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 Chengdu, 610065, China
qjzhao@scu.edu.cn

Abstract

Face images captured by surveillance cameras usually have poor quality, particularly low resolution (LR), which affects the performance of face recognition seriously. In this paper, we develop a novel approach to address the problem of matching a LR face image against a gallery of relatively high resolution (HR) face images. Existing methods deal with such cross-resolution face recognition problem either by importing the information of HR images to help synthesize HR images from LR images or by applying the discrimination of HR images to help search for a unified feature space. Instead, we treat the discrimination information of HR and LR face images equally to boost the performance. The proposed approach learns resolution invariant features aiming to: (1) classify the identity of both LR and HR face images accurately, and (2) preserve the discriminative information among subjects across different resolutions. We conduct experiments on databases of uncontrolled scenarios, i.e., SCface and COX, and results show that the proposed approach significantly outperforms state-of-the-art methods.

1. Introduction

Face recognition, as one of the most ubiquitous biometrics technologies, has been extensively studied in recent decades. The report of MBGC[3] demonstrates that current algorithms perform well on good quality images. However, the performance is far from being satisfactory for images captured in uncontrolled scenarios. Recently, the increasing use of surveillance cameras in common areas has naturally motivated the researchers to pay more attention to recognize faces in surveillance videos. Face images captured by surveillance cameras usually have low resolution (LR) in addition to uncontrolled poses and illumination conditions, whereas the enrolled face images are collected in controlled scenarios with high resolution (HR) and that results in a challenging recognition task. Since face region is limited and the signal to noise ratio of images is very low, discriminatory properties presented in the LR images are poor. Moreover, the huge resolution gap between the probe and gallery images naturally makes the problem even more intractable. Since the probe is not of the same domains of the gallery, feature dimensional mismatch problem emerges. The simple approach is down-sampling the

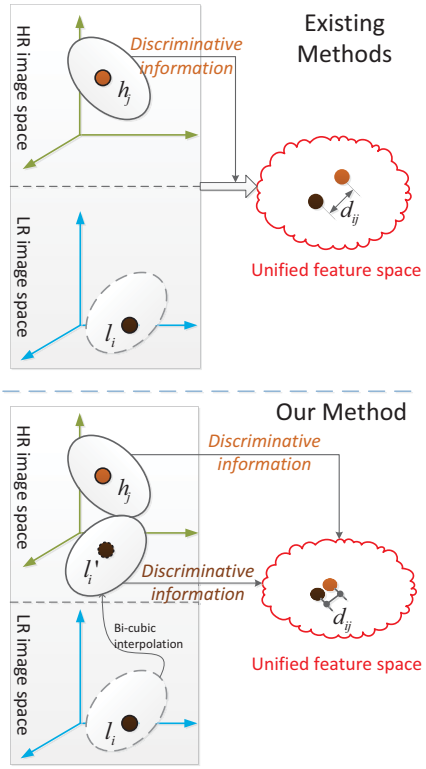

Figure 1. Existing cross-resolution face recognition methods mainly exploit the discriminant information in high resolution (HR) face image space, whereas our proposed method utilizes the discriminant information in both HR and low resolution (LR) face images. As a result, the feature space learned by our proposed method is more robust to resolution variations. h_j and l_i represent HR and LR face images respectively, and l_i' represents upsampled LR face images.

978-1-5090-1869-7/16/$31.00 ©2016 IEEE

gallery HR images and then exploiting the recognition process in the probe LR image space. The computational expense is small, but the resolution gap is eliminated at the cost of losing the discrimination in the gallery. As a result, little attention has been paid to this category.

There are two typical categories [33] to enhance discrimination of LR images and reduce the resolution gap simultaneously. One is super-resolution (SR) approaches, which try to estimate HR images from the LR ones, and hence traditional high resolution face recognition methods could be adopted for recognition. The other is resolution-robust approaches, which devote to build a unified feature space towards resolution variation. Usually, the information of HR face images is used either to help synthesis HR images from LR images, or apply to search for a unified feature space. Obviously, resolution-robust approaches reduce the resolution gap in an implicit way into a unified feature space, whereas SR approaches are searching a same high resolution image space explicitly instead.

