11th IEEE International Conference on
Automatic Face and Gesture Recognition
FG2015

Shared Representation Learning for Heterogeneous Face Recognition

Dong Yi, Zhen Lei, Stan Z. Li
Center for Biometrics and Security Research
Institute of Automation, Chinese Academy of Sciences

Heterogeneous Face Recognition

- Matching face images from different sources
- Sketch vs. VIS (Visual)
  - forensics and security applications
- NIR (Near Infrared) vs. VIS
  - An alternative way to deal with illumination problem
  - Make NIR system be compatible with legacy VIS face images

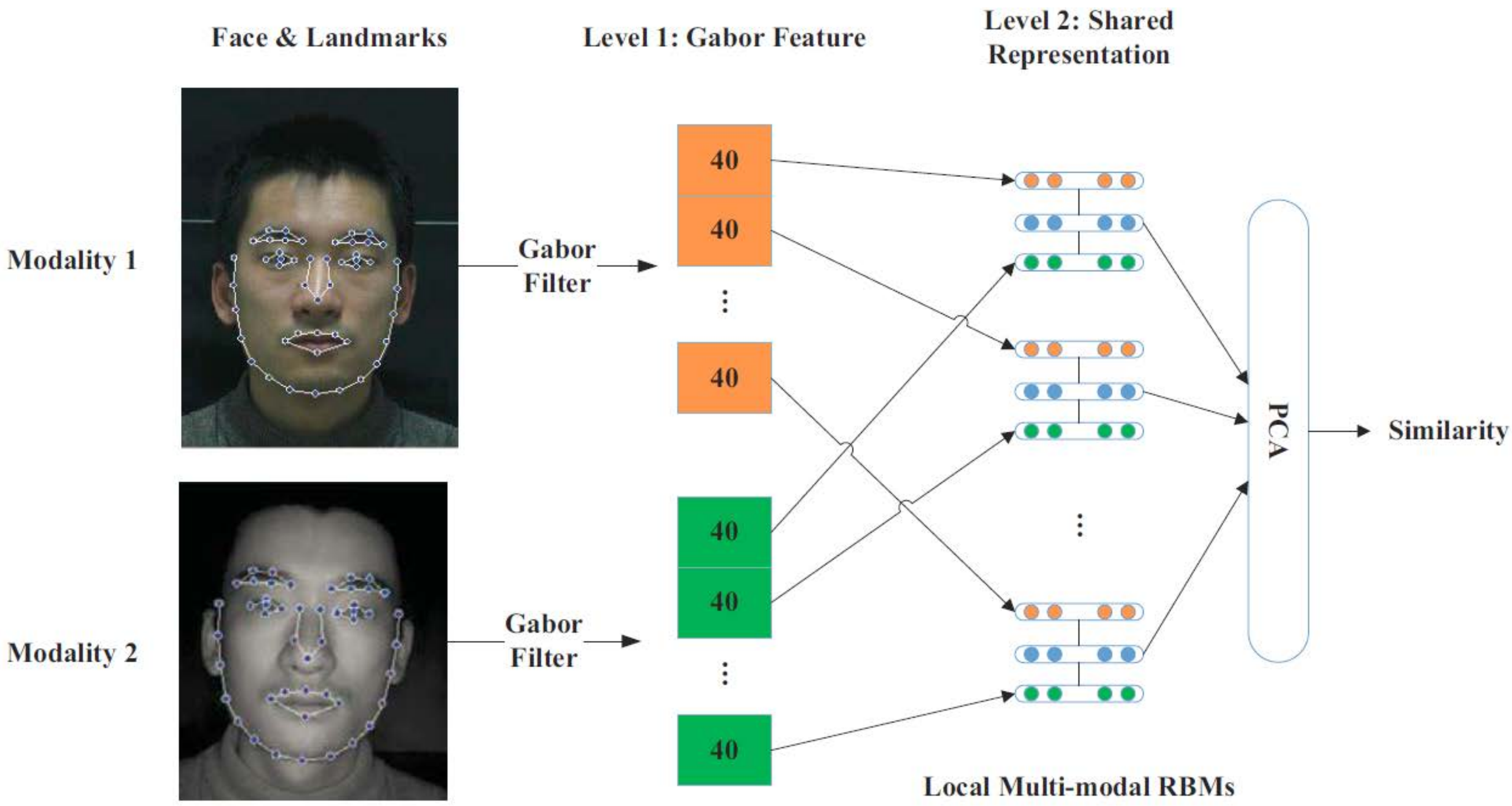
Challenges

- Ordinary VIS-VIS Face recognition is still an unsolved problem
- Nonlinear relation between face images with different modalities
- Limited training samples
- Solution: Multi-modal Restricted Boltzmann Machines (RBMs)

