

# An Evaluation of Super-Resolution for Face Recognition

Yuxi Peng    Berk Gökberk    Luuk Spreeuwiers    Raymond Veldhuis

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

{y.peng,b.gokberk,l.j.spreeuwiers,r.n.j.veldhuis}@utwente.nl

## Abstract

We evaluate the performance of face recognition algorithms on images at various resolutions. Then we show to what extent super-resolution (SR) methods can improve the recognition performance when comparing low-resolution (LR) to high-resolution (HR) facial images. Our experiments use both synthetic data (from the FRGC v1.0 database) and surveillance images (from the SCface database). Three face recognition methods are used, namely Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Local Binary Patterns (LBP). Two SR methods are evaluated. The first method learns the mapping between LR images and the corresponding HR images using a regression model. As a result, the reconstructed SR images are close to the HR images that belong to the same subject and far away from others. The second method compares LR and HR facial images without explicitly constructing SR images. It finds a coherent feature space where the correlation of LR and HR is maximum, and then compute the mapping from LR to HR in this feature space. The performance of the two SR methods are compared to that delivered by the standard face recognition without SR. The results show that LDA is mostly robust to resolution changes while LBP is not suitable for the recognition of LR images. SR methods improve the recognition accuracy when downsampled images are used and the first method provides better results than the second one. However, the improvement for realistic LR surveillance images remains limited.

## 1 Introduction

Face recognition has gained much attention in recent decades [12]. Face recognition systems deliver promising results when using high-resolution (HR) frontal images, but face recognition at a distance remains challenging. The face regions of images acquired at a distance are usually small and of low quality. To deal with the low-resolution (LR) problem of face images, super-resolution (SR) methods can be applied to increase the resolution of an image.

SR was initially intended to construct HR images for visual enhancement. These methods have achieved great success, but the objective of most SR methods is to construct high-frequency details that is insufficient for the recognition of LR images. Recently, some SR methods have been developed specially for face recognition problem. Hennings-Yeomans et al. [5] built a model for SR based on Tikhonov regularization and a linear feature extraction stage. This model can be applied when images from training, gallery and probe sets have varying resolutions. This approach was extended in [6] by adding a face prior to the model and using relative residuals as measures of fit. In [3], Biswas et al. proposed an approach using multidimensional scaling to improve

the matching performance of LR images. Their method finds a transformation matrix so that the distance between transformed features of LR images can be as close as possible to the corresponding HR images. Identity information about subjects is also used to make sure that the distance is small between data from the same class. Li et al. [8] proposed a method to obtain coupled mappings that project both HR and LR image features to a unified feature space in which direct comparison of HR and LR is possible. The objective function is built to cluster the projections of LR images and their corresponding HR images in the new feature space. A face recognition system for long video sequences is presented by Nasrollahi and Moeslund [9]. In [9], key-frames are first selected and then a hybrid SR method is applied. The images that are closest to full-frontal and with higher quality score are chosen as the key-frames. Multiple images from the key-frames are used to construct HR images.

In this paper, we first evaluate the performance of three standard face recognition algorithms (PCA [11], LDA [2] and LBP [1]) at various image resolutions and then apply two SR methods to LR images in order to observe their contribution to the identification performance. In our experiments, two face databases are used: the first face database (FRGC v1.0 [10]) contains high-quality images captured at controlled situations. The second database, SCface [4], contains surveillance quality facial images captured at three different distances. The performance of two SR methods (DSR method [13] and Huang and He’s method [7]) are compared with standard face recognition without SR. The remainder of this paper is organized as follows. In Section 2, the two SR methods are introduced. Experimental setup and identification test results are presented in Section 3. Section 4 concludes the paper.

## 2 Super-resolution methods

Two state-of-the-art SR methods are chosen in our experiments. They are explained in detail in the following sections.

### 2.1 DSR method

In [13], Zou and Yuen proposed a simple but effective SR method for very low resolution images which is compatible with various face recognition methods. This method is called discriminative super resolution (DSR). They introduce a data constraint which clusters the constructed SR images with the corresponding HR images. Identity information about the subject is also used to improve the recognition accuracy.

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

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

where  $\gamma$  is a constant to balance the two terms. We set  $\gamma$  to 1 in our experiments. The first term of (1) minimizes the distance between HR and the space of LR projected by  $R$ . The second term  $d(R')$  is represented as

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

where  $\lambda_i$  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.

Thus, for a given LR image  $I_{input}$ , we first apply  $I_{SR} = R I_{input}$  to obtain SR image  $I_{SR}$ , and then use  $I_{SR}$  for face recognition.

## 2.2 NMCF method

In [7], 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.