It is generally believed that HR images preserve more reliable discriminant information than LR images. Hence the discrimination of LR images, i.e., intra-class and inter-class variations, is usually overlooked by existing cross-resolution face recognition methods. In contrast, this paper proposes a novel method considering the HR and LR images equally, for the sake of the better information utilization (see Fig. 1). To make the comparison achievable, we increase resolution of LR images through interpolation method, i.e., bicubic. It neither suffered high computational cost nor a collection of training image pairs from the same scene. Furthermore, we mix the real HR images with the upsampled HR ones to learn resolution invariant features in a supervised way with deep convolutional network. Finally, we employ cosine distance metric to obtain recognition results. The proposed approach has several advantages: (i) Since resolution invariant features can be learned offline from training data, it is fast and suitable for databases of large size; (ii) we do not ask for a series of image pairs from the same scene for training, and any image is available for training; (iii) it has good generalization ability since no test data is used for training.

The remainder of the paper is organized as follows. In Section II, we review related work and highlight differences in our approach. In Section III, we state the problem. The proposed approach is presented in Section IV. We detail our experimental results and discussions in Section V. In Section VI, we discuss our conclusions.

2. Related Work

Existing works on low resolution face recognition (LR face recognition) can be categorized as SR approaches that recognize face into the HR image space and resolution-based approaches where face images are compared in a unified space.

Super-resolution Approaches have been developed within last decade, which synthesize HR images from the LR images. Nasrollahi K et al. [23] apply a learning based SR algorithm to improve the magnification factor by combining multilayer perceptron with maximum a posteriori. However, the method fails to optimize face images from recognition perspective, and asks for a collection of images from the same scene for training. There are many works [11, 36] integrate visual quality and recognition discrimination. In simultaneous SR and recognition (S^2R^2) [11], the features for recognition are available as prior information to realize SR and recognition simultaneously. The framework is promising, however, the performance is largely affected by the reconstruction process and feature extraction module. More recently, Zou et al. [36] propose discriminative super-resolution (DSR) to solve the very low resolution problem through learning the relationship between training HR images and LR images. Two constraints, data constraint and discriminate constraint are designed to improve the visual quality and recognition discrimination simultaneously. Data constraint estimates the reconstruction error, and discriminate constraint improves the recognition performance. However, the method is time-consuming because it needs online optimization.

Resolution-robust Approaches have gradually attracted attention. Local Frequency descriptor (LFD) [17] presents both blur-robust magnitude and phase information in the low frequency domain to boost the recognition performance. The engineered feature, i.e., LFD, tries to depict the essence attributes of LR images, whereas the majority methods [18, 35, 16, 30, 27, 21] prefer original intensity to search for a unified feature space. Then how to use discrimination of images naturally becomes important. Multi-dimensional scaling (MDS) [6] utilizes discrimination information among HR images as guidance, requiring that the distances between gallery (HR) and probe (LR) images of unified space approximate the distances that probe images captured in the same conditions as the gallery images. An experiment is conducted on the part of the SCface database [9]. However, only the probe images nearly frontal are tested. An improved version [5] applies the automatic facial landmark localization to the LR images to make it more robust to pose variations. Ren at al. [27] propose coupled kernel embedding (CKE) method to preserve the local relationship among images in the HR space and minimize the dissimilarities captured by the kernel Gram matrix in the LR and HR spaces simultaneously .

Recently, based on Ref. [6, 5], Mudunuri and Biswas [22] develops a reference-based approach to address the time-consuming cause by stereo matching, and shows

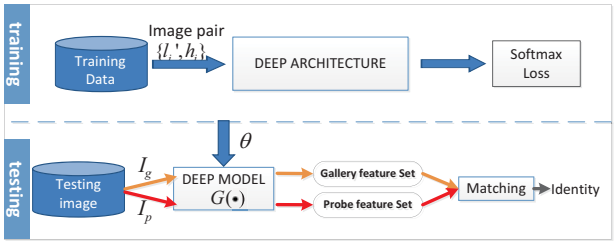

Figure 2. Block diagram of the proposed approach. l_i' and h_i represent upsampled LR face images and HR images, respectively. θ represents the learned parameters. I_g and I_p are gallery and probe face images. $G(\cdot)$ is the deep model based feature extraction function.

