Coupled Discriminative Feature Learning for Heterogeneous Face Recognition

Yi Jin[1]   Jiwen Lu[2]   Qiuqi Ruan[1]

[1] Beijing Jiaotong University, Beijing, China

[2]Advanced Digital Sciences Center, Singapore

(presented by Weihong Deng)

Outline

Motivation : Heterogeneous Face Recognition

Sketch

What is Heterogeneous Image ? Different Modalities

High and Low resolution

VIS NIR

2D 3D

Motivation : HFR

- Three steps of HFR

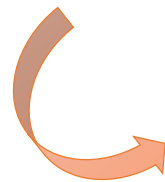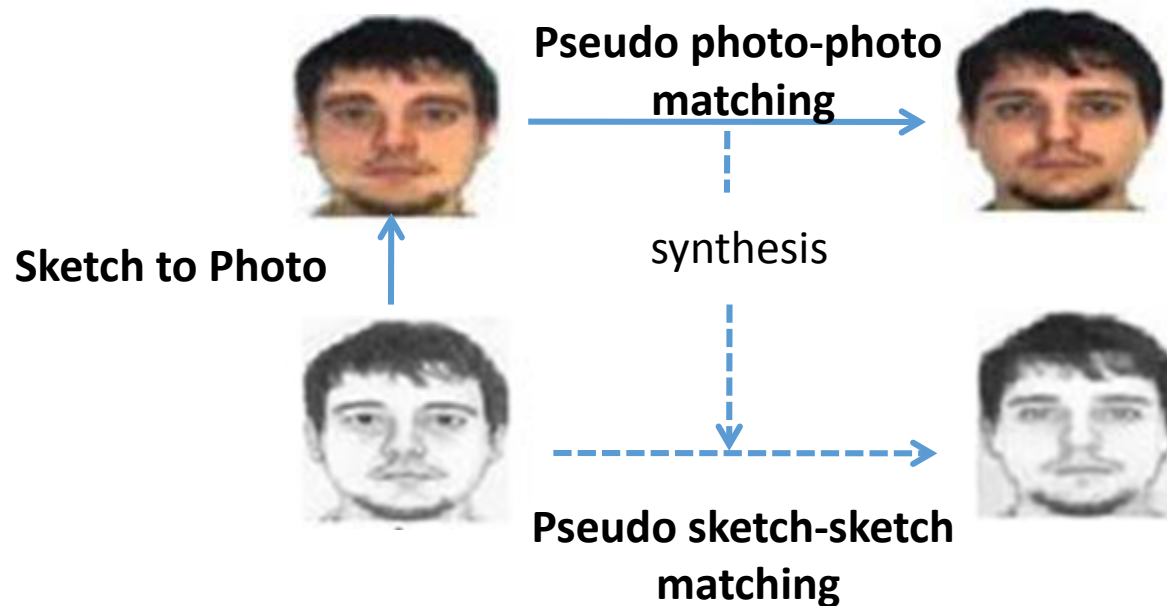| Data preprocessing	→	Cross-modality feature extraction	→	Cross-modality matching
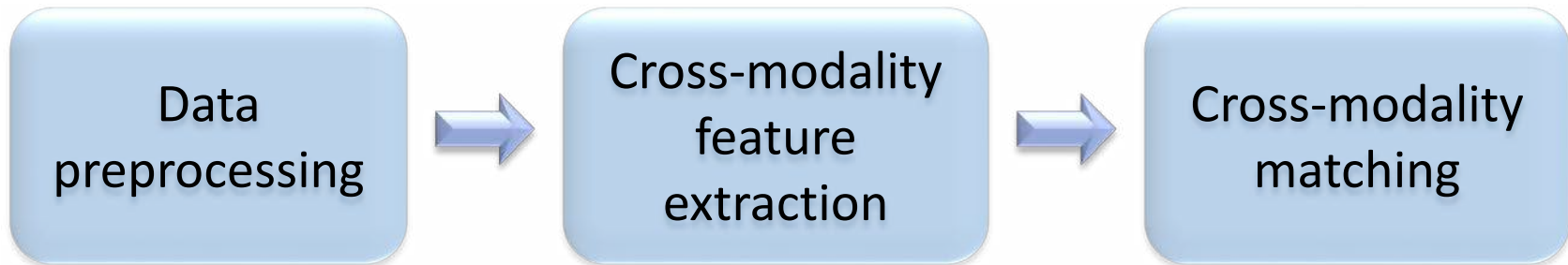
- What's the difficulty ?

Large appearance differences between different image modalities.

**Aim at reducing the cross-modality difference**

Related work: HFR

- Homogenous image synthesis methods

Pseudo photo-photo matching

Sketch to Photo

synthesis

Pseudo sketch-sketch matching

Face photo-sketch synthesis and recognition. PAMI 2009, Wang et al.

Face sketch synthesis algorithm based on E-HMM and selective ensemble. TCSVT 2008, Gao et. al.

Semi-coupled dictionary learning with applications to image super-resolution and photosketch synthesis . CVPR 2012, Wang et. al.

Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition .ICCV 2013, Huang et. al.

