Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition

Wei Zhang, Xiaogang Wang, and Xiaoou Tang
The Chinese University of Hong Kong

Outline

- Introduction
  - Face sketch recognition
  - Inter-modality image matching
- Our Approach
  - Coupled encoding
  - Maximum mutual information criterion
  - Building coupled information-theoretic tree
  - Coupled information-theoretic encoding descriptor
- Experimental Results
- Conclusions and Future Work

Face Sketch Recognition

- Match a face sketch drawn by the artist with face photos in the database
- Application in law enforcement
  - If the photo of a suspect is not available, the best substitute is the sketch drawn by the art according to description of the witness
- **It is a inter-modality image matching problem**

Face sketch

Face photo database

Inter-Modality Image Matching

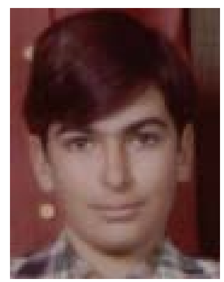
- There exists an unknown **transform** between images of different **“styles”**

Face sketch recognition | **Infrared face recognition** | **face recognition across ages** | **face recognition across resolutions** | **Object matching across cameras**

Style A


Photo


Optical image


Age 1

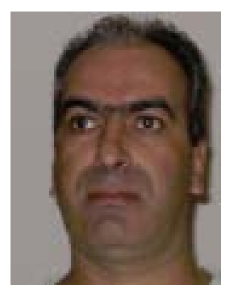

High resolution


Camera 1

Style B


Sketch


Infrared image


Age 2


Low resolution


Camera 2

Solutions of Inter-Modality Matching

- *Preprocessing stage*: **transform** images from one "style" to the other

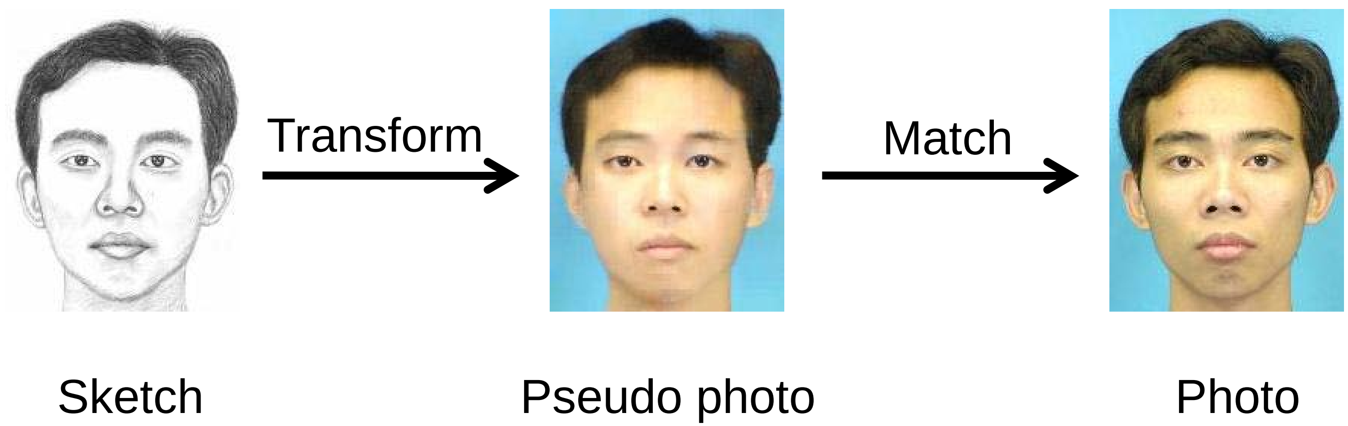


Photo-Sketch synthesis: X. Wang and X. Tang, "Face Sketch Synthesis and Recognition," *IEEE Trans. on PAMI*, Vol. 31, pp. 1955-1967, 2009.

Solutions of Inter-Modality Matching

- *Preprocessing stage*: **transform** images from one modality to the others
  - **Synthesis is a harder problem than recognition**

Face synthesis from infrared images: J. Chen, D. Yi, J. Yang, G. Zhao, S. Z. Li, M. Pietikainen, "Learning Mappings for Face Synthesis from Near Infrared to Visual Light Images," *CVPR,* 2009.

Face synthesis across ages: J. Suo, S. C. Zhu, S. Shan, and X. Chen, "A Compositional and Dynamic Model for Face Aging," *IEEE Trans. on PAMI*, Vol. 32, pp. 385-401, 2010

Face synthesis across ages: U. Park, Y. Tong, and A. K. Jain, "Age-Invariant Face Recognition," *IEEE Trans. on PAMI*, Vol. 32, pp. 947-954, 2010

Brightness transfer function between cameras: O. Javed, K. Shafique, and M. Shah, "Appearance Modeling for Tracking in Multiple Non-overlapping Cameras," *CVPR,* 2005.

Solutions of Inter-Modality Matching

- *Classification stage*: design advanced classifiers to reduce the gap between features extracted from images of different modalities
  - The inter-modality difference between the extracted features may be too large for the classifiers

Sketch-Photo recognition : B. Klare, Z. Li, and A. K. Jain, “Matching Forensic sketches to mugshot photos ,” *IEEE Trans. on PAMI*, Vol. 33, pp. 639-646, 2011.

Infrared-optical recognition: Z. Lei and Z. Li, “Coupled Spectral Regression for Matching Heterogeneous face,” *CVPR*, 2009.

Sketch-Photo recognition & Infrared-optical recognition:
D. Lin and X. Tang, “Inter-Modality Face Recognition,” *ECCV*, 2006.