state-of-the-art performance. Shi and Qi. [29] applies local optimization for each training sample according to the relationship among images and incorporates local geometry together to build the global structure. The experimental performance on the SCface is better than other methods. Mixed-resolution biometric comparison (MRBC) [24] considers the combined statistics of different resolution images and uses likelihood ratio framework to solve the comparison between images of different resolutions. The result has achieved state-of-the-art performance on the SCface.

The reviewed works contribute to utilize discrimination among HR images to guide a good performance. On the contrary, we consider the discrimination among HR images and LR images at the same time (see Fig. 1). The results show that the added LR discrimination helps obtain a better performance than other methods. We will detail our approach in the following sections.

3. Problem Statement

In the LR face recognition task, the essential problem is to find a proper unified feature space to extract resolution invariant feature. Then we can compute the similarity among samples using common distance metrics, i.e., euclidean, cosine through feature space.

The definition of formulas is given as follows: $l_i \in \mathbb{R}^m$ represents a LR image, $h_j \in \mathbb{R}^M$ represents a HR image, and $i = 1, 2, ...N_l$, $j = 1, 2, ...N_h$, where N_l is the number of LR and N_h is the number of HR images. In this work, N_l is equal to N_h. The distance metric can be computed in the unified feature space by:

$$d_{ij} = D(G(f_{sr}(l_i)), G(h_j)) \quad (1)$$

where $G : \mathbb{R}^M \rightarrow \mathbb{R}^d$ is the function that project the images in the HR image space into the unified feature space, and $f_{sr} : \mathbb{R}^m \rightarrow \mathbb{R}^M$ is a super resolution function that increases the resolution of LR image from the $\mathbb{R}^m$ image space to the $\mathbb{R}^M$ image space. We choose bicubic interpolation method as the function f_{sr} in this work.

As for feature extraction, big data motivates the deep learning-based method to outperform engineered features. Inspired by this, we apply the deep convolutional network to learn resolution invariant features. Since the deep learning neural network is a data driven method, the absence of LR and HR image pairs of the same subject in reality makes the deep neural network paralyzed. Consequently, we generate a HR and a LR images from the same source images by downsampling. In this way, we solve the cross-resolution face recognition problem indirectly (see Fig. 2).

4. Proposed Approach

It remains an unachievable mission to collect a large amount of face images of various resolutions in the real-world for network training, so we choose to imitate kinds of resolution variability of face pairs as an alternate way. In order to do so, we obtain face pairs through downsampling from the source image data by different factors. There are 13671 classes, 438,139 images chosen from available public datasets. Details of the training data will be described in the Section "Training Data Description".

All face images are preprocessed through face detection and face landmarking [32], and aligned by affine transformation through four landmarks, i.e., left eye center, right eye center, nose tip and mouth center. The important entities in this work are listed as follows:

- Size of HR images: 60×55
- Size of LR images: 30×24
- Baseline-HR: network trained on only HR images
- Baseline-LR: network trained on only LR images
- Two Resolutions: mix 60×55 with 30×24
- Four Resolutions: mix 60×55, 30×24, 40×35 with 50×40

'Baseline-HR' and 'Baseline-LR' are used as baseline of performance evaluation using network. 'Two Resolutions' and 'Four Resolutions' are two proposed experimental settings which take two and four resolutions as input to train the network, respectively.

4.1. Resolution-invariant Deep Network

To take discrimination of HR and LR face images equally into consideration, we mix LR images with HR ones of the gray scale as the input of the Network. Moreover, the resolution of LR images is increased to HR images by f_{sr}.

