International Conference on Biometrics (ICB 2016)
June 13-16, 2016 Halmstad, Sweden

Towards Resolution Invariant Face Recognition in Uncontrolled Scenarios

Dan Zeng, Hu Chen, Qijun Zhao

National Key Laboratory of Fundamental Science on Synthetic Vision, College of Computer Science, Sichuan University, China

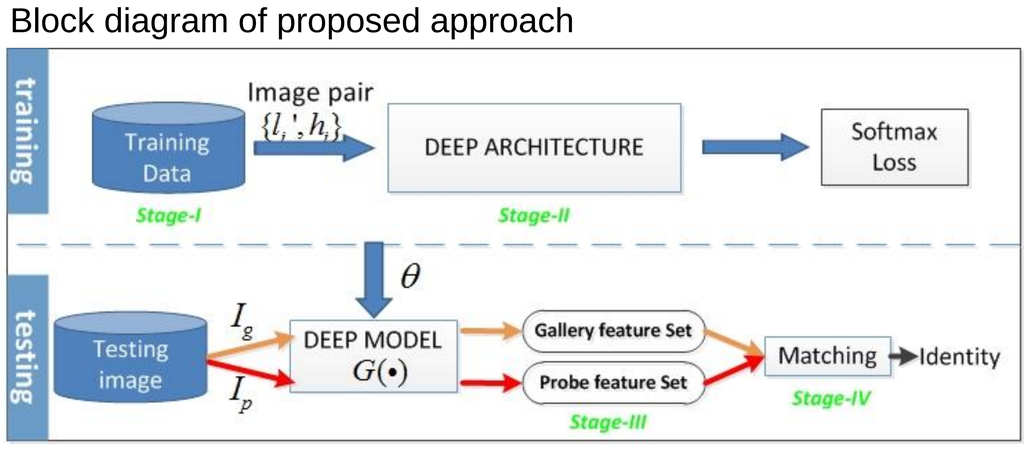
1. Introduction

- Face images captured by surveillance cameras usually have low resolution (LR), whereas the enrolled face images are collected in controlled scenarios with high resolution (HR) and that results in a challenging recognition task.
- Difficulties in cross-resolution face recognition
  - Variations in resolution, uncontrolled poses and illumination conditions, etc.

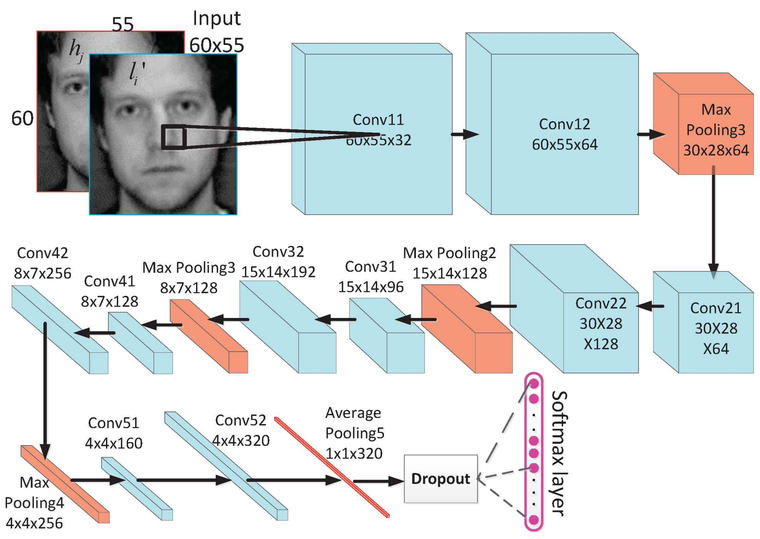
Images (SCface database) are captured in the acquisition distances marked as DC (digital camera), d3 (1.00m), d2 (2.60m) and d1 (4.20m).

2. Proposed Resolution Invariant Approach

- Block diagram of proposed approach



- Stage-I: Training Data Preparation
  - Size of HR images: 60x55, and size of LR images: 30x24.
  - Faces are detected, cropped & normalized to various resolutions, and LR images are up-sampled and combined with HR images.
- Stage-II: Resolution-invariant Deep Network (RIDN)



a) Input: 13,671 classes and 438,139 images;

b) optimization to satisfy the following goals:

- *classify the identity of both LR and HR face images accurately;*
- *preserve the information among subjects across different resolutions;*

- Stage-III: Feature Extraction
  - $G = Conv(x, w, b)$, $G(\bullet)$ is the feature extraction function, x is the input face, w and b denote parameters to be learned from the 1st to Pooling5 layer. The dimension of feature is 320.
- Stage-IV: Matching
  - The distance between a probe (l_i) image and HR gallery (h_j) is computed: $d_{ij} = \mathrm{Cos}\,ine(G(f_{sr}(l_i)), G(h_j))$
  - f_{sr} denote the operation of up-sampling.