Given a training set of HR and LR image pairs ( $\{I^H, I^L\} = \{I_i^H, I_i^L\}_{i=1}^N$ ), firstly PCA features are extracted to reduce computational costs. Next, CCA is used to project PCA features to a coherent feature space. In this feature space, 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. Define  $C_{11} = E[\hat{X}^H(\hat{X}^H)^T]$  and  $C_{22} = E[\hat{X}^L(\hat{X}^L)^T]$  as the within-set covariance matrices, and  $C_{12} = E[\hat{X}^H(\hat{X}^L)^T]$  and  $C_{21} = E[\hat{X}^L(\hat{X}^H)^T]$  as the between-set covariance matrices, where  $E[\cdot]$  stands for mathematical expectation. 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 base matrices  $V^H$  comprises eigenvectors of  $R_1$  and  $V^L$  comprises eigenvectors of  $R_2$  when their corresponding eigenvalues are sorted in descending order. The coherent features of HR and LR images are

$$C^H = (V^H)^T \hat{X}^H, C^L = (V^L)^T \hat{X}^L. \quad (3)$$

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  $\Phi$  is a multiquadratic basis function (see [7] for details). As a result, the weight matrix  $W$  can be solved by  $W = C^H(\Phi + \tau I_{id})^{-1}$ .  $\tau I_{id}$  is included because  $\Phi$  is not always invertible.  $\tau$  is a small positive value, such as  $10^{-3}$ , and  $I_{id}$  is the identity matrix.

In the testing stage, the coherent features of HR gallery images are first computed. For a LR probe image  $I^p$ , we compute the coherent features  $c^p$  and then apply the learnt mapping to obtain the SR features of the probe image by

$$c_{SR} = W \cdot [\varphi(\|c_1^l - c^p\|) \dots \varphi(\|c_N^l - c^p\|)]^T. \quad (4)$$

where  $\varphi(\|c_i - c_j\|) = \sqrt{\|c_i - c_j\|^2 + 1}$ . Finally, the above features are fed to the nearest neighbor classifier for recognition.

## 3 Experimental results

In this section, we present identification results of the selected face recognition algorithms at various image resolutions and evaluate the performance of SR methods.

### 3.1 Low-resolution face recognition performance

In this section we provide the recognition performance of PCA, LDA and LBP on face images at various resolutions. In our identification experiments, we use 2820 facial images from the FRGC v1.0 database [10]. Original high-quality face images are resized to  $70 \times 60$ ,  $56 \times 48$ ,  $42 \times 36$ ,  $28 \times 24$ ,  $14 \times 12$  and  $7 \times 6$  pixel resolutions using bicubic interpolation. See Figure 1(a). 1228 images of 139 persons are selected as training images. Two images per person from the remaining images are used as gallery (242 images) and the others are used as probe images (1350 images). Distance measures employed in 1-nearest neighbor classifier for PCA, LDA and LBP are L1 norm, cosine angle and Chi square, respectively. We use uniform patterns and (8, 1) neighborhood

for the LBP approach where face images are divided into 49 regions ( $7 \times 7$  grid); but for images at the resolutions  $14 \times 12$  and  $7 \times 6$ , the images are divided into 16 and 4 regions. The identification rates are shown in Figure 1(b).

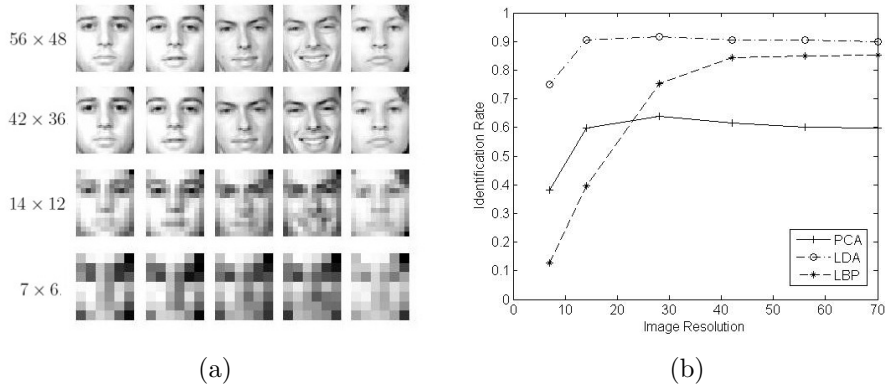


Figure 1: (a) Sample face images from the FRGC v1.0 database at various resolutions, (b) identification rates

As Figure 1(b) shows, LDA outperforms PCA and LBP at all image resolutions. Both PCA and LDA accuracies decrease sharply when the image resolution is lower than  $14 \times 12$ . On the other hand, LBP has shown to be sensitive to resolution differences: correct classification rate of LBP is only 12.8% for images at resolution  $7 \times 6$  but the performance rises as the resolutions get higher and becomes stable when reach the resolution  $42 \times 36$ .