Motivated by Ref. [34], Resolution-invariant Deep Network (RIDN) are trained in a multi-class face recognition task to classify the identity of a face image. Fig. 3 shows the detailed architecture of the Network which takes $60 \times 55 \times 1$

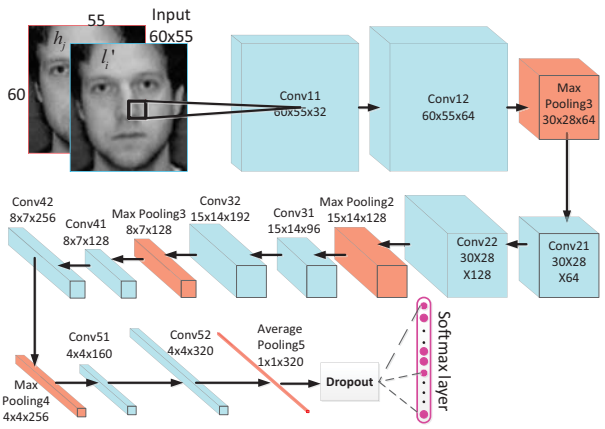

Figure 3. Structure of the proposed Resolution-Invariant Deep Network (RIDN) for resolution-robust feature extraction.

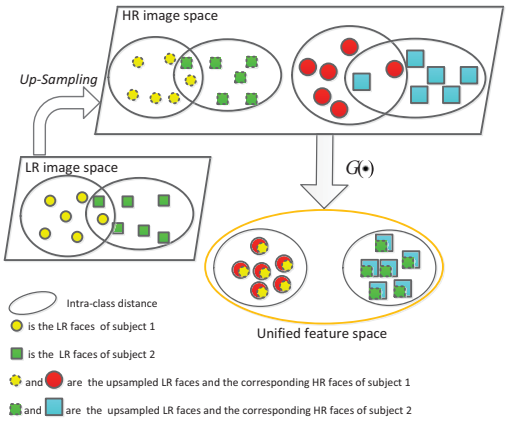

Figure 4. Basic idea of the proposed method. LR images are upsampled to HR image space, and then used together with HR images to learn a unified feature space. This makes the obtained feature space robust to resolution variations.

input and predicts n (e.g.,n = 13,671) identity classes. The dimension of feature extracted from the Pooling5 layer is fixed to 320. As Fig. 3 exhibits, the Network contains ten convolutional layers, two of which are bundled in a group, then Max Pooling layers followed every group except the last one is followed by a Average Pooling layer, followed by Fully Connect layer and Softmax layer indicating identity classes.

The output of RIDN is an $n-way$ softmax layer, which outputs a probability distribution over the n classes:

$$y_p = \frac{exp(y'_p)}{\sum_{k=1}^{n} exp(y'_k)} \tag{2}$$

where $y'_k = \sum_{k=1}^{320} x_p \cdot w_{k,p} + b_k$ combines the 320 dimensional feature linearly as the input of neuron k, its output is y_k. The network is trained to minimize the cross-entropy loss, denoted as:

$$L = -\sum_{y=1}^{n} -y_p \log \hat{y_p} = -\log \hat{y_t} \tag{3}$$

where y_p is the target probability distribution, $y_p = 0$ for all p except $y_p = 1$ for the target class t. $\hat{y_p}$ is the predicted probability distribution. The loss is minimized over the parameters by computing the gradient of L, and stochastic gradient descent (SGD) is used in back-propagation. To accurate classify classes, the Network layer form discriminative identity-related features (i.e. features with large interpersonal variations).

4.2. Feature Extraction

In RIDN, Deep Model including w and b related to each layer is learned through training phase. Feature extraction process is denoted as $G = Conv(x, w, b)$, where $G(\cdot)$ is the feature extraction function, x is the input face, w and b denote parameters to be learned from the first layer to Pooling5 layer (see Fig. 2).

Actually, mixing LR images with HR images to train Deep Model is just like the process searching for a unified feature space. Intuitively, the parameters θ are optimized to satisfy the following goals: (1) LR face images should be classified accurately; (2) HR face images should be classified accurately; (3) images pairs of different resolutions should be classified by their identities; All three constraints are realized in RIDN. The conceptual description of the optimization process is exhibited in Fig. 4.