3. Database and Experiment Protocol

- Databases & Experiment Protocol

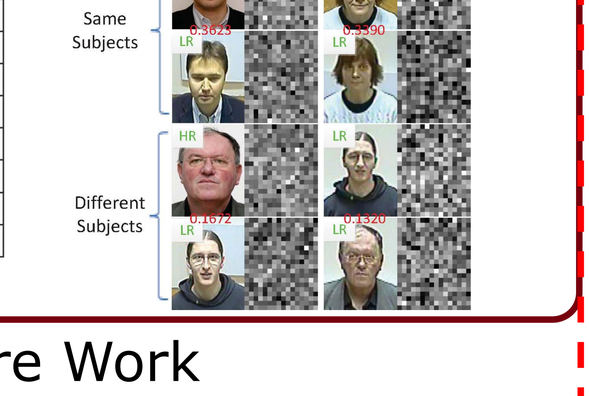
Database	Description	Experiment Protocol
SCface	130 subjects are captured by five surveillance cameras at three distance. Besides, one mugshot image per subject taken by the digital camera is also included. There are totally 2080 images.	1)d1/d2/d3: 4.20m/2.60m/1.00m. 2)DC-di: gallery (DC) – probe (di). 3)di-dj: gallery (distance di) – probe (dj). 4) The following settings are considered: • DC-d3/d2/d1 • d3-d2
COX	1000 subjects, each of which has a HR still image and four uncontrolled scenario.	1) Frames of video2 and video4 are taken as probe images. 2) Images taken under controlled scenario are enrolled as gallery.

4. Evaluation Results

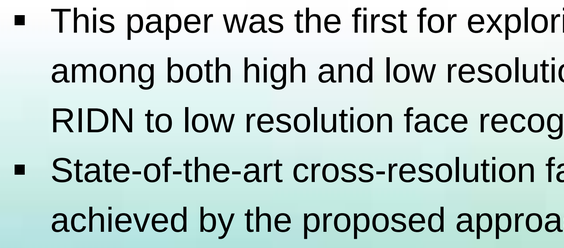
- Results-I on SCface
  - Protocol DC-d3

Method	Rank-1 Accuracy
Reference-based Approach[1]	69.45%
MDS HR-LR	61.14%
LSML	59.25%
GMA-MFA	27.00%
Baseline-HR	49.50%
Baseline-LR	55.00%
Proposed Two Resolutions	**71.25%**
Proposed Four Resolutions	**74.00%**

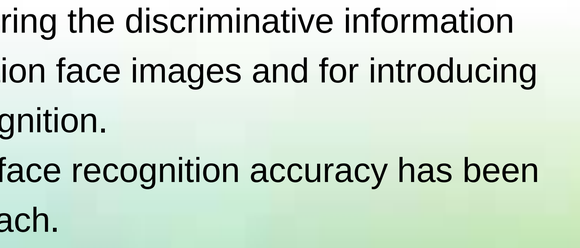
  - Protocol DC-d2/d1



  - Protocol d3-d2



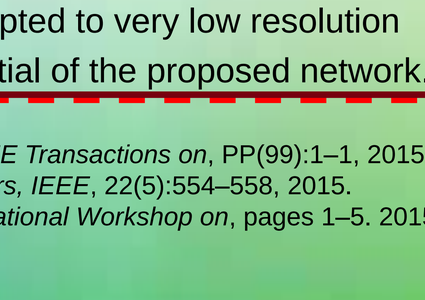
  - Protocol d3-d2



- Results-II on COX

Method	Video2	Video4
PSCL	33.20%	N\A
LERM	42.80%	N\A
CAR	55.00%	28.86%
Baseline-HR	42.21%	42.82%
Baseline-LR	46.87%	44.78%
Proposed Two Resolutions	**57.24%**	**56.49%**
Proposed Four Resolutions	**63.94%**	**63.29%**

- Visualization-III



5. Summary & Future Work

- We have proposed to solve cross-resolution face recognition problem by extracting resolution invariant features from the unified HR and LR training face images through RIDN.
- This paper was the first for exploring the discriminative information among both high and low resolution face images and for introducing RIDN to low resolution face recognition.
- State-of-the-art cross-resolution face recognition accuracy has been achieved by the proposed approach.
- Future work: To explore features adapted to very low resolution problem; To study the extreme potential of the proposed network.

[1] S. Mudunuri and S. Biswas, " Low resolution face recognition across variations in pose and illumination," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PP(99):1–1, 2015.
[2] J. Shi and C. Qi, " From local geometry to global structure: Learning latent subspace for low-resolution face image recognition," *Signal Processing Letters, IEEE*, 22(5):554–558, 2015.
[3] Y. Peng, L. Spreeuwers, and R. Veldhuis, " Likelihood ratio based mixed resolution facial comparison," In *Biometrics and Forensics (IWBF), 2015 International Workshop on*, pages 1–5. 2015.