Comparison of Super-Resolution Benefits for Downsampled Images and Real Low-Resolution Data

Yuxi Peng   Luuk Spreeuwers   Berk Gökberk   Raymond Veldhuis

University of Twente
Faculty of EEMCS
P.O. Box 217, 7500 AE Enschede, The Netherlands

y.peng@utwente.nl

Abstract

Recently, more and more researchers are exploring the benefits of super-resolution methods on low-resolution face recognition. However, often results presented are obtained on downsampled high-resolution face images. Because downsampled images are different from real images taken at low resolution, it is important to include real surveillance data. In this paper, we investigate the difference between downsampled images and real surveillance data in two aspects: (1) the influence of resolution on face recognition accuracy, and (2) the improvement of accuracy that can be achieved by super-resolution on these images. Specifically, we will test the following hypotheses: (1) face recognition performance on real images is much worse than on downsampled images, and (2) super-resolution improves the performance of downsampled images more than real images.

Our experiments are conducted using videos from the HumanID database. In each video, the target person's face is captured while he is walking towards the surveillance camera. We detect the faces in the video frames using a Viola-Jones face detector. Then we select face images of four different resolutions: two low-resolution and two high-resolution. The high-resolution images are used for gallery and generating downsampled images. We perform two types of face recognition experiments. In the first type of experiments, three face recognition methods are evaluated for images with different resolution. The three methods are (1) Principal Component Analysis, (2) Linear Discriminant Analysis, and (3) Local Binary Patterns. In the second type, we apply two super-resolution methods: (1) a model based method and (2) a feature based method on the low-resolution (both real and downsampled) images and then compute the face recognition accuracy.

1 Introduction

Face recognition at a distance is a challenging topic in the face recognition domain. The face images captured at a distance usually are small and of low quality, thus, they are not suitable for most of the face recognition systems which are developed for high quality images. To improve the face recognition performance of these low quality images, one way is to use super-resolution (SR) to enhance the resolution. Some SR methods have been developed specially for face recognition.

In [5], a model for SR is built based on Tikhonov regularization and a linear feature extraction stage. It includes the face features which would be extracted for face

recognition as a prior information. In [6], canonical correlation analysis is used to project high-resolution (HR) and low-resolution (LR) image pairs to a coherent feature space and then radial based functions is applied to find the mapping between them. A data constraint is developed in [14] to minimize both the distances between the constructed SR images and the corresponding HR images and the distances between SR images from the same class. In [3], multidimensional scaling is used to transform the HR gallery and LR probe images to a common space so that the distance between them approximate the distance between their corresponding HR images. Another SR method based on morphable model is proposed in [13].

The LR images used in most SR papers are downsampled from HR images. Recently, a few researchers also use data from real surveillance cameras. In [14] and [3], the proposed methods are also evaluated using images from the SCface database [4] which contains surveillance quality facial images captured at three different distances. In [7], videos are captured and enhanced with SR methods for face recognition. A real-world outdoor video dataset captured by a PTZ camera is used in [12].

In this paper, we investigate the difference between downsampled and real LR images for face recognition. The experiments are conducted using images from the HumanID database [8]. This database contains videos in which the target person's face is captured while he/she is walking towards the camera. We select images with four different resolutions from the videos and we also generate LR images by downsampling. Then we conduct both standard face recognition and super-resolution experiments. Three face recognition algorithms (PCA [10] , LDA [2] and LBP [1]) and two SR methods (RL/DSR method [14] and NMCF method [6]) are investigated.

The remainder of this paper is organized as follows. In Section 2, hypotheses are proposed regarding the difference between downsampled and real LR images. Section 3 gives a brief introduction to the two SR methods that are used in our experiments. Experimental setup and verification test results are presented in Section 4. Section 5 concludes the paper.

2   Downsampled vs. real low-resolution image

We expect downsampled images have better performance for face recognition than real LR images intuitively. One of the main reasons is that the registration is much poorer for real LR images. Finding the landmarks on HR faces is relatively easy. Downsampled images can make use of these landmarks for registration. But for LR images, the landmarks usually cannot be located precisely. Thus, we make following hypotheses:

(1) Face recognition performance on real LR images is much worse than on downsampled images;

(2) Super-resolution improves the performance of downsampled images more than real LR images.