4.3. Matching

The test images are sampled to 60×55 by f_{sr} after alignment. During the matching, the features of the both the LR probe and HR gallery images are obtained through feature extraction function $G(\cdot)$. If $G(f_{sr(l_i)})$ and $G(h_j)$ denote the features corresponding to an LR probe and an HR gallery image, the distance between them is computed as the cosine distance between their features as follows:

$$d_{ij} = Cosine(G(f_{sr}(l_i)), G(h_j)) \tag{4}$$

Since the $G(\cdot)$ can be learned offline training, so the proposed method is fast and suitable for databases of large size.

5. Experimental Evaluation

In this section, we first introduce the training data, then report the results on SCface and COX databases to verify the efficacy of the proposed approach.

To validate how much the results are influenced by deep learning, 'Baseline-HR' and 'Baseline-LR' are regarded as the baseline of RIDN network which takes only HR or LR images as input. Moreover, to verify whether more different

Figure 5. Example face images in the CASIA-WebFace database. Each row corresponds to one person.

Table 1. Summary of public databases used for training in this study.

Database	# Subjects	# Images
CASIA-WebFace [34]	10069	292004
FERET [26]	1195	3285
CAS-PEAR-R1 [7]	1040	8510
FRGC v2 [25]	466	15828
Multi-PIE [10]	337	111281
MUCT [20]	176	2095
Faces94 [1]	153	3033
AR [19]	100	586
PIE [31]	68	333
ORL [4]	40	389
Pointing04 [8]	15	611
Grimace [2]	12	184

resolutions of images for training bring about better face recognition performance, we add images with another two resolutions, i.e., 50×40 and 40×35, for training, namely "Four Resolutions".

5.1. Training Data

We gather 12 public face databases to construct the training data. Table.1 lists these face databases.

Since the pose problem is not the main concern factor, the images whose poses are less than 30° in the yaw orientation and 15° in the pitch orientation are chosen. As Fig. 5 shows, face images suffered from illumination variations, expression variations etc.. Besides, the imbalance of the number per class on training database is striking, ranging from 2 images to 400+ images. All this makes it challenging to construct training data.

5.2. Results on SCface

To our best knowledge, many works related to LR face recognition test on Multi-PIE [10]. However, lacking of real

(a)

(b)

Figure 6. Samples in (a) SCface and (b) COX databases. In SCface, images of each row are captured in the same acquisition distances marked as DC (digital camera), $d3$ (distance3), $d2$ (distance2) and $d1$ (distance1).

LR images in Multi-PIE, so the LR images are downsampled from HR images for face recognition. Since there is a big difference between the simulation and real surveillance environment, the methods can not be applied to the real-world.

We test the efficacy of our method on the SCface database which contains face images of 130 subjects taken in uncontrolled indoor environment and some subjects are exhibited in Fig. 6(a). The face images are captured by five surveillance cameras at three distances, $4.20m$ ($d1$), $2.60m$ ($d2$) and $1.00m$ ($d3$), and one frontal mugshot image per

Table 2. Rank-1 recognition accuracy of the proposed approach and some existing algorithms on SCface database.

Method	Rank-1 Accuracy
Reference-based Approach[22]	69.45%
MDS HR-LR[5]	61.14%
LSML[16]	59.25%
GMA-MFA[28]	27.00%
Baseline-HR	49.50%
Baseline-LR	55.00%
Proposed Two Resolutions	**71.25%**
Proposed Four Resolutions	**74.00%**

subject taken from digital camera (DC) is also included. The cameras are placed slightly above the subject's head, and the humans are not required to look at a fixed point during the recordings, thus the collected images are blurred. Moreover, pose and lighting variations as well as poor quality are different according to different cameras and different distances. Face images captured at $4.20m$ are of the most poor quality compared with the $2.60m$ and $1.00m$.

We use some acronyms to denote the acquisition distances in the experiments as follows. (1)$d1/d2/d3$: The images are captured by surveillance camera at a distance of $4.2m/2.6m/1.0m$; (2) $DC - di$: The images captured by digital camera are enrolled as gallery, and those captured at di distance by surveillance camera are used as probe; (3)$di - dj$: The images captured by di distance are enrolled as gallery, and those captured at dj distance are used as probe.