Related work: HFR

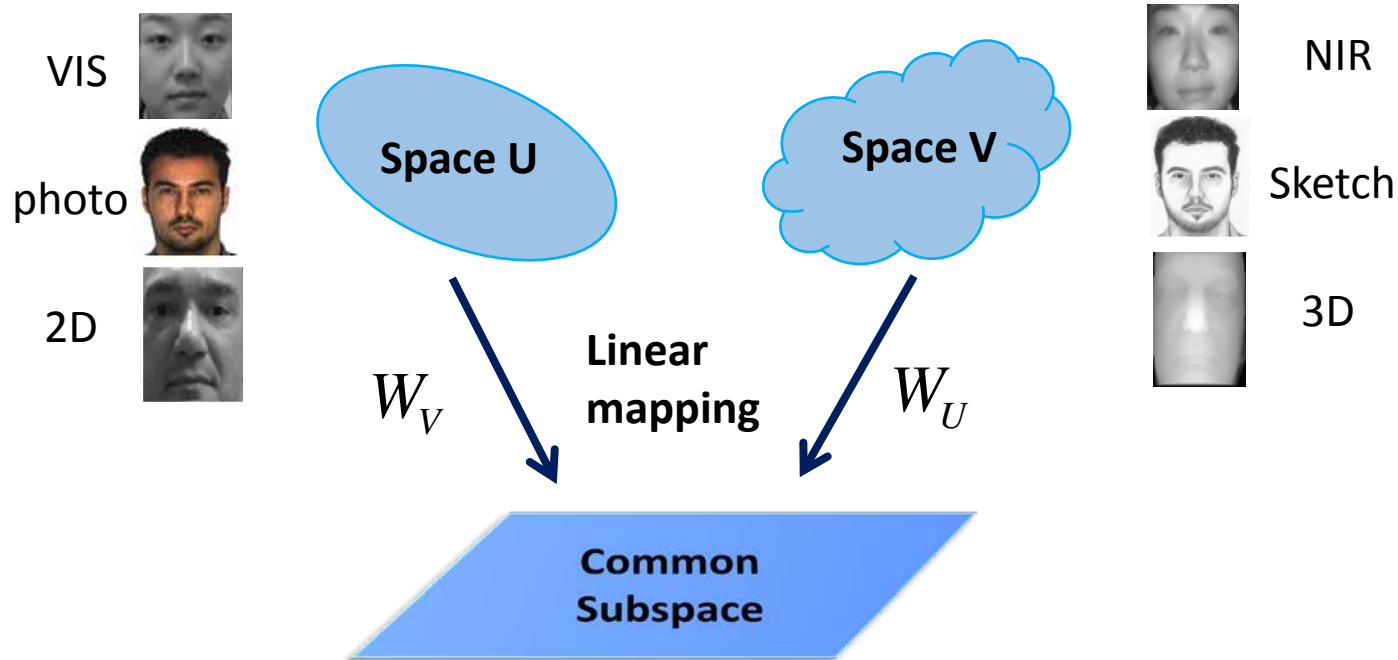- Invariant Feature Extraction

  - Hand-crafted feature descriptors

    - Such as, LBP and its variations, SIFT, HOG, Gabor filters
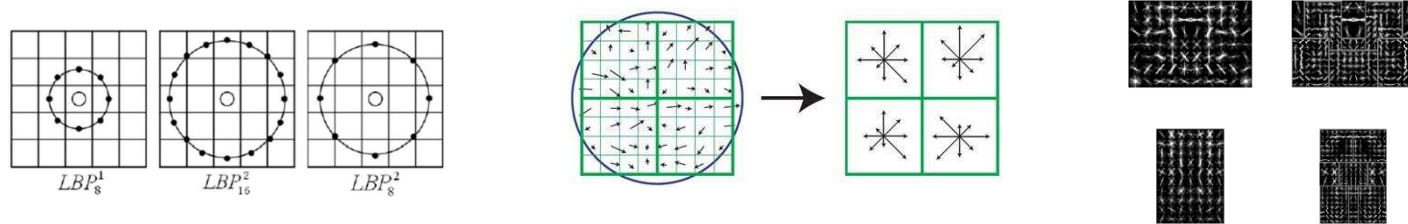
Heterogeneous face recognition using kernel prototype similarities. PAMI 2013, Klare, et. al.

Heterogeneous face recognition from local structures of normalized appearance. ICB 2009, Liao et. al.

Matching NIR face to VIS face using transduction, TIFS 2014, Zhu et. al.

Coupled information-theoretic encoding for face photo-sketch recognition. CVPR 2011, Zhang et al.

Learning modality-invariant features for heterogeneous face recognition. ICPR 2012, Huang et al.

- Common Subspace Learning

VIS

photo

2D

Space U

Space V

NIR

Sketch

3D

W_V

Linear mapping

W_U

Common Subspace

- Inter-modality face recognition . ECCV 2006, Lin et. al.
- Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch . CVPR 2011, Sharma et. al.
- Coupled discriminant analysis for heterogeneous face recognition. TIFS 2012, Lei et. al.
- Coupled spectral regression for matching heterogeneous faces. CVPR 2009, Lei et. al.

Our works

- Why feature learning?

  - Learn Data-driven descriptor, prior knowledge is not required.

  - The cross-modality differences can be directly reduced at the image-pixel level.

- What are the benefits?

  - Can be easily involved in different HFR applications.

  - Emphasizing both the discriminant and correlation information.

  to Improving recognition performance.