Related Work

- Synthesis based model
  - Eigen-transformation, CVPR 2002
  - Patch-wise LLE, CVPR 2003
  - Transforming VIS to NIR, ICB 2009
- Discriminative model
  - DoG filter, LBP, HOG, ICB 2009, PAMI 2011
  - Coupled information theoretic encoding, CVPR 2011
  - Coupled discriminant analysis, TIFS 2012

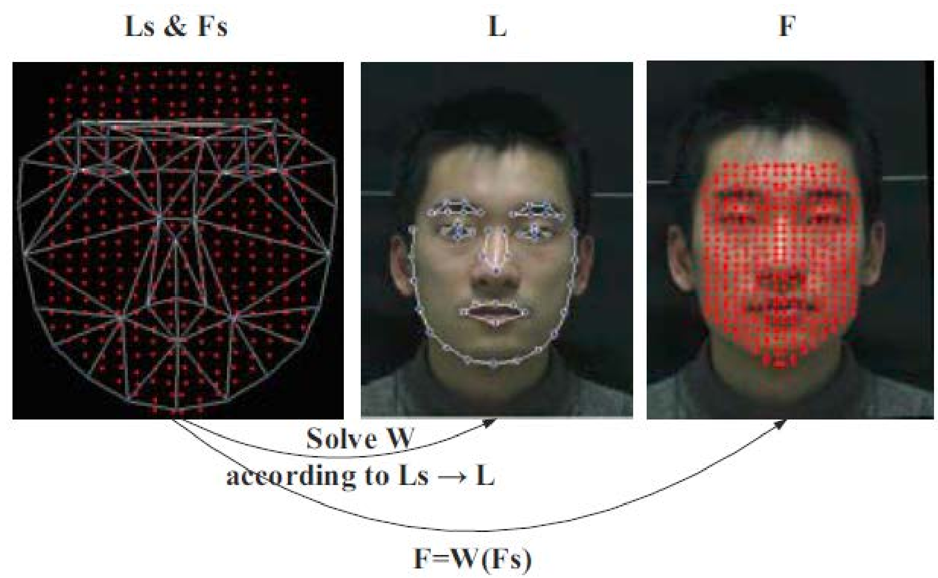
Our Method

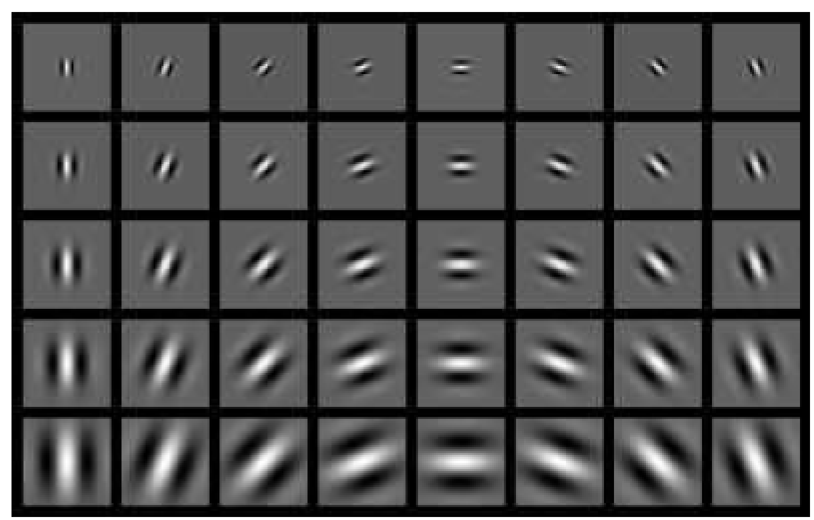


Face & Landmarks
Level 1: Gabor Feature
Level 2: Shared Representation
Modality 1
Modality 2
Gabor Filter
Gabor Filter
40
40
40
40
40
40
PCA
Similarity
Local Multi-modal RBMs