Face recognition across ages: N. Ramanathan, R. Chellappa, “Face Verification across Age Progression,” *IEEE Trans. on Image Processing*, Vol. 15, pp. 3349-3361, 2006.

Solutions of Inter-Modality Matching

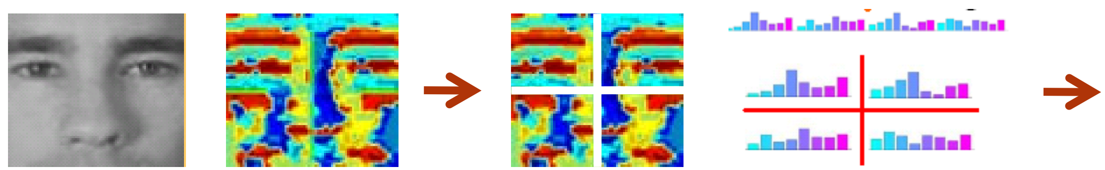
- Our approach: reduce the modality gap at the *feature extraction stage*

Outline

- Introduction
  - Face sketch recognition
  - Inter-modality image matching
- Our Approach
  - Coupled encoding
  - Maximum mutual information criterion
  - Building the coupled information-theoretic tree
  - Coupled information-theoretic encoding descriptor
- Experimental Results
- Conclusions and Future Work

Recognition Pipeline

Quantized local features (codes) → Histograms in local regions → Classifier



PCA + LDA classifier

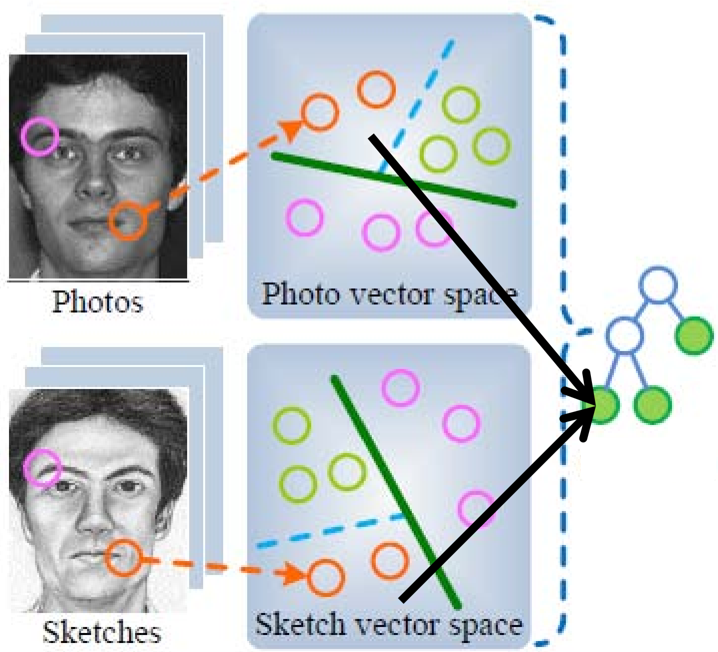
Face recognition: T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Applications to Face Recognition,” *IEEE Trans. on PAMI*, Vol. 28, 2006.

Object detection: N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” *CVPR*, 2005.

Object recognition: S. Lazebnik, C. Schmid, and J. Ponce, “Beyond Bag of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,” *CVPR*, 2006.

Coupled Encoding

- Encoding for a single modality: k-means, mean shift (Jurie & Triggs ICCV'05), random projection tree (Wright & Hua CVPR'09), random forest (Shotton CVPR'08)
- ***Coupled encoding for cross-modality quantization***



Coupled projection tree: mapping local structures to **one codebook** with two coding functions C_p and C_s for photos and sketches

Coupled Projection Tree

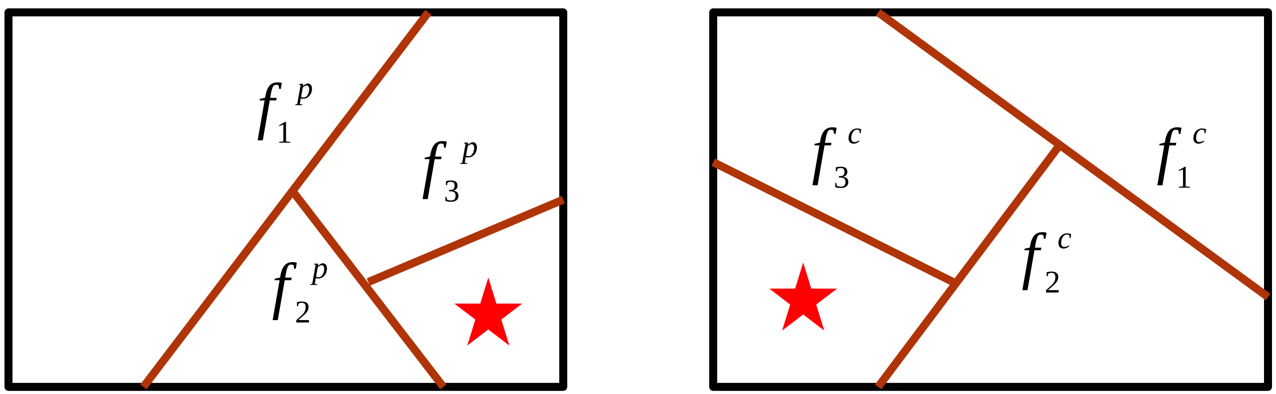


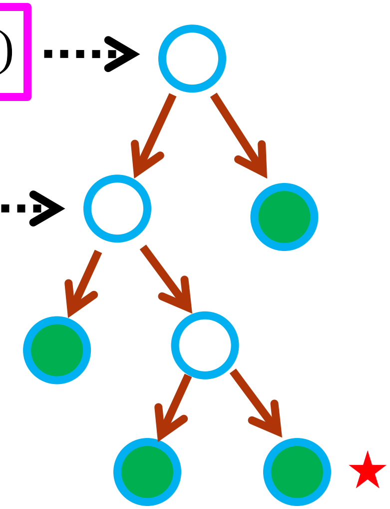
Photo feature space

Sketch feature space

$$f_1^p\left(x^p\right) = sign(w_1^p x^p - \tau_1^p) \qquad f_1^c\left(x^c\right) = sign(w_1^c x^c - \tau_1^c)$$