We conduct extensive experiments with different combinations of these images to evaluate the performance of our proposed method under different settings and compare it with state-of-the-art methods of the same condition. (i) Our experiment stays the same with the experimental setting in Ref. [22] which considers $DC-d3$. There are 80 subjects at distance3 randomly picked for testing (a total of 400 probe images). The recognition accuracy of our method and other methods on SCface are shown in Table.2.

(ii) We adopt $d3-d2$ to estimate LR face recognition performance followed the experimental setting in Ref. [29]. 50 subjects at distance2 are used as probe images, and subjects for distance3 are enrolled in the gallery. The comparison of different methods are presented in Fig. 7.

(iii) Experimental setting $d3 - d2$ is used as a testing protocol. The comparative results are shown in Fig. 8.

(iv) To prove the high generalization of the proposed approach, all data in SCface dataset are used for testing. We conduct experiment on both $DC - d2$ and $DC - d1$. The two results are reported in Fig. 9. Since no works report recognition performance under these experimental settings, no comparison results are shown. The recognition accuracy

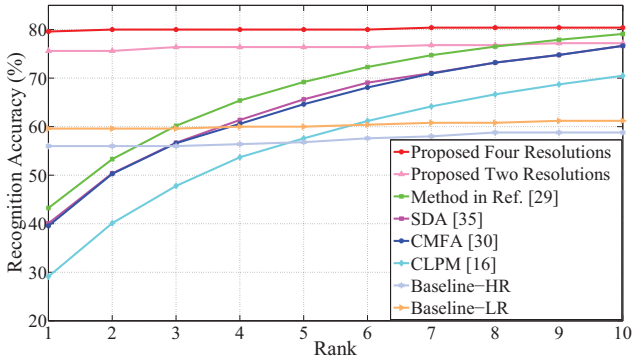

Figure 7. Cumulative Matching Characteristic (CMC) curves of different methods on the SCface database.

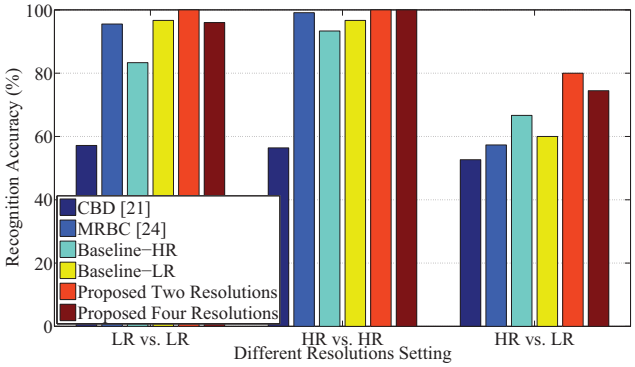

Figure 8. Comparison of the proposed approach to MRBC [24] and CBD [21] in terms of Rank-1 recognition accuracy on the SCface database.

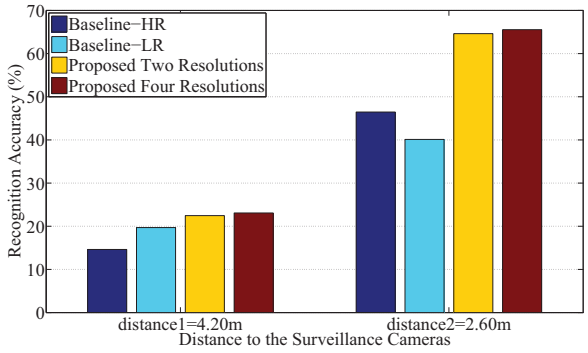

Figure 9. Rank-1 recognition accuracy of the proposed approach on the SCface database when all data in the database are used for testing.

of images at distance2 is close to 70%. Since quality of face images captured at distance1 is more poor than distance2, the recognition performance between two distances is vary significantly (see Fig. 9).

From the above, the proposed approach is significantly more stable and has a higher recognition performance than

Table 3. Rank-1 recognition accuracy of the proposed approach and state-of-the-art method on COX database. "N\A' means the result is not available.