Level 1 Features

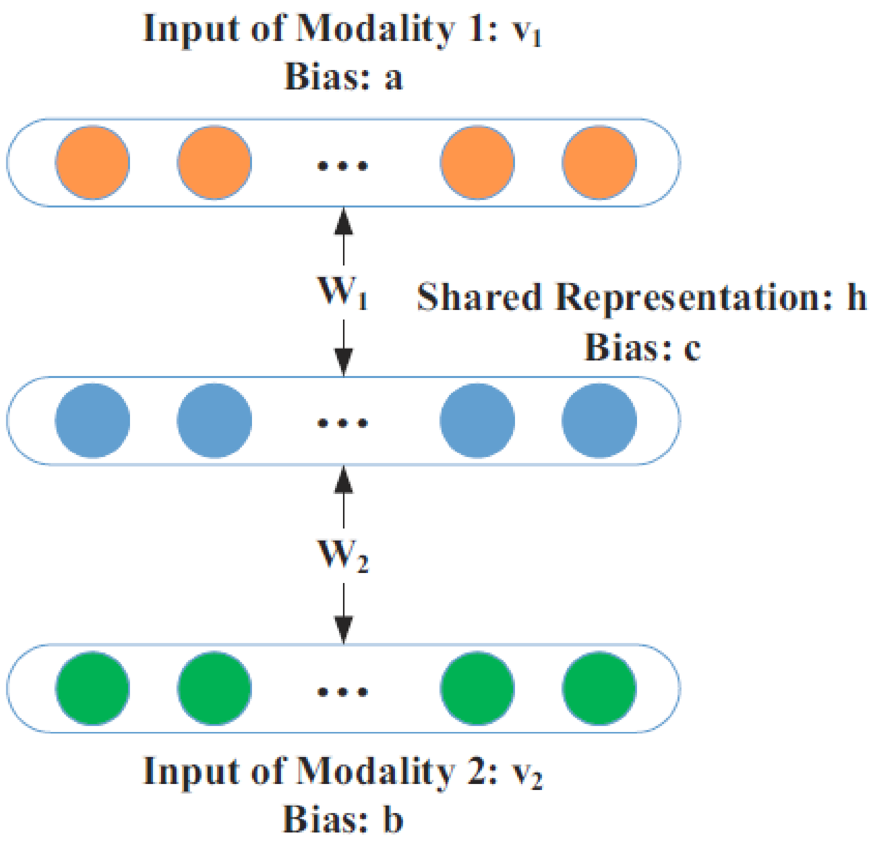
- Identity adaptive feature points
  - Localize sparse facial landmarks (48 pts)
  - Generate adaptive feature points by RBF warping (176x2 pts)
- Gabor filtering
  - 5 scales and 8 orientations
- PIE robust local features
  - 176x2x40 dimensions





Level 2 Representations

- Build the relationship between two modalities by multi-modal RBMs
- Train RBM for each feature point
- By using facial symmetry, need train 176 RBMs
- The structure of one RBM is 40-80-40
- Dimension: 176x2x80