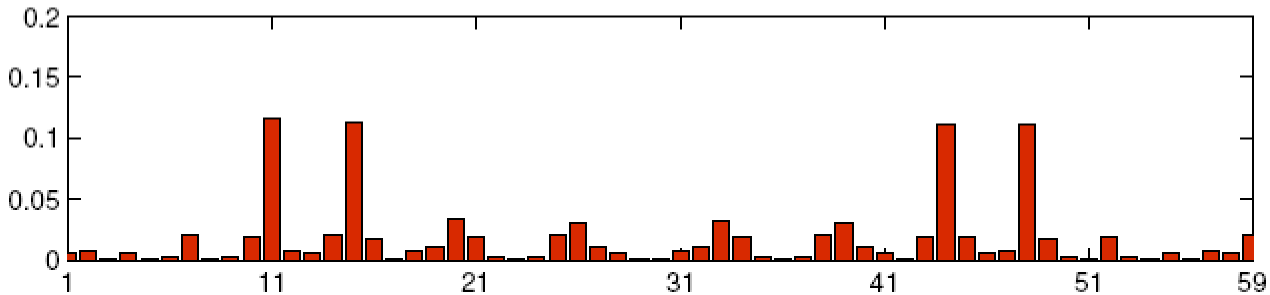
$$f_2^p\left(x^p\right) = sign(w_2^p x^p - \tau_2^p) \qquad f_2^c\left(x^c\right) = sign(w_2^c x^c - \tau_2^c)$$

...



Maximum Mutual Information Criterion

- What are good quantized local features
  - ***High discriminative power***: codes uniformly distribute across different subjects



LBP code frequency histogram

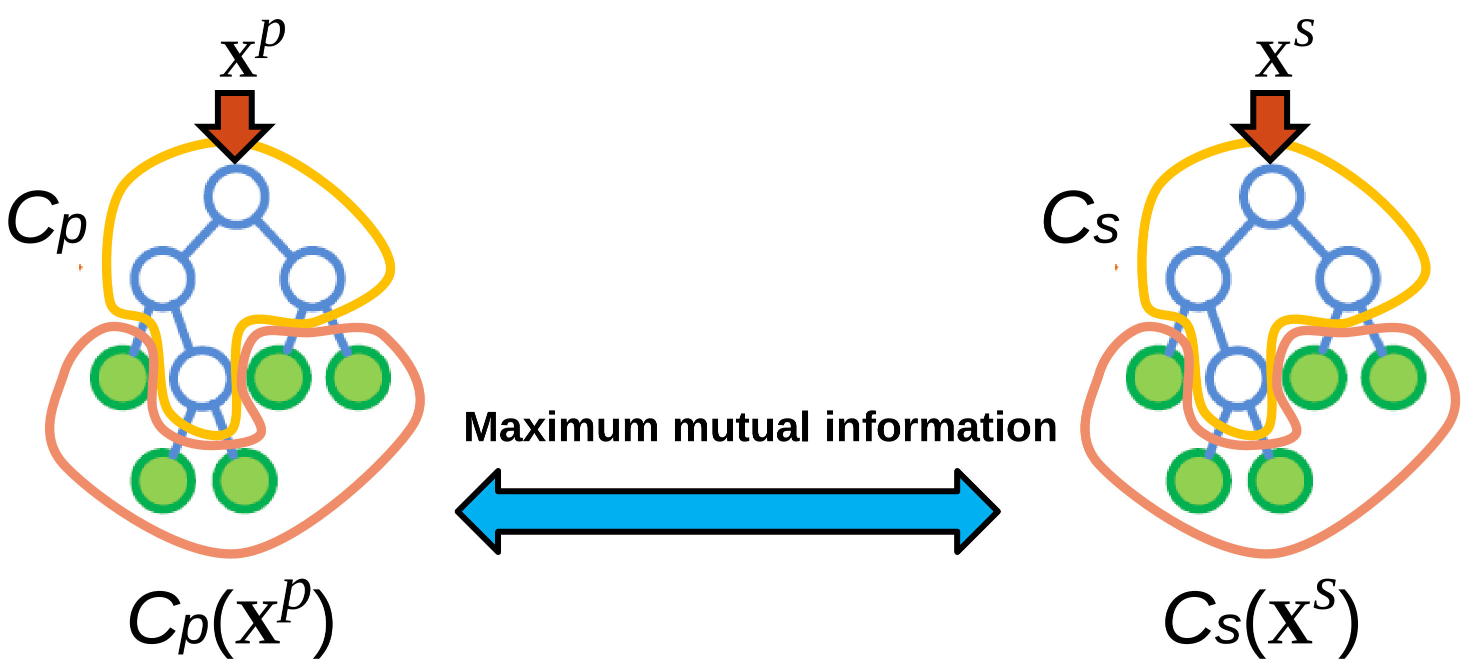
  - ***Low inter-modality gap***: the codes of the same subject's photo and sketch are highly correlated
  - These two requirements lead to the ***maximum mutual information criterion***

$$I(\mathcal{C}_p(\mathbf{X}^p);\mathcal{C}_s(\mathbf{X}^s)) = H(\mathcal{C}_p(\mathbf{X}^p)) - H(\mathcal{C}_s(\mathbf{X}^p)|\mathcal{C}_p(\mathbf{X}^s))$$