Related work: feature learning

- ## feature learning based methods

  - ### Such as, coupled common subspace learning,

    Learning coupled feature spaces for cross-modal matching. ICCV 2011, Wang et al.

    Generalized multiview analysis: A discriminative latent space. CVPR 2012, Sharma et al.

    - ### dictionary learning

      Semi-coupled dictionary learning with applications to image super-resolution and photosketch synthesis. CVPR 2012, Wang et al.

      Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. CVPR 2012, Huang et al.

    - ### deep learning

      Deep canonical correlation analysis. ICML 2013, Andrew et. al.

      Multimodal deep learning. ICML 2011, Ngiam et. Al.
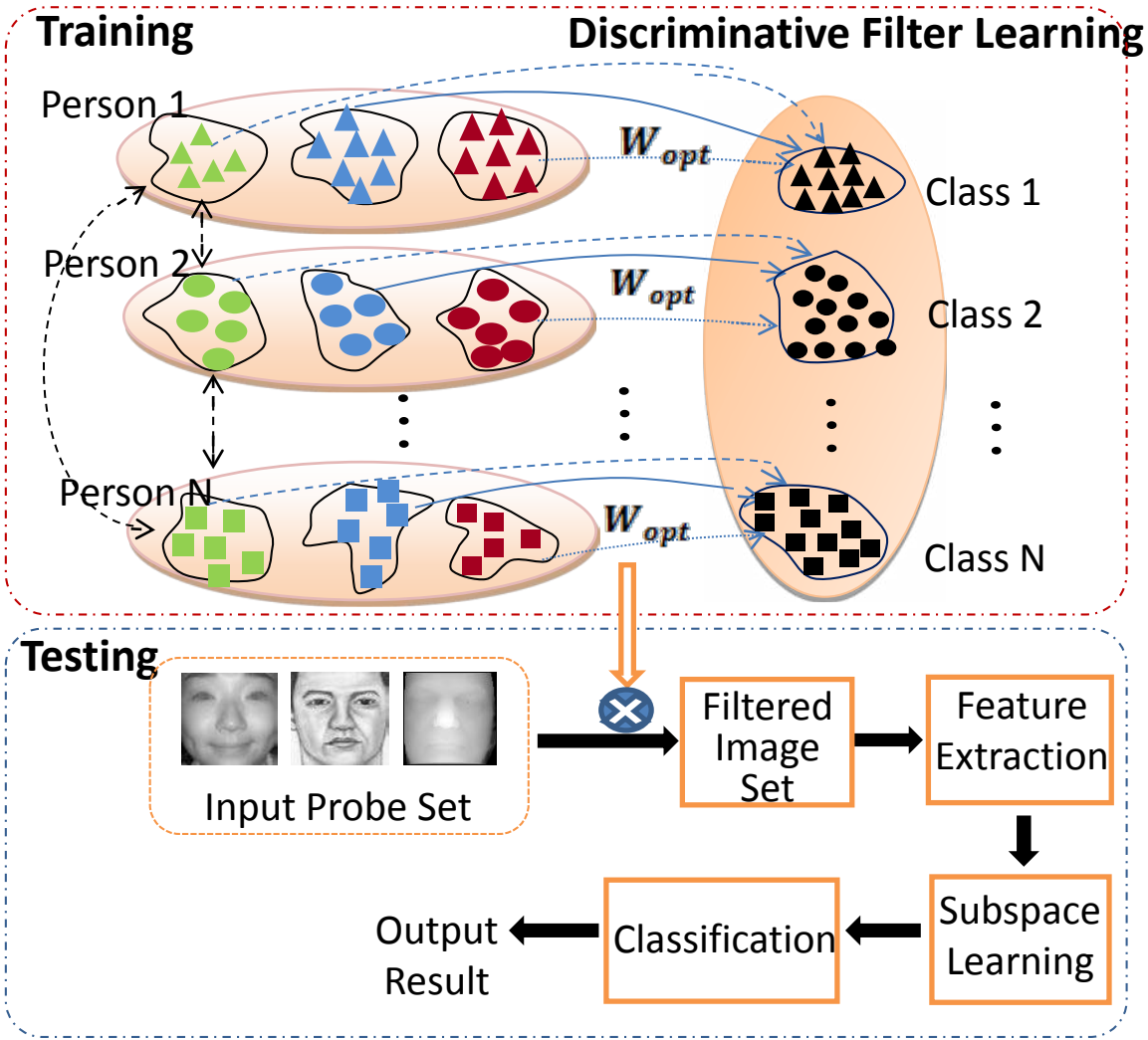
      - ### hashing

        Predictable dual-view hashing. ICML 2013, Rastegari et. Al.

      - ### cross-modal metric learning

        CMML: A new metric learning approach for cross modal matching. ACCV 2012, Mignon et. Al.

Motivation:

Training
Discriminative Filter Learning

Person 1

Person 2

Person N

W_{opt}

W_{opt}

W_{opt}

Class 1

Class 2

Class N

The basic idea of feature learning to seek optimal image filters to obtain discriminative face representation. The red, green and blue points represent face samples captured from different modalities, such as the NIR, VIS, Sketch or 3D images. The samples with the same shape represent face images from the same person.