Multi-Modal RBM

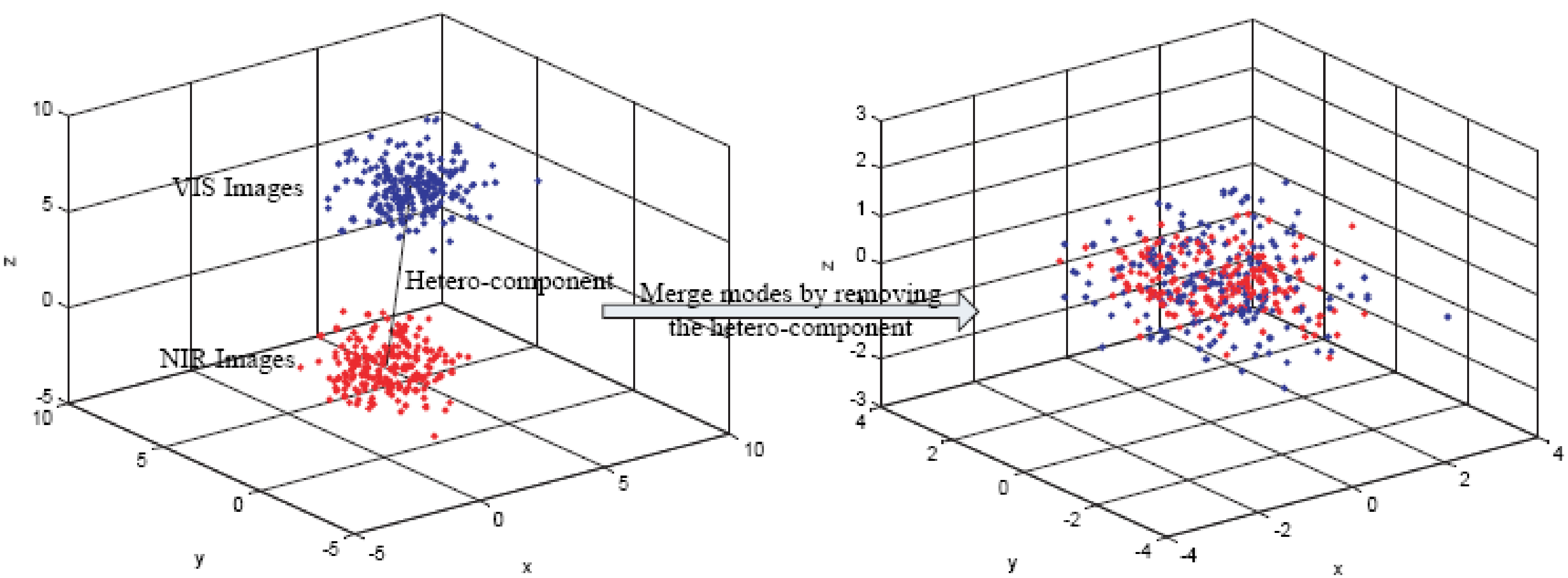
- The parameters are learned by using mean-field inference and an MCMC procedure
- Generate the hidden (shared) representation ***h*** by alternating Gibbs sampler

$$
\begin{aligned}
E(\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \mathbf{h}; \theta) = & \frac{1}{2}(\hat{\mathbf{v}}_1 - \mathbf{a})^T (\hat{\mathbf{v}_1} - \mathbf{a}) + \\
& \frac{1}{2}(\hat{\mathbf{v}}_2 - \mathbf{b})^T (\hat{\mathbf{v}_2} - \mathbf{b}) - \\
& \mathbf{c}^T \mathbf{h} - \hat{\mathbf{v}}_1^T \mathbf{W}_1 \mathbf{h} - \hat{\mathbf{v}}_2^T \mathbf{W}_2 \mathbf{h},
\end{aligned}
$$

- Refer: Nitish Srivastava and Ruslan Salakhutdinov. “Multimodal learning with deep boltzmann machines”. NIPS , 2012.

Removing Heterogeneous Principle Components

- The correlation factor between the first PC and $\mathbf{h} = \mathbf{m}_{NIR} - \mathbf{m}_{VIS}$ is 0.97
- Remove the first several PCs



Experiments

- Take NIR-VIS face recognition as main experiments
  - CASIA HFB dataset
  - CASIA NIR-VIS 2.0 dataset
- Sketch-VIS experiments
  - CUFS dataset
  - CUFSF dataset
- Performance: Rank1 recognition rate and ROC