Method	Video2	Video4
PSCL [12]	33.20%	N\A
LERM [14]	42.80%	N\A
CAR [15]	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%**

other methods.

5.3. Results on COX

The COX [13] consists of 1,000 subjects, each of which has a HR still image and four uncontrolled LR video clips. Each video clip generally contains 25 frames. Still images are taken under controlled environment and video clips are captured by two different camcorders at two different distances, and have variations on resolution, head pose, illumination and expression (see Fig. 6(b)). The dataset is frequently used in Video-based face recognition, since it is a favourable simulation to video surveillance scenarios.

In order to verify the usefulness of our approach we regard each frame as a probe image and compare performance with state-of-the-art video-based face recognition method. Our testing protocol seems more strict than other referred methods, because they follows the testing protocol in Ref. [13] that using 300 subjects for training and the remaining 700 subjects for testing, but we test on 1000 subjects of the whole database without further training for testing. For simplicity, we perform experiment on video2 and video4 as a representative. The result is shown in Table.3.

5.4. Visualization

In order to further comprehend the proposed approach, we visualize the resolution invariant features of face images. As Fig.10 shows, the proposed approach can not only learn the discrimination among images of different resolutions, but also learn the discriminant information among low resolution face images.

6. Conclusion

The results obtained by the proposed approach are promising. It supports the proposed hypothesis that considering information of LR is necessary. Compared with state-of-the-art methods, our method obtains a better result, and the probable reason is that discrimination of LR images help realize three constraints to learn resolution invariant

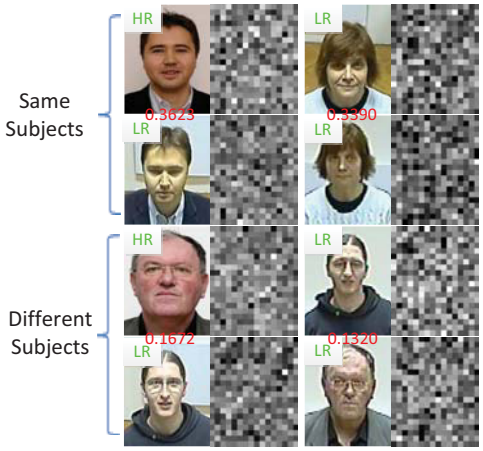

Figure 10. Visualization of the learned 320-dimensional resolution-robust feature for example face images. We rearrange features as 20×16. The cosine similarities of these image pairs are also shown. 'HR' and 'LR' marked on the images indicate if the images are high resolution or low resolution.

feature. According to performance curves, the accuracy of 'Baseline-LR' is better than that of 'Baseline-HR', which seemingly indicates extracted LR features have more effects on cross-resolution recognition than HR features do.

In this paper, we have proposed to solve the cross-resolution face recognition problem by extracting resolution invariant features from the unified high resolution and low resolution training face images via Resolution-Invariant Deep Network (RIDN). To the best knowledge of us, this work is the first (i) for exploring discriminative information among both high and low resolution face images, and (ii) for introducing RIDN to low resolution face recognition.

7. Acknowledgement

This work has been supported by the National Natural Science Foundation of China (NO.61202161) and National Key Scientific Instrument and Equipment Development Projects of China (NO. 2013YQ49087904).

References

[1] http://cswww.essex.ac.uk/mv/allfaces/faces94.html.
[2] http://cswww.essex.ac.uk/mv/allfaces/grimace.zip.
[3] http://face.nist.gov/mbgc/.
[4] http://www.cam-orl.co.uk/.
[5] S. Biswas, G. Aggarwal, P. J. Flynn, and K. W. Bowyer. Pose-robust recognition of low-resolution face images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(12):3037–3049, 2013.
[6] S. Biswas, K. W. Bowyer, and P. J. Flynn. Multidimensional scaling for matching low-resolution face images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(10):2019–2030, 2012.

[7] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The cas-peal large-scale chinese face database and baseline evaluations. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 38(1):149–161, 2008.

[8] N. Gourier and J. Letessier. The pointing 04 data sets. In *Proceedings of Pointing 2004, ICPR International Workshop on Visual Observation of Deictic Gestures*, pages 1–4, 2004.