Testing

Input Probe Set

Filtered Image Set → Feature Extraction

Subspace Learning

Classification

Output Result

Methodology: CDFL

- ## Formulation

Given a $p \times q$ image I and $\phi(I)$ is the filtered image of I. Let α be the discriminative image filter vector, the value of the filtered image at position k is

$$\varphi(I)^k = \alpha^T I^k$$

Inspired by LBP, the pixel difference vector (PDV) [39,40] is defined as:

$$d(I)_{ij}^k = [((\dot{I})_{ij}^{k_1} - (I)_{ij}^k), ((I)_{ij}^{k_2} - (I)_{ij}^k), \ldots, ((I)_{ij}^{k_L} - (I)_{ij}^k)]^T.$$

The goal of CDFL is to learn an optimal filter α, which makes the feature of filtered images more discriminative.
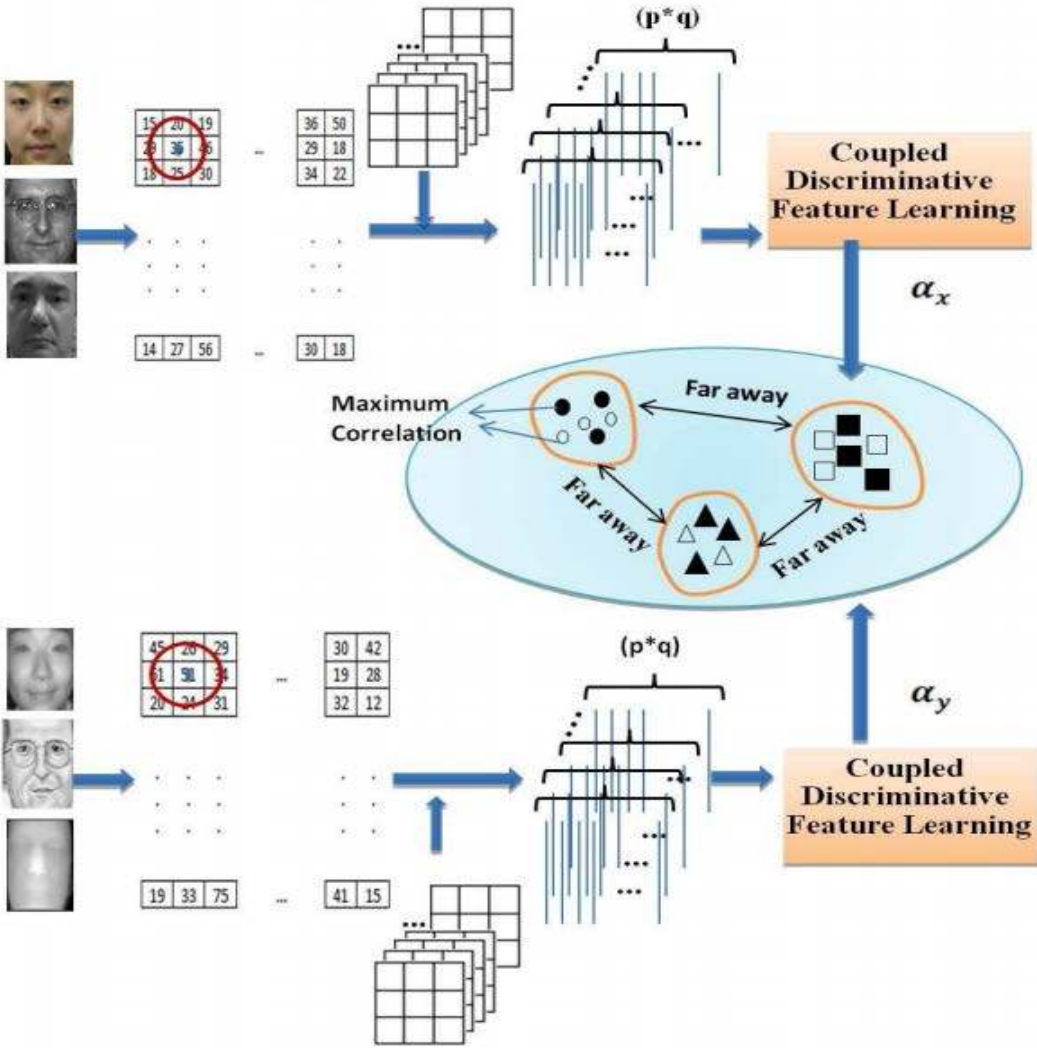
$$d\varphi(I)_{ij}^k = \alpha^T d(I)_{ij}^k$$

Discriminant image filter learning for face recognition with local binary pattern like representation. CVPR, 2012, Lei et. al.

Learning discriminant face descriptor. TPAMI 2014, Lei et. al.

Methodology: CDFL

- Our principle of feature learning in CDFL:

The principle of feature learning in CDFL. The core idea of CDFL is to build an optimal filter eigenvector α to reduce the modality gap at the image pixel level. The input face image is first divided into many patches according to a predefined neighborhood. Then, the pixel difference vector (PDV) is extracted at each pixel as the center, therefore, we get a $L \times (p \times q)$ pixel difference matrix. Finally, the filtered image is obtained by the projecting the pixel difference vector (PDV) at each position k into the CDFL subspace.