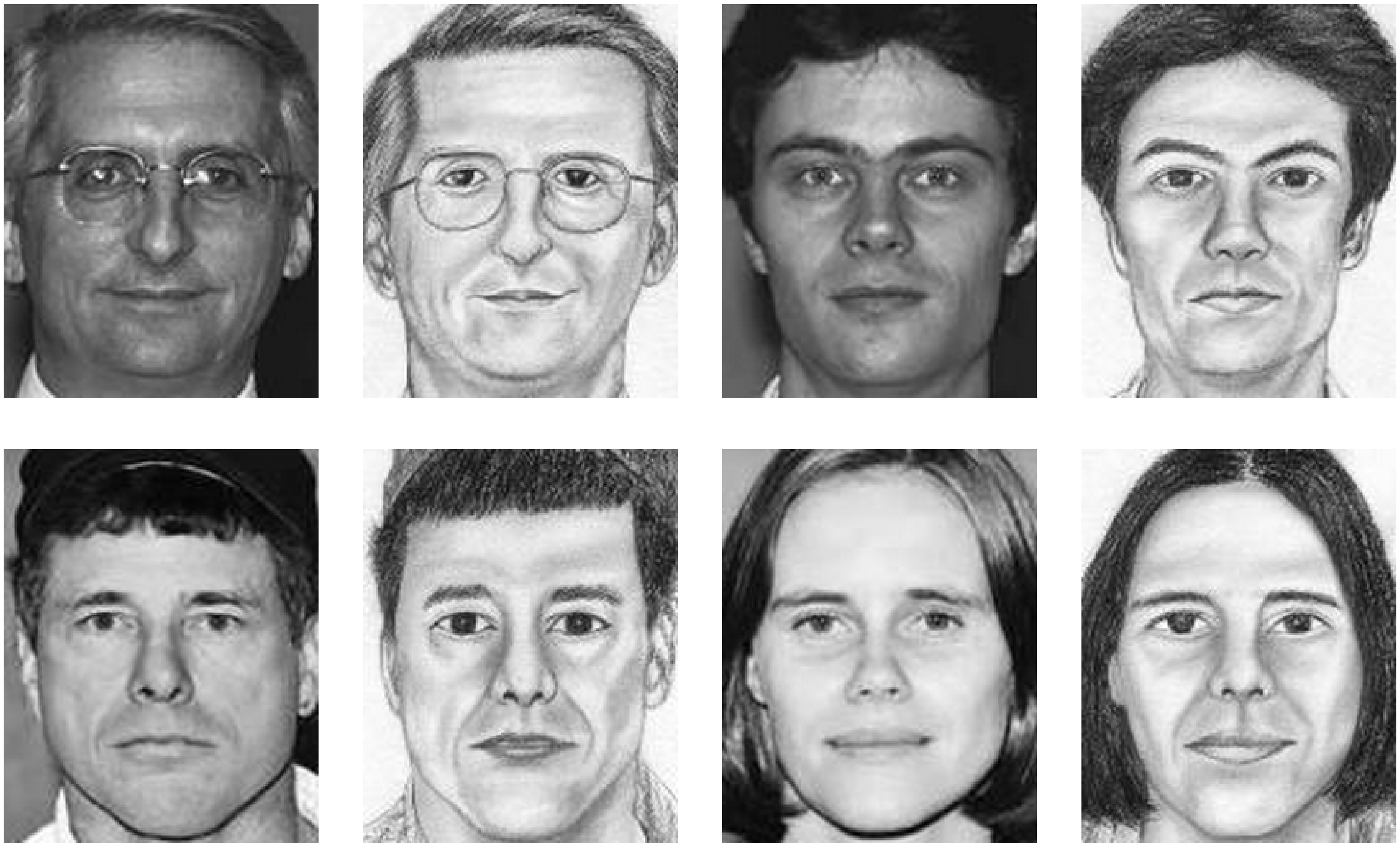
Datasets Information

Name	Modalities	No. of Subjects	No. of Images	Variations
CASIA HFB	NIR, VIS	202	5097	E, G, D
CASIA NIR-VIS 2.0	NIR, VIS	725	17580	P, E, G, D
CUFS	Sketch, VIS	606	1212	-
CUFSF	Sketch, VIS	1194	2388	-

Datasets used in our experiments. The last column denotes the main variations they addressed, *i.e.,* Pose(P), Expression(E), Eyeglasses(G), Distance(D).

Sample Images

Sample Images



Results on CASIA HFB

RANK1 RECOGNITION RATES AND VR@FAR=0.1% OF VARIOUS METHODS ON VIEW 2 OF CASIA HFB.

	Rank1	VR
Gabor	59.47±6.72%	33.51±5.70%
Gabor + Remove 20 PCs	94.87±1.72%	71.70±6.42%
Gabor + RBM	98.12±1.13%	84.50±3.75%
Gabor + RBM + Remove 11 PCs	99.38±0.32%	92.25±1.68% [1]
NN [12]	88.8%	48.78±3.87%
SR [12]	93.4%	77.56±2.96%
NN + SR [12]	92.2%	79.05±4.48%
Cognitec [12]	93.8%	85.62±2.17%
NN + SR + Cognitec [12]	97.6%	93.45±0.96%
C-DFD [15]	92.2%	65.5%
P-RS [14]	-	95.8±6.15% [2]