An additional hypothesis is made for the training approach. Three different training configurations are possible. In the first configuration, the images from the training set are downsampled to the same resolution as the probe images. The second one upsamples the images of the probe set to the same resolution as the training images. In the third configuration, images from the probe set are also upsampled, but the training images are first downsampled to the same resolution as the probe images and then upsampled to the original resolution. We would expect the first and third approaches to perform better than the second one.

3 Super-resolution methods

3.1 RL and DSR method

In [14], Zou and Yuen proposed a SR method for very low-resolution images which is compatible with various face recognition methods. They introduce a data constraint which clusters the constructed SR images with the corresponding HR images.

Given a set of HR and LR image pairs ($\{I_i^h, I_i^l\}_{i=1}^N$), the relation R is modeled as

$$R = \arg\min_{R'} \sum_{i=1}^N \left\| I_i^h - R'I_i^l \right\|^2. \tag{1}$$

This is called relationship learning (RL). It minimize the distance between HR and the space of LR projected by R.

To acquire better results in face recognition, identity information about the subject is used based on RL. This is called discriminative super-resolution (DSR). A second term is added to (1), see (2).

$$R = \arg\min_{R'} \frac{1}{N} \sum_{i=1}^N \left\| I_i^h - R'I_i^l \right\|^2 + \gamma d(R'). \tag{2}$$

where γ is a constant to balance the two terms. We set γ to 1 in our experiments. $d(R')$ is represented as

$$d(R') = \frac{1}{N(\lambda_i = \lambda_j)} \sum_{\lambda_i = \lambda_j} \left\| I_i^h - R'I_j^l \right\|^2 - \frac{1}{N(\lambda_i \neq \lambda_j)} \sum_{\lambda_i \neq \lambda_j} \left\| I_i^h - R'I_j^l \right\|^2 \tag{3}$$

where λ_i is the class label of I_i. This makes sure the reconstructed HR images are clustered with the images from the same class and far away from those from other classes.

In the testing phase, for both RL and DSR, we first apply $I_{SR} = RI_{input}$ to a given LR image I_{input} to obtain SR image I_{SR}, and then use I_{SR} for face recognition.

3.2 NMCF method

In [6], Huang and He proposed a SR method where canonical correlation analysis (CCA) is used to project the PCA features of HR and LR image pairs to a coherent feature space. Radial based functions (RBFs) are then applied to find the mapping between the HR/LR pairs. This method finds nonlinear mappings on coherent features. Thus, we will refer to this method as NMCF method in this paper.

In the training stage, firstly PCA features of HR and LR image pairs are extracted to reduce computational costs. Next, these PCA features are projected to a coherent feature space by CCA where the correlation between HR and LR features is maximum. This provides better condition for finding the mappings in the next step. Let $\hat{X}^H$ and $\hat{X}^L$ be the PCA features of HR and LR subtracted by the mean, and C_{11} and C_{22} be the within-set covariance matrices, and C_{12} and C_{21} be the between-set covariance matrices. Compute $R_1 = C_{11}^{-1}C_{12}C_{22}^{-1}C_{21}$ and $R_2 = C_{22}^{-1}C_{21}C_{11}^{-1}C_{12}$. The coherent features of HR and LR images are

$$C^H = (V^H)^T \hat{X}^H, C^L = (V^L)^T \hat{X}^L. \tag{4}$$

V^H and V^L comprises eigenvectors of R_1 and R_2 when their corresponding eigenvalues are sorted in descending order.

Then RBFs are applied to approximate the mapping between HR and LR coherent features. The function approximation is represented as $C^H = W\Phi$ where W is a weighting coefficient matrix and Φ is a multiquadratic basis function (see [6] for details). Thus, W can be solved by $W = C^H(\Phi + \tau I_{id})^{-1}$. τ is a small positive value and I_{id} is the identity matrix.

In the testing stage, for a given LR probe image I^p, first compute the coherent features c^p and then apply the learnt mapping to obtain the SR features by

$$c_{SR} = W \cdot [\varphi(\|c_1^l - c^p\|) \ldots \varphi(\|c_N^l - c^p\|)]^T \tag{5}$$

where $\varphi(\|c_i - c_j\|) = \sqrt{\|c_i - c_j\|^2 + 1}$. The SR features are fed to the nearest neighbor classifier together with the coherent features of HR gallery images for recognition.