- Our goal:

  The objective function of CDFL is defined as the following:

  $$\max \quad \alpha^T [\theta(S_b - S_w) + (1 - \theta)\Phi] \alpha$$

  $$s.t. \quad \alpha^T \Phi_1 \alpha = 1 \ \& \ \alpha^T \Phi_2 \alpha = 1$$

  $$\alpha = [\alpha_x; \alpha_y].$$

$$S_w = \begin{bmatrix} S_w^{xx} & S_w^{xy} \\ S_w^{yx} & S_w^{yy} \end{bmatrix}, \quad S_b = \begin{bmatrix} S_b^{xx} & S_b^{xy} \\ S_b^{yx} & S_b^{yy} \end{bmatrix}$$

$$\Phi = \begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} C_{xx} & 0 \\ 0 & 0 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0 & 0 \\ 0 & C_{yy} \end{bmatrix}$$

Methodology: Algorithm

Algorithm 1 The Coupled Discriminative Feature Learning (CDFL) Algorithm

Input:
$\mathbf{I} = \{I_1, I_2, \cdots, I_{ij}, \cdots, I_N\}$: a set of $p \times q$ image samples;
i : the ith sample;
j : the jth class;
L : the number of choosen neighbours.

Output:
$\alpha = [\alpha_x; \alpha_y]$: the learned optimal image filters;
function CDFL($\mathbf{I}, L$)

    Step 1. Compute the Pixel Difference Vectors (PDV) of the original image set according to Eq. (2).

    Step 2. Get the Pixel Difference Matrix (PDM) of the original image set: $PDM = [PDV_1; PDV_2; \cdots; PDV_{ij}; \cdots; PDV_M]$.

    Step 3. Compute the $\tilde{S}_w$ and $\tilde{S}_b$ as the Eq. (15).

    Step 4. Compute the $\tilde{\Phi}$, $\tilde{\Phi}_1$ and $\tilde{\Phi}_2$, which are defined according to Eq. (19).

    Step 5. Solve α by taking the eigenvectors corresponding to the leading eigenvalues of Eq. (28).

    Step 6. The optimal solution α is obtain by splitting into two parts , as $\alpha = [\alpha_x; \alpha_y]$.
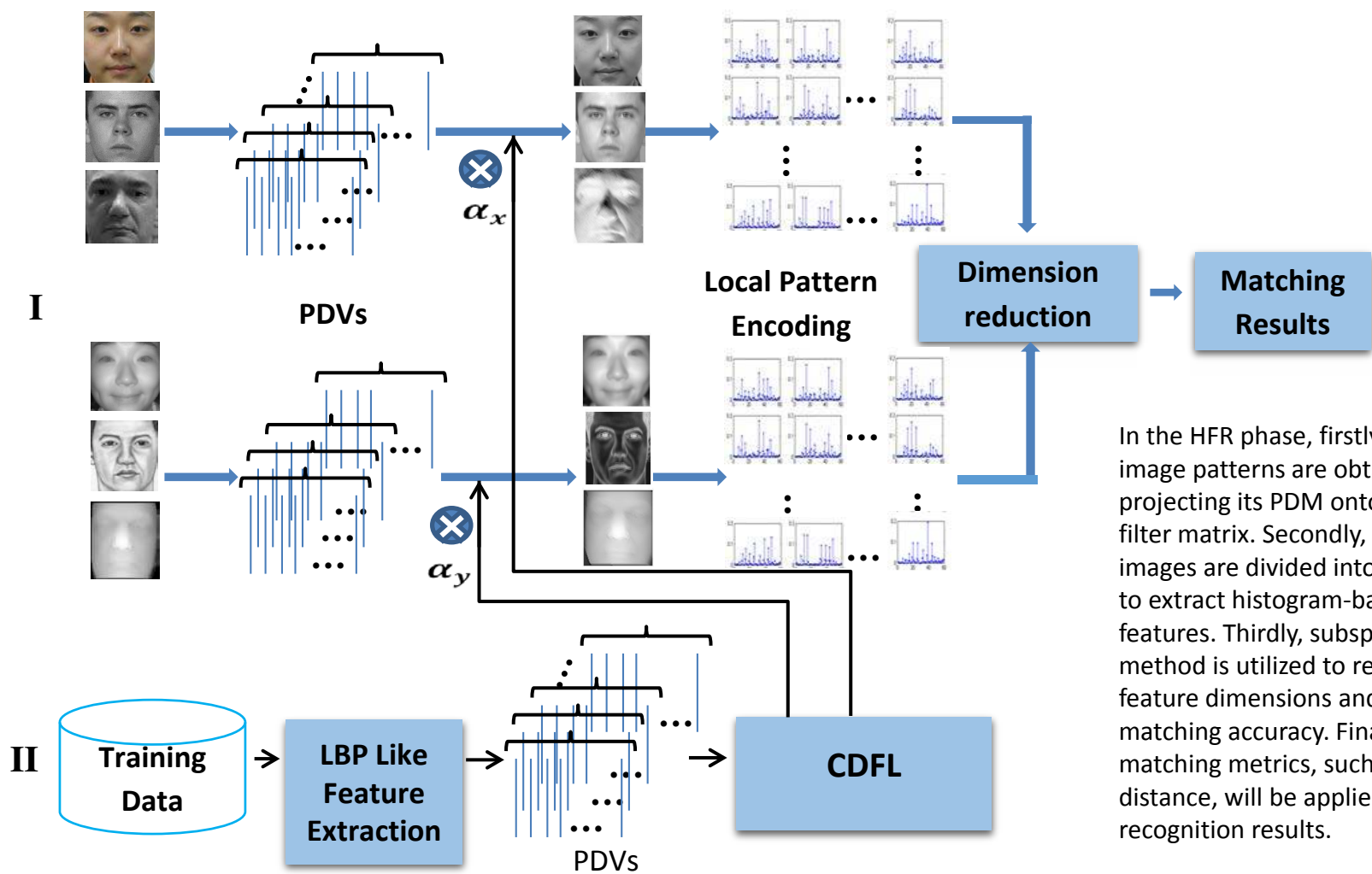
    Step 7. **return** $result : \alpha_x; \alpha_y$