[1] 96.33% with z-score normalization

[2] 133 subjects for training, 67 subjects for testing

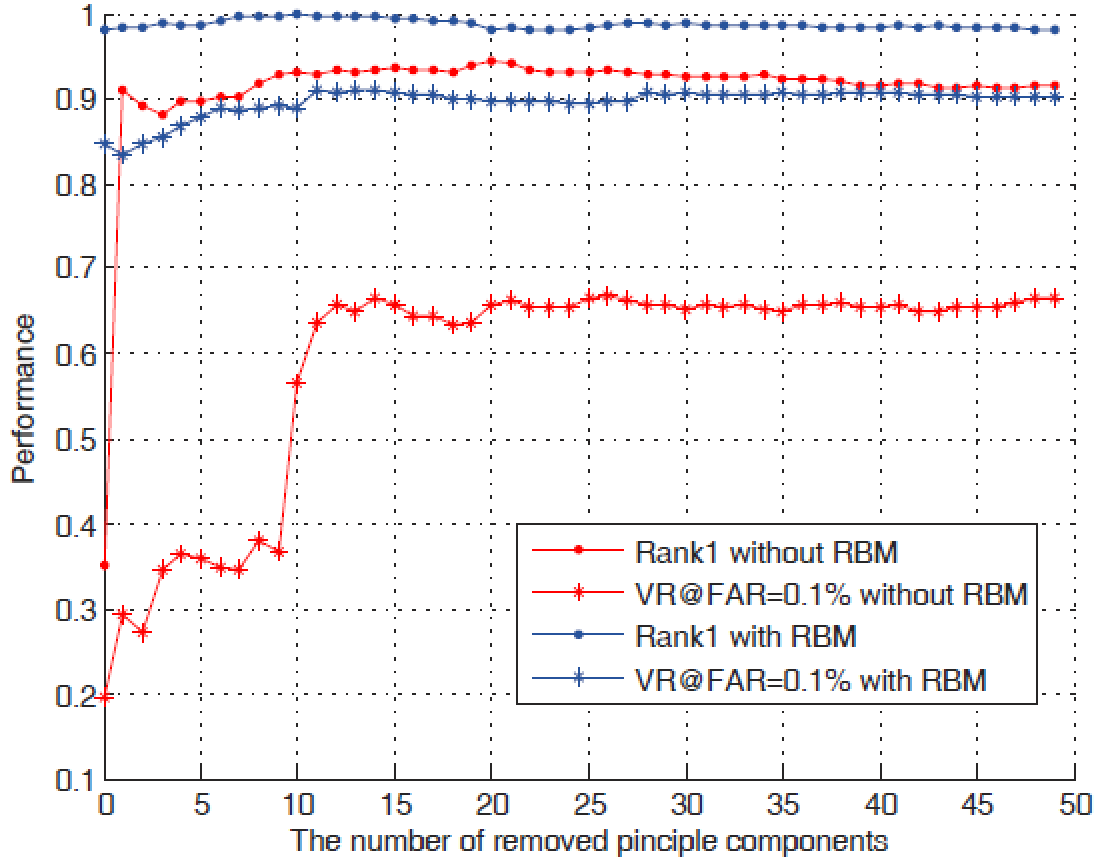
Global, Convolutional and Local RBMs

- Global: fully connected RBM
- Conv: locally connected RBM with shared weights
- Local: locally connected RBM with unshared weights