Training a Couple Information-Theoretic Tree

- Training set: vector pairs $\mathcal{X} = \{(\mathbf{x}_i^p, \mathbf{x}_i^s), i = 1, ..., N\}$

$$\mathbf{X}^p = [\mathbf{x}_1^p, ..., \mathbf{x}_N^p] \quad \mathbf{X}^s = [\mathbf{x}_1^s, ..., \mathbf{x}_N^s]$$



$$I(\mathcal{C}_p(\mathbf{X}^p); \mathcal{C}_s(\mathbf{X}^s)) = H(\mathcal{C}_p(\mathbf{X}^p)) - H(\mathcal{C}_p(\mathbf{X}^p)|\mathcal{C}_s(\mathbf{X}^s))$$

Training a Couple Information-Theoretic Tree

- Parameter searching for a node k: $w_k^p, w_k^s, \tau_k^p, \tau_k^s$
  - Under the Gaussian assumption, w_k^p and w_k^c has the closed form solution

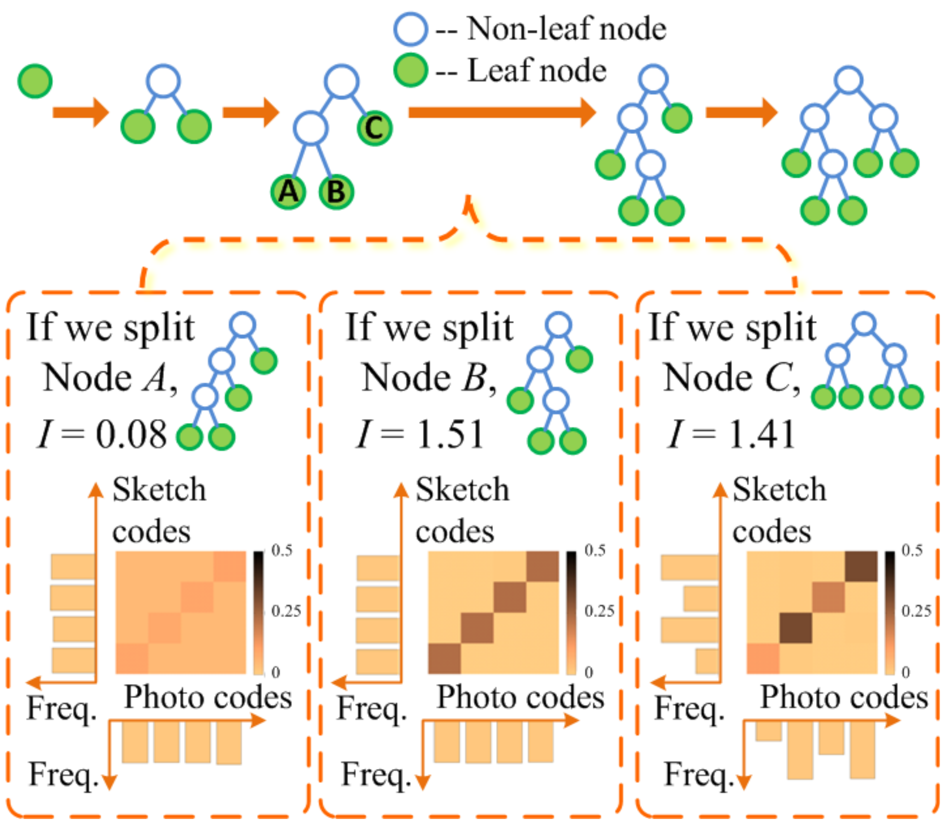
$$\max_{w_k^p, w_k^s} \frac{\left(w_k^p\right)^T C_k^{p,s} w_k^s}{\sqrt{\left(w_k^p\right)^T C_k^p w_p \left(w_k^s\right)^T C_k^s w_k^s}}$$

Where C_k^p and C_k^s are the covariance matrices of photo vectors and sketch vectors assigned to node k in **X**^{p} and **X**^{s} respectively; $C_k^{p,s}$ is the covariance matrix between photo vectors and sketch vectors.

  - τ_k^p and τ_k^s are found by brute force search

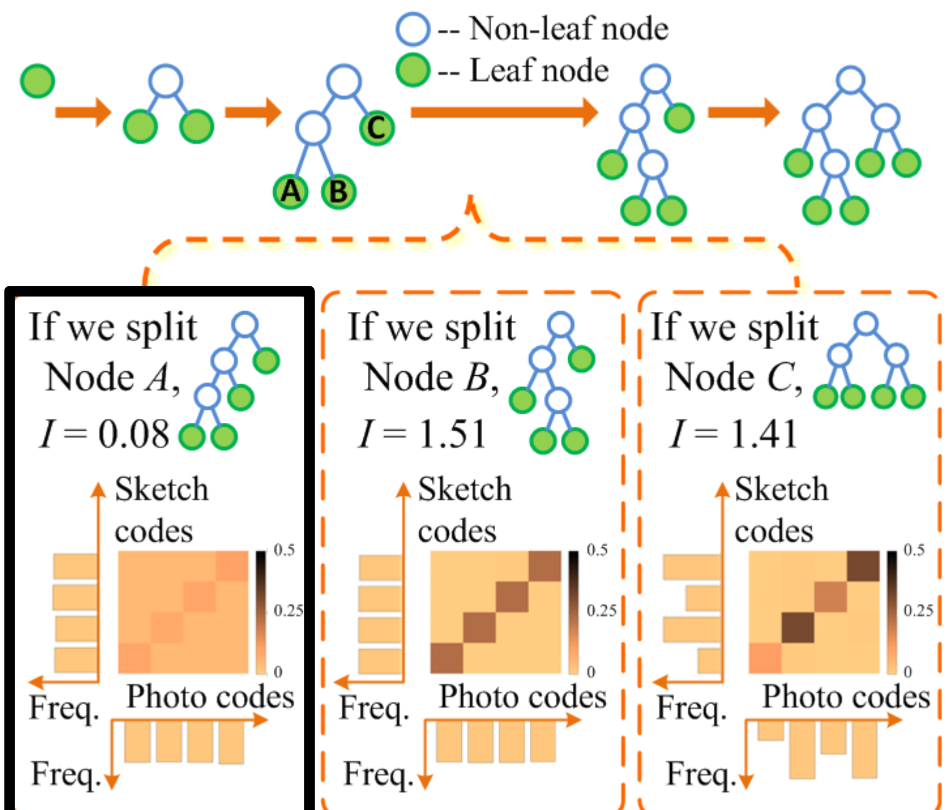
Training a Couple Information-Theoretic Tree