end function

$$\Sigma \alpha = \lambda \mathbf{M} \alpha.$$

$$\Sigma = [\theta(S_b - S_w) + (1 - \theta)\Phi]$$

$$\mathbf{M} = (\Phi_1 + \Phi_2) = \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix}$$

Methodology: The CDFL approach

I   PDVs α_x α_y

Local Pattern Encoding

Dimension reduction → **Matching Results**

II   Training Data → LBP Like Feature Extraction → PDVs → CDFL

In the HFR phase, firstly, the new image patterns are obtained by projecting its PDM onto the learned filter matrix. Secondly, the filtered images are divided into image regions to extract histogram-based local features. Thirdly, subspace learning method is utilized to reduce the feature dimensions and improve the matching accuracy. Finally, direct matching metrics, such as cosine distance, will be applied to get the recognition results.

The whole process of Coupled Discriminative Feature Learning (CDFL) approach for HFR.

- Samples of three heterogeneous applications

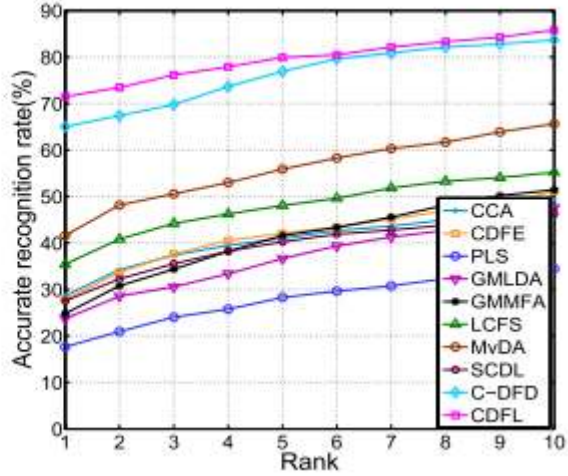(a). VIS vs. NIR   (b). Photo vs. Sketch   (c). 2D Photo vs. 3D range

- The CASIA NIR-VIS 2.0 face database. CVPR, 2013, Li et. al.
- Face photo-sketch synthesis and recognition. TPAMI 2009, Wang et. al.
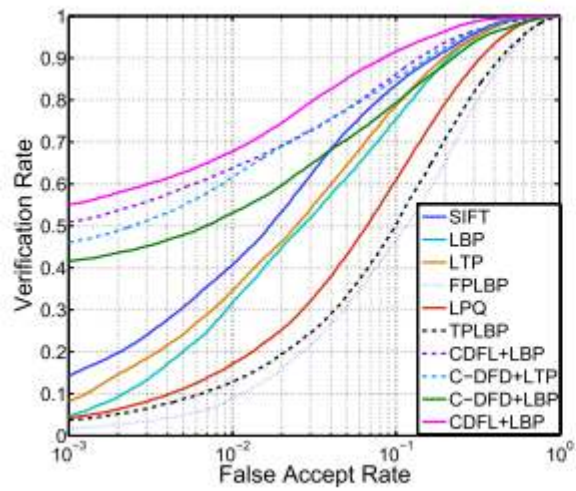- Overview of the face recognition grand challenge. CVPR 2005, Phillips et al.
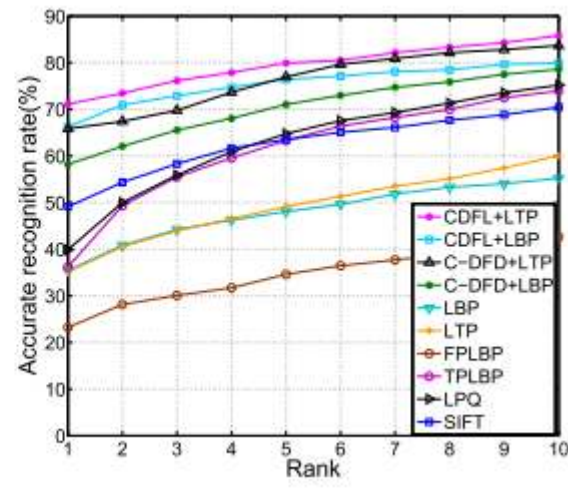
- Performance on CASIA NIR-VIS 2.0

(a)

(b)

(c)