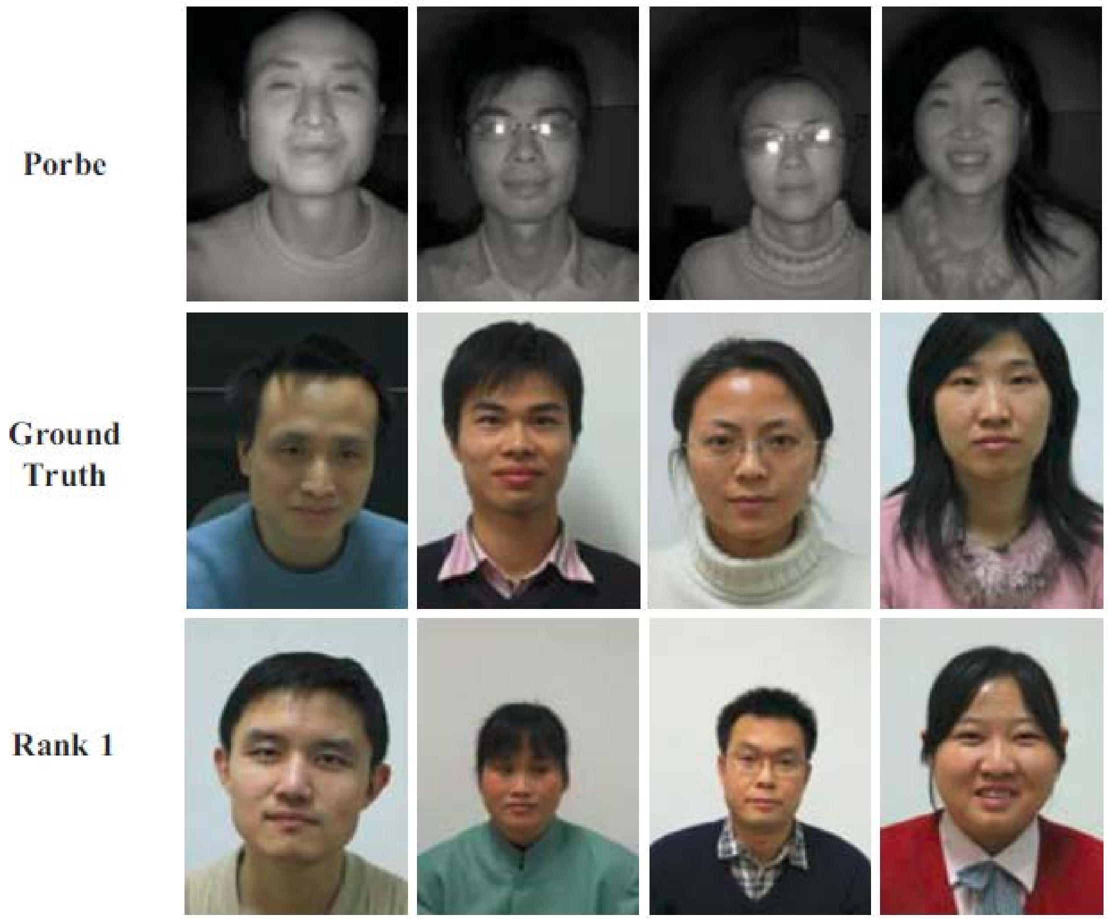
THE COMPARISON OF GLOBAL, CONVOLUTIONAL AND LOCAL RBMS ON VIEW 1 OF CASIA HFB. THE 3RD COLUMN IS VR@FAR=0.1% ON THE TRAINING SET OF VIEW 1. THE 4TH COLUMN IS VR@FAR=0.1% ON THE TESTING SET OF VIEW 1.

	Architecture	VR (Train)	VR (Test)
Global	7040-3520-7040	99.94%	1.549%
Conv.	40-80-40	73.31%	71.79%
Local	176×(40-80-40)	99.45%	90.85%

The Number of Removed PCs



Some Cases of Failure



Results on CASIA NIR-VIS 2.0

RANK1 RECOGNITION RATES AND VR@FAR=0.1% OF VARIOUS METHODS ON VIEW 2 OF CASIA NIR-VIS 2.0.

	Rank1	VR
Gabor	36.18 ± 2.56%	33.37 ± 2.29%
Gabor + Remove 20 PCs	75.54 ± 0.75%	71.40 ± 1.21%
Gabor + RBM	84.22 ± 0.86%	78.39 ± 1.45%
Gabor + RBM + Remove 11 PCs	86.16 ± 0.98%	81.29 ± 1.82%
PCA + Sym + HCA [18]	23.7 ± 1.89%	19.27%
Cognitec [5]	58.56 ± 1.19%	-
DSIFT + LDA [5]	73.28 ± 1.10%	-

Results on CUFS and CUFSF

- The first method can perform perfectly on CUFS dataset
  - Rank1 = 100%
  - VR@FAR=0.1% = 100%
- Perform good on CUFSF dataset
  - Rank1 = 98.59%
  - Rank1 = 98.70%, CITE (Coupled information theoretic encoding, CVPR 2011)

Summary

- Proposed a local to global learning framework for heterogeneous face recognition
- The first time to use multi-modal RBMs to learn shared representations of heterogeneous face images
- By plugging the local RBMs into the framework, we obtained the state-of-the-art results for NIR-VIS and Sketch-VIS problems

Thank you!

Q&A