4 Experimental results

Our experiments are conducted on the HumanID database [8]. We use the videos in which the target person's face is captured while he is walking towards the surveillance camera, see Figure 1. The faces from the videos are detected by the Viola-Jones face detector [11]. For each video, we select four face images with different resolution: 70×70, 50×50, 30×30 and 25×25. Thus, we have about 400 images for each resolution. The images with resolution 70×70 are used as gallery images. The images with the remaining three resolutions are used as probe sets. We also generate images with different resolutions by downsampling images of resolution 50×50. Images with six different resolutions are generated: 30×30, 25×25, 20×20, 15×15, 10×10 and 5×5. To train the face recognition classifier and the super-resolution system, we use downsampled images from the FRGC database [9]. We use the eye coordinates for registration of the face images. The FRGC database provides eye coordinates of its images while we manually click on the eyes for the images from the HumanID database.

Figure 1: Sample image frames from the HumanID database.

4.1 Face recognition experiments

In this section we provide the face recognition performance of PCA, LDA and LBP on both downsampled and real LR face images. Distance measures employed in 1-nearest neighbor classifier for PCA, LDA and LBP are L1 norm, cosine angle and Chi square, respectively. Three different training configurations are applied for our face recognition experiments:

Config. 1: the images of the training set are downsampled to the same resolution of

the probe images;

Config. 2: the training images are with resolution 70×70, and the probe images are upsampled to resolution 70×70 using bicubic interpolation.

Config. 3: the images of the training set are first downsampled to the same resolution of the probe images, then upsampled back to the resolution of 70×70 using bicubic interpolation. The probe images are also upsampled to the resolution of 70×70.

Config. 2 and 3 are also the baseline configurations for SR. Because LBP does not require a training set, we only apply Config. 1 and 2 for LBP.
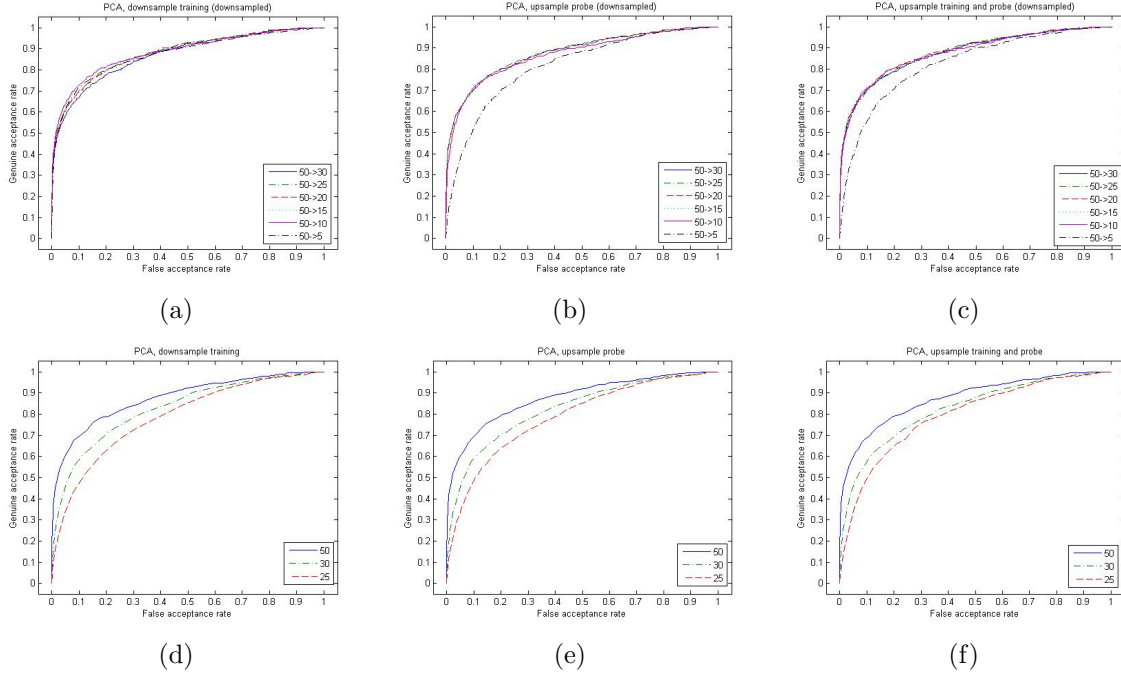
Figure 2: ROC curves of PCA results, downsampled vs. real: (a) downsampled, Config. 1; (b) downsampled, Config. 2; (c) downsampled, Config. 3; (d) real, Config. 1; (e) real, Config. 2; (f) real, Config. 3.
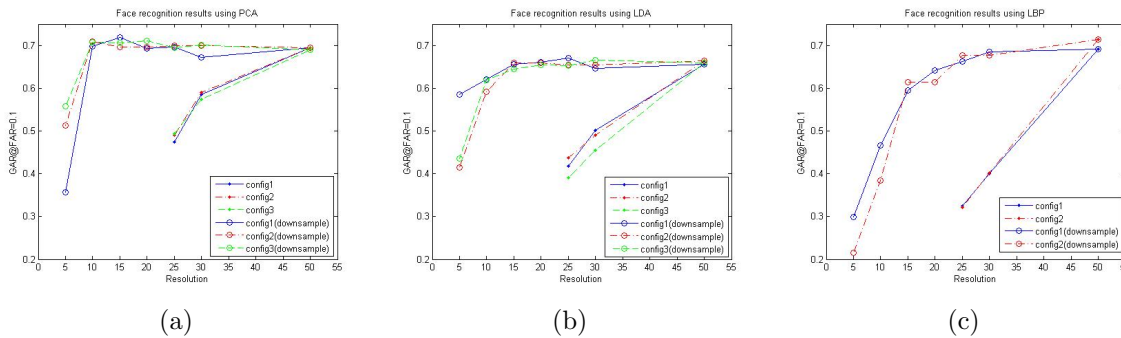