(d)
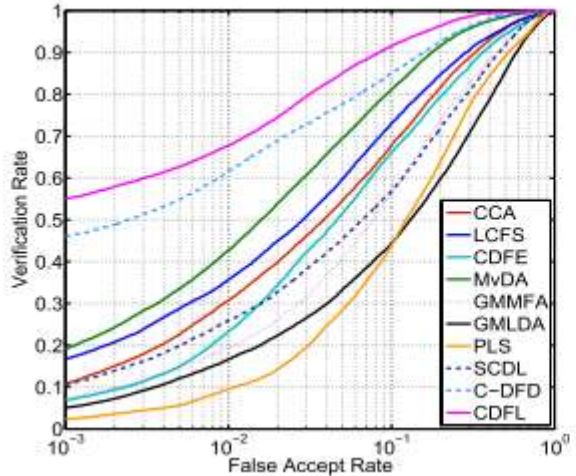
Experimental results

- Computational time
  - PC with an Intel 3.2GHZ i5 CPU, 8G RAM

COMPARISON OF THE FEATURE EXTRACTION TIME
ON CASIA VIS-NIR 2.0 DATABASE

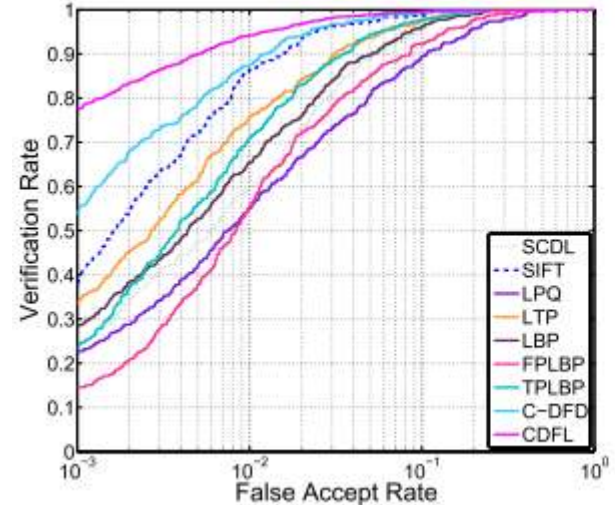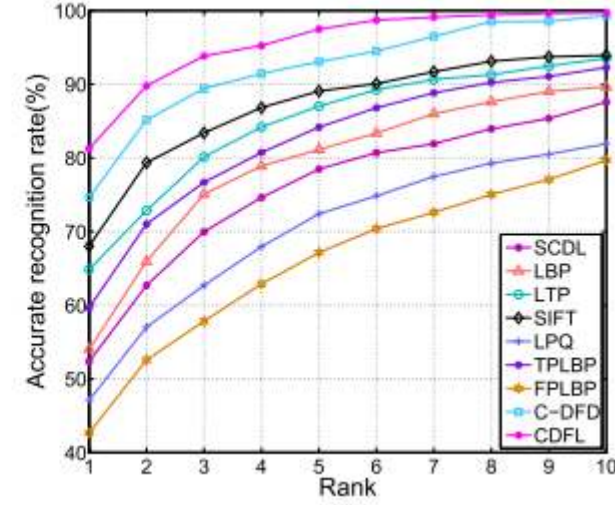| Methods	Training Time(s)	Methods	Training Time(s)
CCA [19]	2.31 ± 0.14	LBP [24]	29.58 ± 0.93
PLS [11]	91.26 ± 1.58	LTP [42]	59.65 ± 1.01
CDFL [8]	26.65 ± 0.51	TP-LBP [44]	187.12 ± 4.47
MvDA [43]	32.36 ± 1.02	FP-LBP [44]	191.88 ± 5.25
LCFS [27]	14.68 ± 0.61	LPQ [50]	254.12 ± 6.72
GMMFA [28]	12.57 ± 0.39	SIFT [25]	841.91 ± 4.65
GMLDA [28]	12.21 ± 0.30	C-DFD [40]	51.48 ± 0.75
CDFL	$\mathbf{27.20 \pm 0.97}$	–	–

- Performance on CUFSF

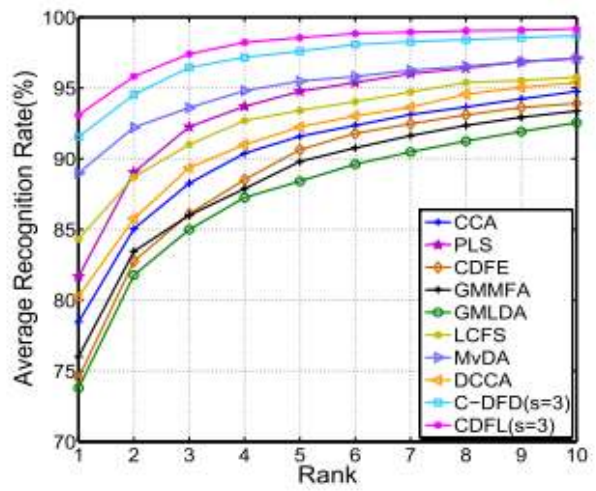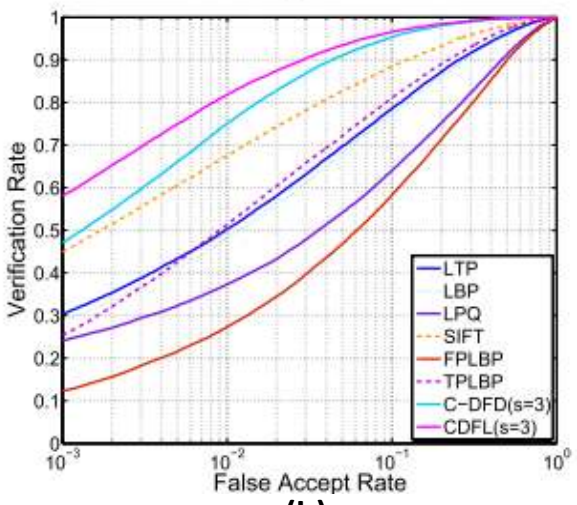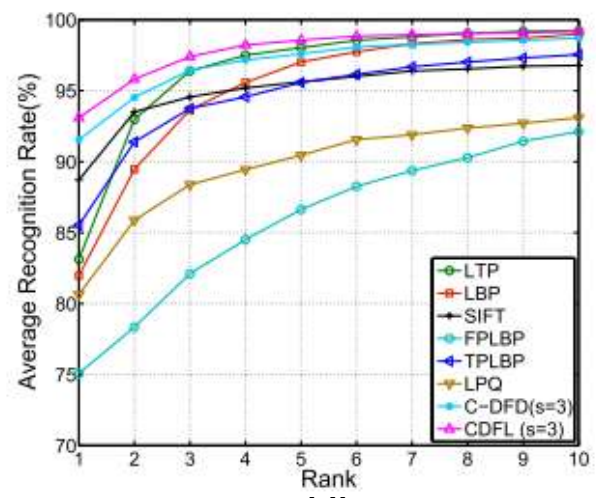(a)