- Tree structure searching
  - Search the node whose splitting can maximize the mutual information



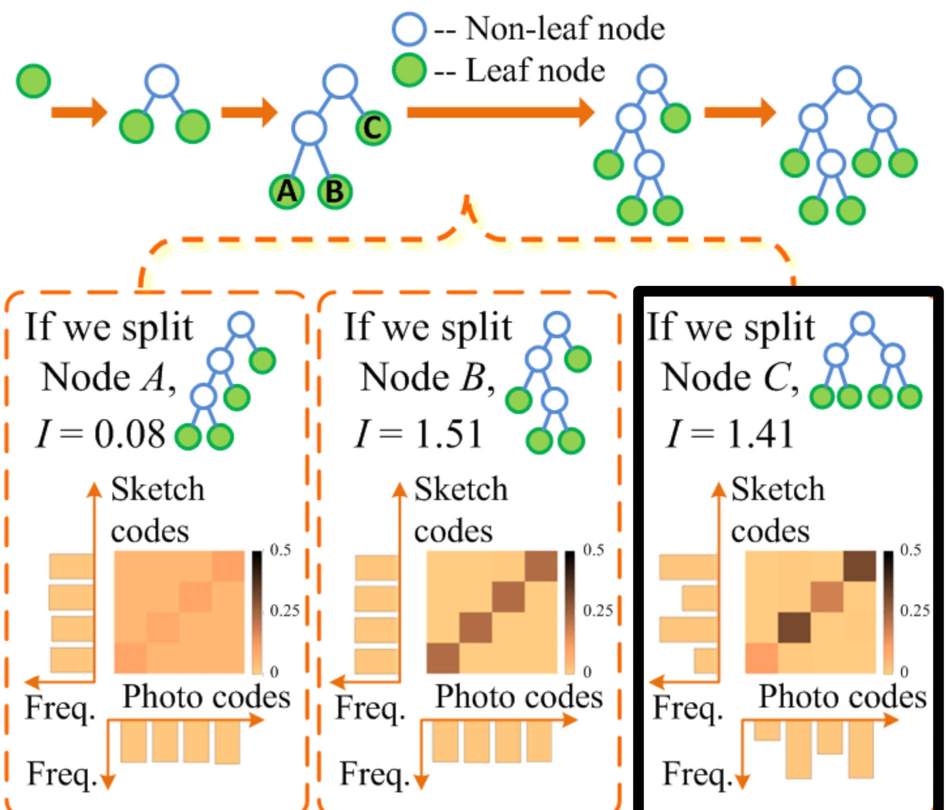
Training a Couple Information-Theoretic Tree

- Tree structure searching
  - Search the node whose splitting can maximize the mutual information



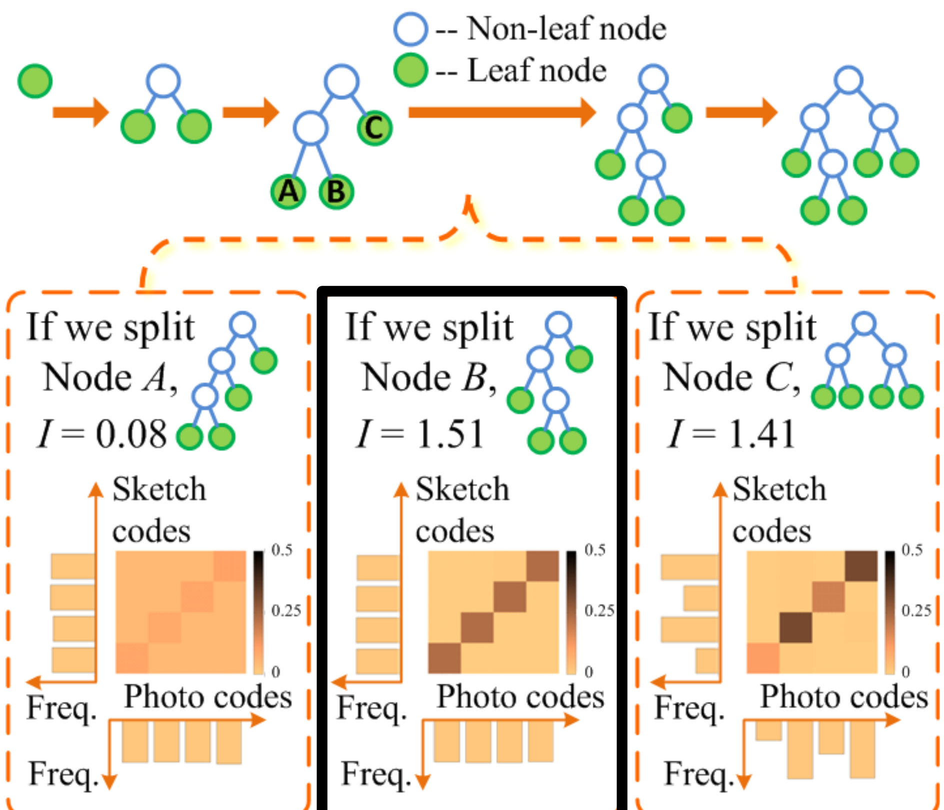
Training a Couple Information-Theoretic Tree