Figure 3: Face recognition results, GAR@FAR=0.1: (a) PCA, (b) LDA, (c) LBP.

We conducted verification experiments. The ROC curves of PCA is shown in Figure 2. We also compute the genuine acceptance rates (GAR) when false acceptance rates (FAR) equal to 0.1 for PCA, LDA and LBP, see Figure 3.

As we can see, resolution changes do not have much influence for downsampled images, but for real LR images, the recognition results decrease significantly as the

resolution becomes lower. At resolution 30×30 and 25×25, the GARs of downsampled images are almost the same as for resolution 50×50, but the GARs of real LR images are much lower. The performance of different face recognition classifiers are of similar trend while LBP is more sensitive to resolution changes than PCA and LDA. Different training configurations influence little when the image resolution is higher than 10×10.

4.2 Super-resolution experiments

In this section, we apply the SR methods that were introduced in Section 3 to explore the benefits of SR. Images from the FRGC database are used to train the SR system. The training configurations are similar to Config. 2 and 3, but RL and DSR methods are used for upsampling LR images instead of bicubic interpolation. The NMCF method is only designed for Config. 2. The ROC curves of RL and DSR methods with PCA in Config. 2 are shown in Figure 4. The GARs when FARs equal to 0.1 are presented in Figure 5.
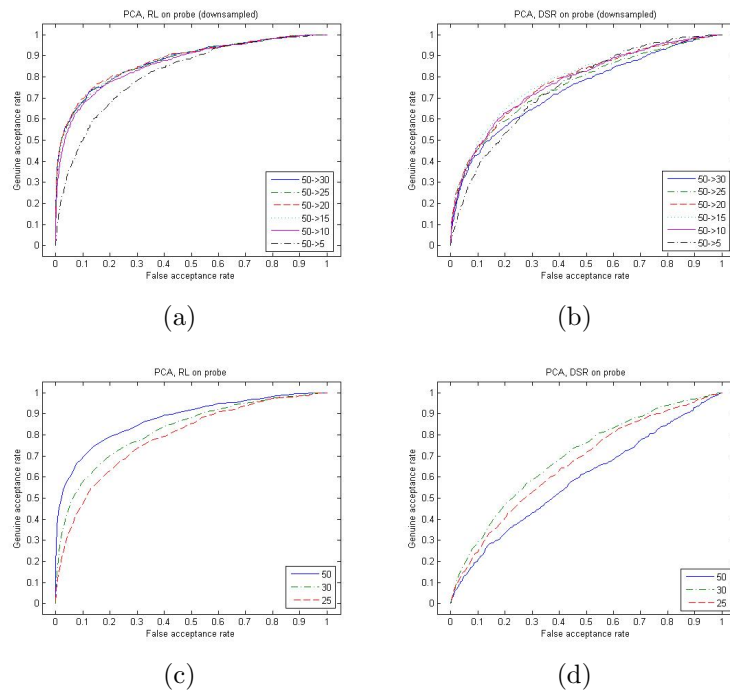