[9] M. Grgic, K. Delac, and S. Grgic. Scface–surveillance cameras face database. *Multimedia tools and applications*, 51(3):863–879, 2011.

[10] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010.

[11] P. H. Hennings-Yeomans, S. Baker, and B. Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In *Computer Vision and Pattern Recognition, 2008 IEEE Conference on*, pages 1–8. IEEE, 2008.

[12] Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, and X. Chen. A benchmark and comparative study of video-based face recognition on cox face database. *Image Processing, IEEE Transactions on*, 24(12):5967–5981, 2015.

[13] Z. Huang, S. Shan, H. Zhang, S. Lao, A. Kuerban, and X. Chen. Benchmarking still-to-video face recognition via partial and local linear discriminant analysis on cox-s2v dataset. In *Computer Vision–ACCV 2012*, pages 589–600. Springer, 2013.

[14] Z. Huang, R. Wang, S. Shan, and X. Chen. Learning euclidean-to-riemannian metric for point-to-set classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1677–1684, 2014.

[15] Z. Huang, X. Zhao, S. Shan, R. Wang, and X. Chen. Coupling alignments with recognition for still-to-video face recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3296–3303, 2013.

[16] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2288–2295. IEEE, 2012.

[17] Z. Lei, T. Ahonen, M. Pietikäinen, and S. Z. Li. Local frequency descriptor for low-resolution face recognition. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 161–166. IEEE, 2011.

[18] B. Li, H. Chang, S. Shan, and X. Chen. Low-resolution face recognition via coupled locality preserving mappings. *Signal Processing Letters, IEEE*, 17(1):20–23, 2010.

[19] A. Martınez and R. Benavente. The ar face database. *Rapport technique*, 24, 1998.

[20] S. Milborrow, J. Morkel, and F. Nicolls. The muct landmarked face database. *Pattern Recognition Association of South Africa*, 201(0), 2010.

[21] P. Moutafis, I. Kakadiaris, et al. Semi-coupled basis and distance metric learning for cross-domain matching: Application to low-resolution face recognition. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, pages 1–8. IEEE, 2014.

[22] 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.

[23] K. Nasrollahi and T. B. Moeslund. Extracting a good quality frontal face image from a low-resolution video sequence. *Circuits and Systems for Video Technology, IEEE Transactions on*, 21(10):1353–1362, 2011.

[24] 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. IEEE, 2015.

[25] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Computer vision and pattern recognition (CVPR), 2005 IEEE computer society conference on*, volume 1, pages 947–954. IEEE, 2005.

[26] P. J. Phillips, H. Moon, S. Rizvi, P. J. Rauss, et al. The feret evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090–1104, 2000.

[27] C.-X. Ren, D.-Q. Dai, and H. Yan. Coupled kernel embedding for low-resolution face image recognition. *Image Processing, IEEE Transactions on*, 21(8):3770–3783, 2012.

[28] A. Sharma, A. Kumar, H. Daume III, and D. W. Jacobs. Generalized multiview analysis: A discriminative latent space. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2160–2167. IEEE, 2012.

[29] 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.

[30] S. Siena, V. N. Boddeti, and B. V. Kumar. Coupled marginal fisher analysis for low-resolution face recognition. In *Computer Vision–ECCV 2012. Workshops and Demonstrations*, pages 240–249. Springer, 2012.

[31] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Automatic Face & Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 46–51. IEEE, 2002.

[32] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3476–3483. IEEE, 2013.

[33] Z. Wang, Z. Miao, Q. J. Wu, Y. Wan, and Z. Tang. Low-resolution face recognition: a review. *The Visual Computer*, 30(4):359–386, 2014.

[34] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[35] C. Zhou, Z. Zhang, D. Yi, Z. Lei, and S. Z. Li. Low-resolution face recognition via simultaneous discriminant analysis. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–6. IEEE, 2011.

[36] W. W. Zou and P. C. Yuen. Very low resolution face recognition problem. *Image Processing, IEEE Transactions on*, 21(1):327–340, 2012.