- Tree structure searching
  - Search the node whose splitting can maximize the mutual information



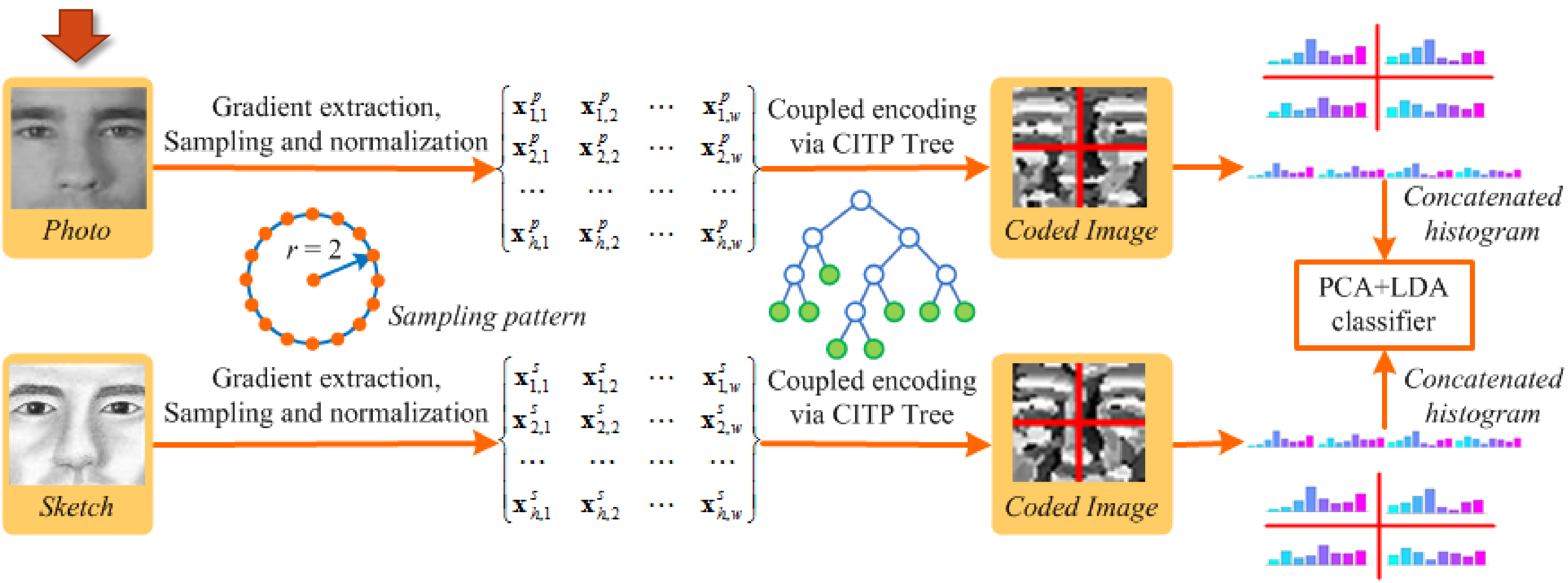
Training a Couple Information-Theoretic Tree

- Tree structure searching
  - Search the node whose splitting can maximize the mutual information