Figure 4: ROC curves of SR results using RL/DSR methods with PCA, Config. 2: (a) dowsampled, RL; (b) dowsampled, DSR; (c) real, RL; (d) real, DSR.

As we can see from these results, all of the SR methods tested in the experiments have no benefits for verification. LBP results drop dramatically after SR. For PCA and LDA, the RL and NMCF methods keep the verification performance at the same level of bicubic interpolation. But the DSR method makes the results worse. The two training configurations also give similar results. To better explain these results, we shown some reconstructed SR images in Figure 6.

Firstly, compare the real LR images with downsampled images at the same resolution, it is easy to find that the downsampled images are clearer and contain more details than the real images. This explains why downsampled images have better performance in LR face recognition. Secondly, the SR images show more details about the faces but also add some artifacts on them. These artifacts are also a problem for face recognition and it may cause the failure of LBP for recognition. Thirdly, as the resolution become lower, the identity information in the images become less, so the SR images become more look like an average face other than the person himself.
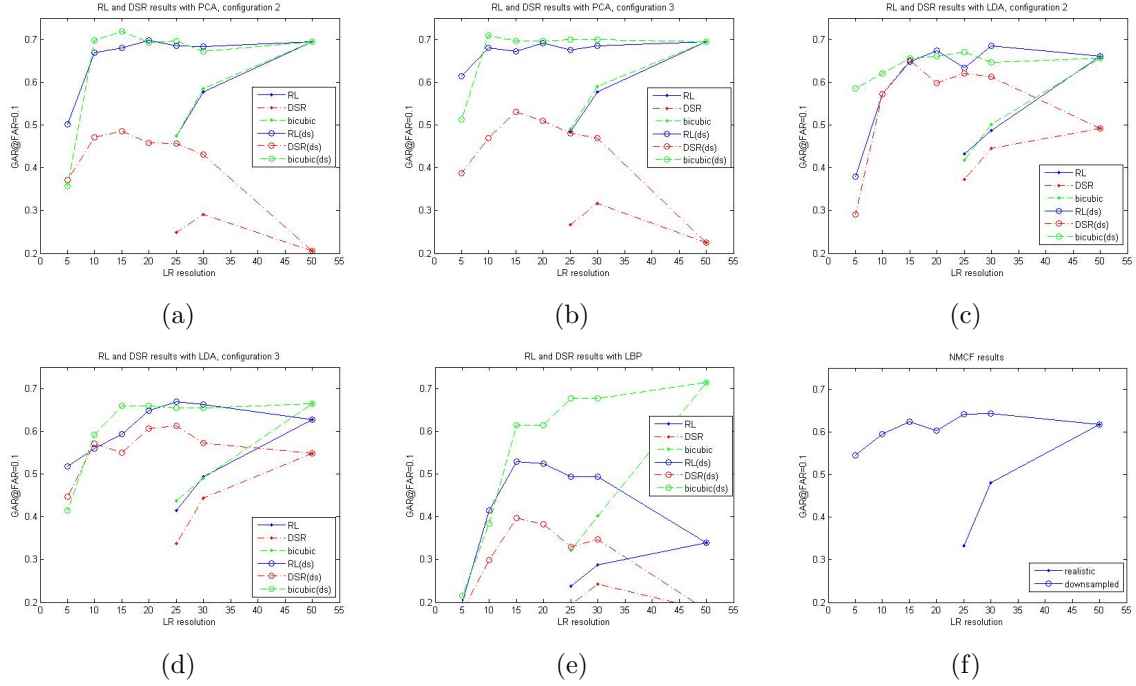
Figure 5: Super-resolution RL/DSR/NMCF results, GAR@FAR=0.1: (a) PCA, Config. 2; (b)PCA, Config. 3; (c) LDA, Config. 2; (d) LDA, Config. 3; (e) LBP results; (f) NMCF method. (ds=downsample)
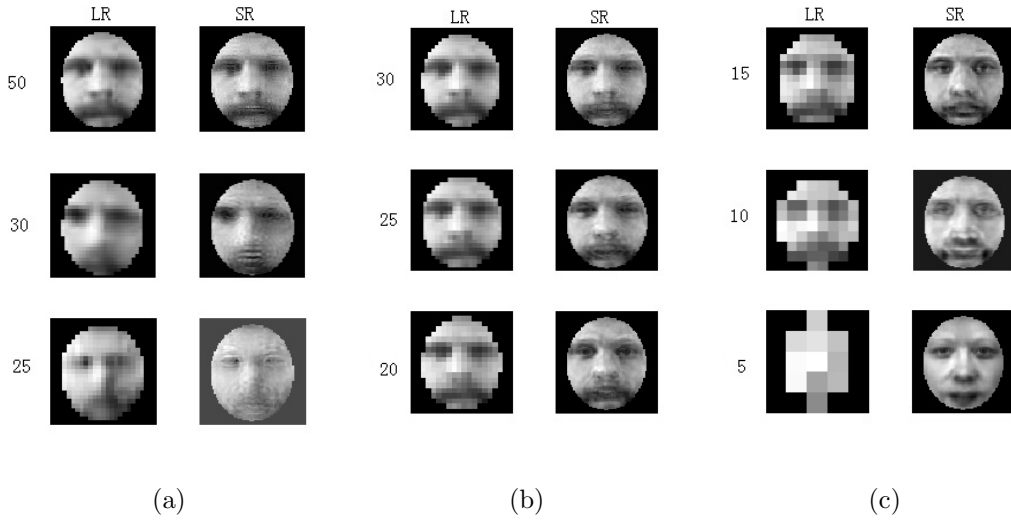

