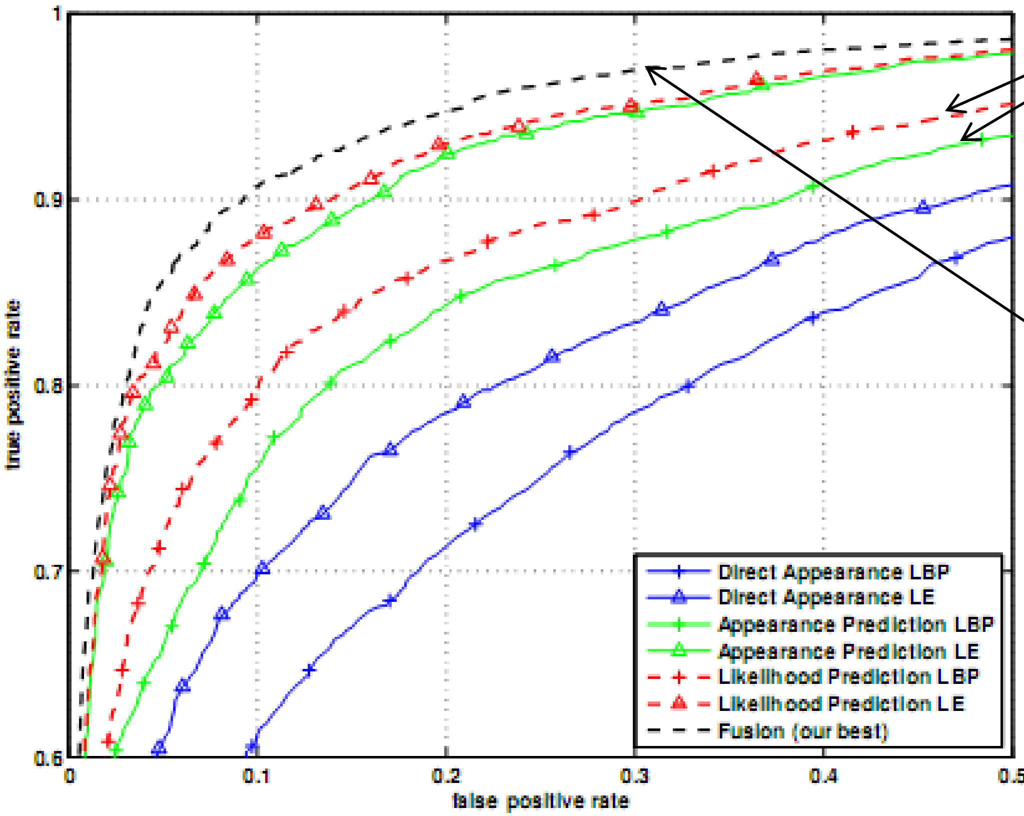
It means: 60% of errors were eliminated

Direct LE descriptor: 84.2%

Experimental results

Result on LFW benchmark

- Again clear improvement

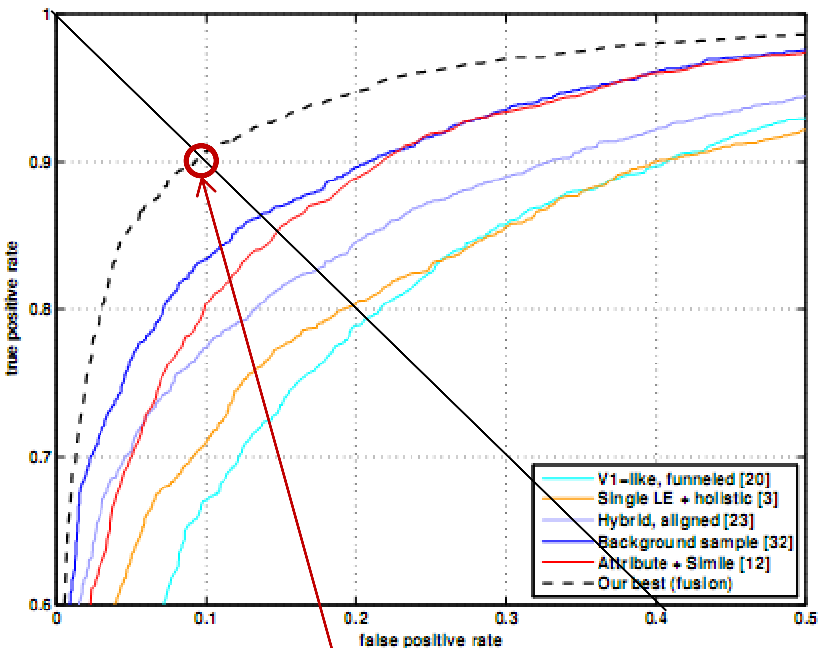


Likelihood a little bit better than appearance

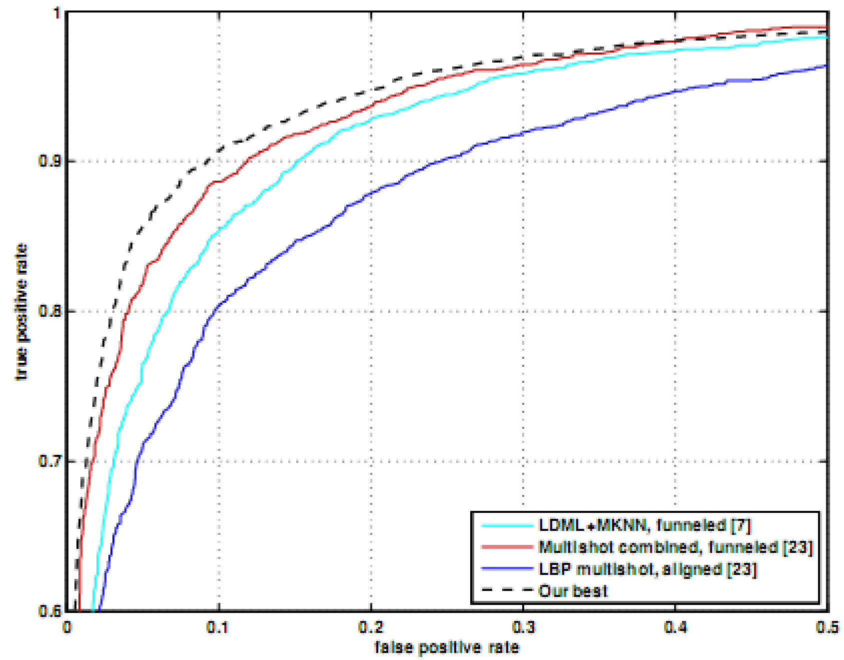
Fusion = appearance & likelihoood fused by linear SVM

Experimental results (LFW benchmark)

Restricted protocol



Unrestricted protocol



- 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

- Q. Yin, X. Tang, and J. Sun. An associate-predict model for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- Z. Cao, Q. Yin, J. Sun, and X. Tang. Face recognition with Learning-based Descriptor. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. *International Conference on Computer Vision (ICCV)*, 2009.

Learning-based descriptor (LE)

“learning-based descriptor” pipeline