We then test the recognition performance on the Surveillance Cameras Face (SCface) Database [4]. The SCface database contains images from 130 subjects taken by five surveillance cameras at three distances, namely 4.20 meters (distance1), 2.60 meters (distance2), and 1.00 meter (distance3). It also contains one frontal mug-shot image for each subject. We crop faces of the original images according to the eye coordinates (see Figure 2).



Figure 2: Sample face images from the SCface database captured at: (a) distance1, (b) distance2, (c) distance3 and (d) mug-shots.

In our recognition experiments, frontal mug-shot images are used as gallery images. For PCA and LDA training, we use images from four cameras at a particular distance and use the remaining camera images as probe set. Using this leave-one-out methodology, we perform five recognition experiments at each distance (distance 1, 2, and 3) and provide the average correct classification rate. Identification rates are shown in Table 1. Additionally, we also utilized FRGC v1.0 database to train PCA and LDA classifiers to test the generalization ability of these methods on the SCface database. Last three rows of Table 1 provides the identification rates with FRGC training. It should be noted that the LBP approach does not need a training set. Therefore LBP results for different training schemes are identical in Table 1.

Table 1: Identification rates [%] for PCA, LDA and LBP on the SCface database.

Training Set	Probe Set	PCA	LDA	LBP
SCface Dist1	SCface Dist1	12.3	24.6	8.3
SCface Dist2	SCface Dist2	18.0	33.5	21.7
SCface Dist3	SCface Dist3	12.3	24.5	19.1
FRGC	SCface Dist1	9.9	10.3	8.3
FRGC	SCface Dist2	18.0	18.0	21.7
FRGC	SCface Dist3	11.1	10.7	19.1

The identification rates of PCA and LDA using FRGC training set were found to be at most 18.0%. Though images captured at distance3 have a higher resolution than those of the other two, their results are worse than at distance2 as a consequence of the pose variation problem: because the subjects get close to the cameras, the images contain mostly the top part of the faces. Moreover, using SCface training sets instead of FRGC training set provides better results and it benefits LDA more than PCA. LBP is more sensitive to resolution changes. The LBP result is worse than that of both PCA and LDA at distance1. However, at distance2 and distance3, LBP outperforms LDA using FRGC training and PCA using both training configurations.

### 3.2 Super-resolution face recognition performance

In the real surveillance situations, HR images are usually pre-stored for training and gallery and LR images are captured later as probe images. To simulate this situation, experiments are conducted using the images from the FRGC database at resolution  $70 \times 60$  for training and gallery, and images at lower resolution for probe sets. The probe images are resized with bicubic interpolation to resolution  $70 \times 60$ . As shown in Figure 3, the identification rates of images at resolution  $7 \times 6$  and  $14 \times 12$  are worse than those of images at higher resolution for all methods. The accuracies do not vary a lot when image resolutions are higher than  $28 \times 24$ . Thus, we will apply SR methods on images at resolution  $7 \times 6$  and  $14 \times 12$  to see their contribution in the following. The LR images chosen for SCface are images captured at distance1 for the SCface database.

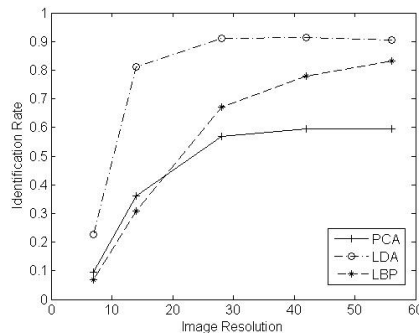


Figure 3: Identification rates for various resolutions with bicubic interpolation on the FRGC database

In SR experiments, we have two training phases: one is to train the SR method (e.g., learning the LR-HR mapping) and the other one is to train the face classifier

(e.g., PCA training). In FRGC experiment, the FRGC training set, which contains 1228 images, is used for both the two training phases. Images at a resolution of  $70 \times 60$  are used as HR. LR images have resolutions  $7 \times 6$  and  $14 \times 12$ . To train the SR system, both HR and LR images are needed, while only HR images are used to train the classifiers. In addition to comparing the performance using SR images and original HR images, we employ a basic bicubic interpolation as a baseline SR method. Comparative identification rates obtained by both SR method are shown in Figure 4.