Figure 6: Reconstructed SR images by RL method from (a) real LR images, resolution 50×50, 30×30 and 25×25; (b) downsampled images, resolution 30×30, 25×25 and 20×20; (c) downsampled images, resolution 15×15, 10×10 and 5×5.

5 Conclusion

We evaluated the difference between downsampled and real LR face images for both standard face recognition and super-resolution. Our results show that face recognition on downsampled images performs much better than real LR images. The face

recognition accuracy hardly decreases when the resolution change is within a certain range. But for real LR images, the verification results drop significantly as the resolution become lower. Moreover, the single image super-resolution methods do not have benefits for face recognition in our experimental configurations. One possible solution is to apply SR method which can make use of the information of multiple images.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037 –2041, dec. 2006.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711 –720, jul 1997.

[3] 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.

[4] M. Grgic, K. Delac, and S. Grgic. Scface — surveillance cameras face database. *Multimedia Tools Appl.*, 51:863–879, February 2011.

[5] P. H. Hennings-Yeomans, S. Baker, and B. V. K. V. Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

[6] H. Huang and H. He. Super-resolution method for face recognition using nonlinear mappings on coherent features. *IEEE Transactions on Neural Networks*, 22(1):121–130, 2011.

[7] K. Nasrollahi and T. Moeslund. Finding and improving the key-frames of long video sequences for face recognition. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1 –6, sept. 2010.

[8] A. O'Toole, J. Harms, S. Snow, D. Hurst, M. Pappas, J. Ayyad, and H. Abdi. A video database of moving faces and people. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5):812–816, 2005.

[9] 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 *In IEEE CVPR*, pages 947–954, 2005.

[10] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586 –591, jun 1991.

[11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511–I–518 vol.1, 2001.

[12] F. W. Wheeler, X. Liu, and P. H. Tu. Multi-frame super-resolution for face recognition. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–6, 2007.

[13] D. Zhang, J. He, and M. Du. Morphable model space based face super-resolution reconstruction and recognition. *Image and Vision Computing*, 30(2):100 – 108, 2012.

[14] W. Zou and P. C. Yuen. Very low resolution face recognition problem. In *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010*, 2010.