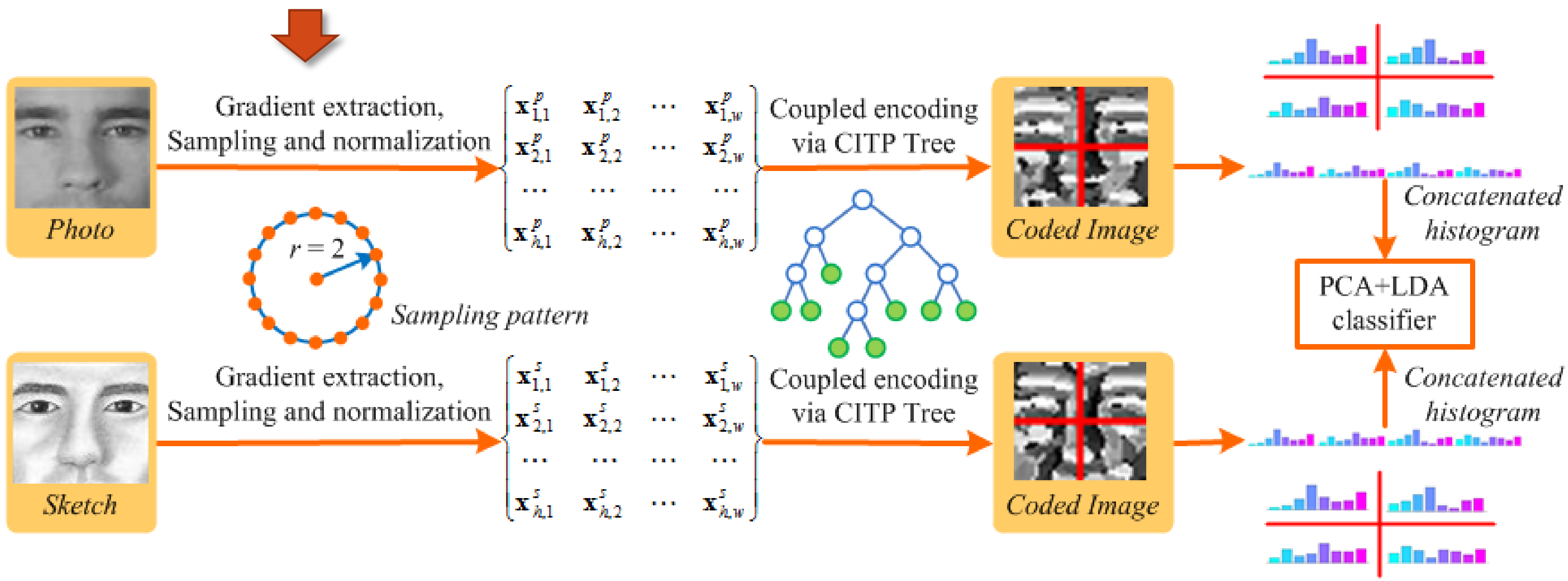
Coupled Encoding Descriptor

- Geometric rectification
- Photometric rectification using Difference-of-Gaussians (DoG) filter



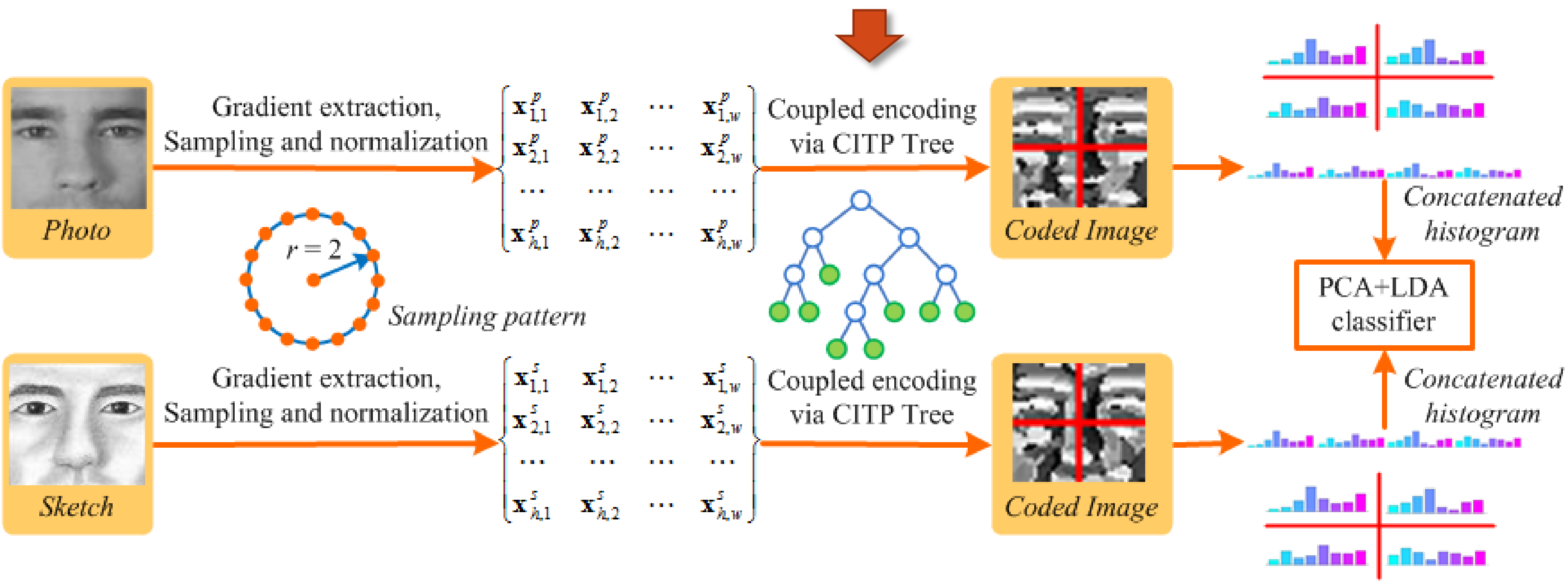
Coupled Encoding Descriptor

- Extracting a local feature vector by sampling the normalized gradients around a pixel