(b)

(c)

(d)
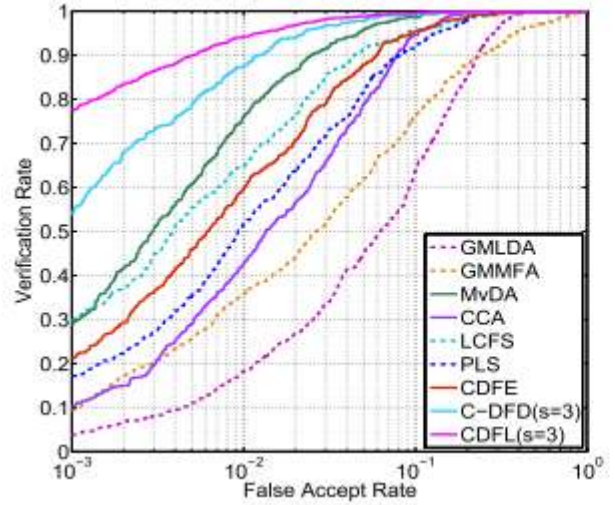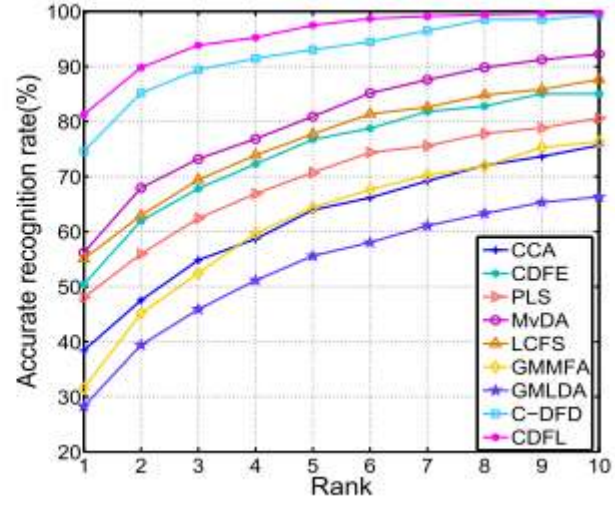
Experimental results

- Performance on FRGC

(a)

(b)

(c)

(d)

Experimental results
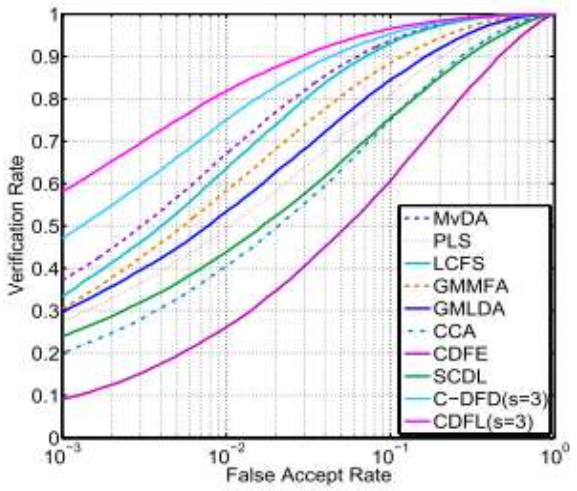
- Observations

  ➢ Feature learning based methods perform better than hand-crafted methods in different HFR applications.

  ➢ By learning both cross-modality discriminant and correlation features, CDFL works effectively on three different HFR scenarios.

Conclusion and Future Work

- Conclusions
  - Coupled Discriminative Feature Learning is proposed for HFR.
  - The most discriminant and correlative image patterns are gathered in CDFL.
  - Experiments on three different HFR scenarios demonstrate the effectiveness and generalization.

- Future Work
  - To investigate the weight of these eigenvectors of CDFL in heterogeneous face analysis.

Thanks for your attention!

Please contact authors via email if you have questions and feedbacks are welcome.