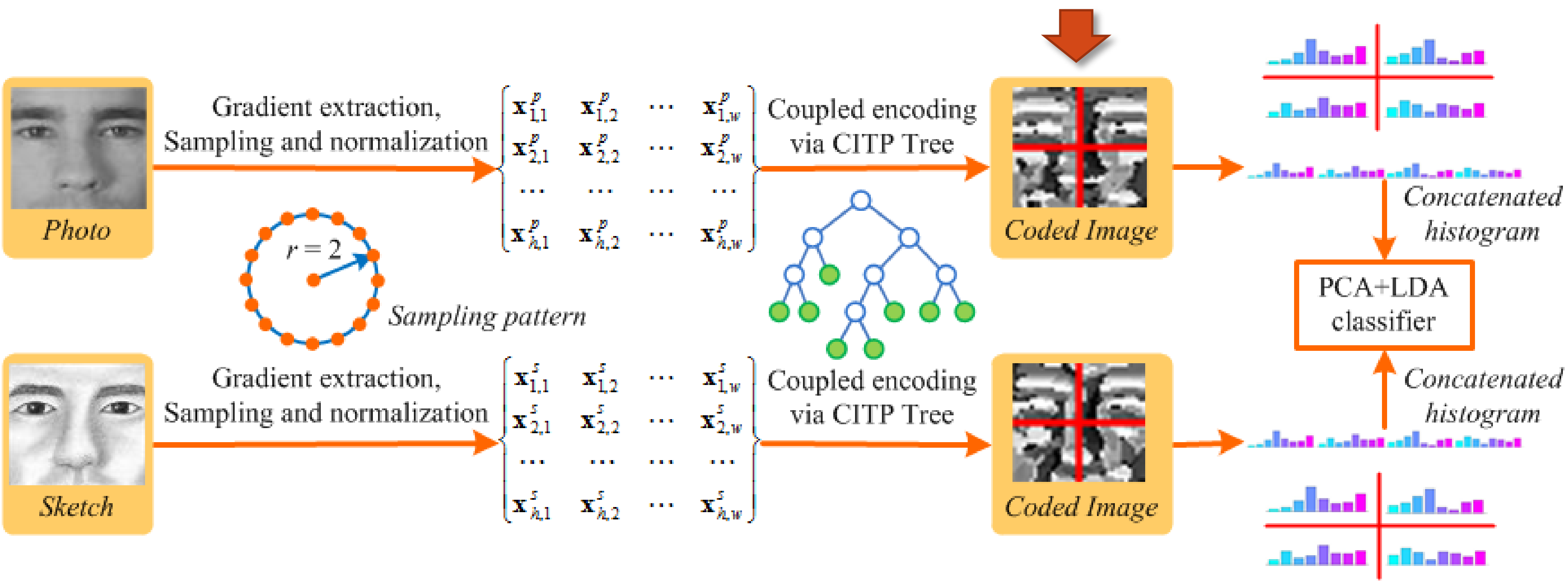
Coupled Encoding Descriptor

- Coupled information theoretic encoding



Coupled Encoding Descriptor

- CITE descriptors: histograms in local regions
- Classifiers: PCA + LDA
- Fusion of distances by different CITE descriptors: Linear SVM



Outline

- Introduction
  - Face sketch recognition
  - Inter-modality image matching
- Our Approach
  - Coupled encoding
  - Maximum mutual information criterion
  - Building coupled information-theoretic tree
  - Coupled information-theoretic encoding descriptor
- Experimental Results
- Conclusions and Future Work

CUHK Face Sketch Database (CUFS)

- 188 people from the CUHK student database

- 123 people from AR database

- 295 people from XM2VTS database

http://mmlab.ie.cuhk.edu.hk/facesketch.html

Experimental Results on CUFS

- 306 persons for training and 300 for testing

Direct match	MRF+ RS-LDA	LFDA (LBP + SIFT)	Ours
6.3%	96.3%	99.47%	99.87%

- MRF + RS-LDA: Wang TPAMI'09
- FLDA (SIFT + LBP): Klare TPAMI'11

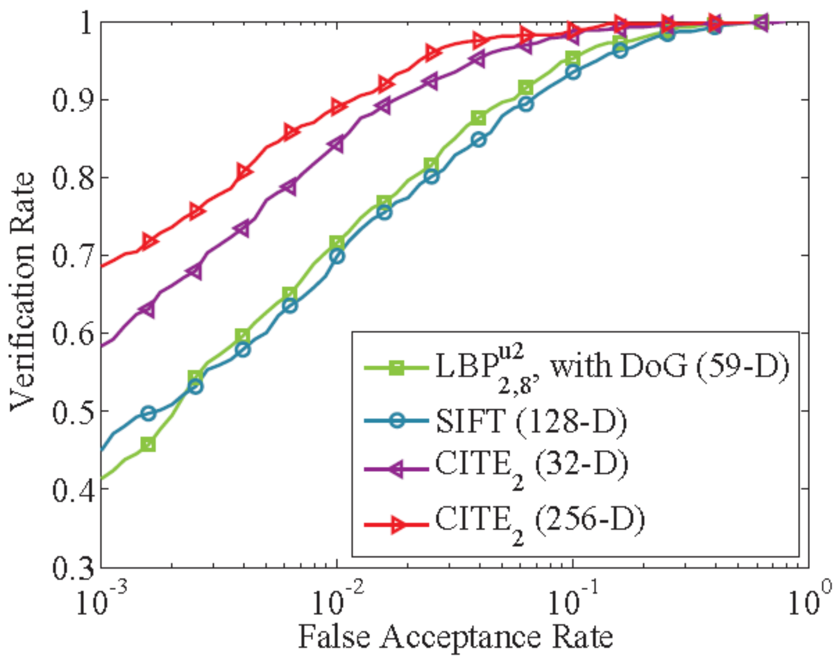
CUHK Face Sketch FERET Database (CUFSF)

- 1,196 people from the FERET database

http://mmlab.ie.cuhk.edu.hk/cufsf/

Experimental Results on CUFSF

- 500 persons from CUFSF are randomly selected for training and 694 for testing



Descriptor comparison

Experimental Results

Verification rate at 0.1% false alarm rate		
MRF + RS-LDA	MRF + LE	LFDA (SIFT + LBP)
29.54%	**43.66%**	**90.78%**
Kernel CSR (LBP)	Kernel CSR (SITF)	Ours
64.55%	**88.18%**	**98.7%**

- MRF based synthesis (Wang TPAMI'09) first transforms photos to sketches and them match with different classifiers, RS-LDA (Wang IJCV'06) or LE (Cao CVPR'10)
- FLDA (SIFT + LBP) is from Klare TPAMI'11
- Kernel CSR: kernel couple spectral regression from Lei CVPR'09
- Ours combines CITE and PCA+LDA

Conclusions and Future Work

- Propose coupled encoding for cross-modality quantization
- Introduce the maximum mutual information criterion to guide the encoding
- Propose a new algorithm of building coupled information-theoretic tree
- The new coupled information-theoretic encoding descriptor significantly outperforms existing approach on face sketch recognition
- Contribute a large scale face sketch database
- Explore other applications of coupled information-theoretic encoding in the future work

Thank you!