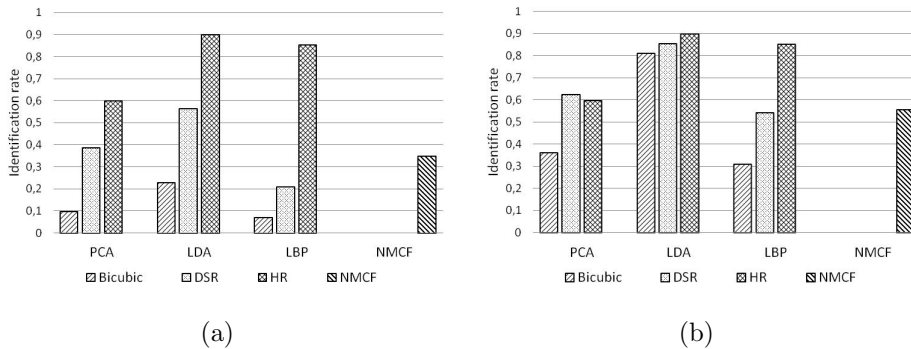


Figure 4: Comparison of DSR method with NMCF method on the FRGC database: (a) LR  $7 \times 6$ , (b) LR  $14 \times 12$ .

DSR method improves the correct classification rates for all face recognition methods at both  $7 \times 6$  and  $14 \times 12$  pixel resolutions: it consistently achieves higher identification rates than the bicubic interpolation method. It is also seen that the relative gain of using a SR method is more visible at lower resolutions. For instance, with LDA classifier, relative performance increase is higher at the  $7 \times 6$  pixel resolution than  $14 \times 12$ . At both resolutions, LDA outperforms the other methods. The second SR method, NMCF approach, also attains better identification rates than the LR results but they are not as good as DSR approach when PCA and LDA are used as the classifier.

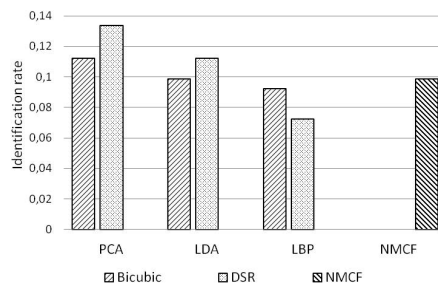


Figure 5: Comparison of DSR method with NMCF method on the SCface database.

For the experiment using the SCface database (Figure 5), images from FRGC are used to train the classifiers. Frontal mug-shots are gallery and images captured at distance1 are used as probe images. DSR method requires that the number of training images must larger than the number of pixels in the LR image. Since images captured at one distance are not enough, we use images captured at both distance2 and distance3 to train the SR system. The HR training images are resized to the same resolution as images at distance2, and the LR training images are downsampled from the HR images to the same resolution as images at distance1.

In SCface database experiments, PCA outperforms LDA and LBP. DSR method can improve the results of PCA and LDA but not for LBP. NMCF approach provides better results than DSR methods with LBP but worse than the others.

In addition, some SR images constructed by the DSR method are shown in Figure 6. The SR images of FRGC have much better quality than the SR images of SCface. The SR images of SCface have much more artifacts that lead to the poor quality of local features, thus, reduce the accuracy of LBP.

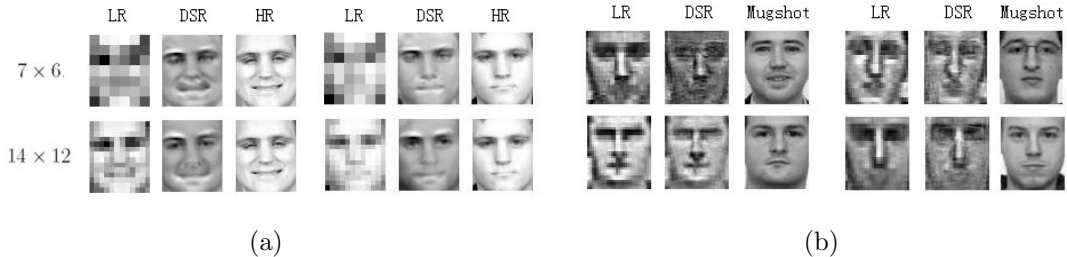


Figure 6: SR images constructed by DSR method (a) on the FRGC database (b) on the SCface database.

As a summary, we provide a comparison of three main schemes described above on the FRGC database (see Table 2): 1) matching of LR probe to LR (downsampled) gallery, 2) matching of upsampled (using bicubic interpolation) probe to HR gallery image and 3) matching of SR probe to HR gallery image. The classifiers are trained with LR images in the first scheme while they are trained with HR images in the last two schemes. The upsampling scheme provides the worst results for both LR resolutions. The SR scheme provides better results than the LR scheme when PCA and LBP are used. However, all the SR results are not as good as the LDA results of the first scheme.

Table 2: Comparison of the identification rates [%] of three schemes on the FRGC database.

LR resolution	Probe	Gallery	PCA	LDA	LBP
$7 \times 6$	LR	LR	38.3	75.1	12.8
$7 \times 6$	Bicubic	HR	9.6	22.8	6.8
$7 \times 6$	DSR	HR	38.6	56.4	20.9
$14 \times 12$	LR	LR	59.9	90.5	39.5
$14 \times 12$	Bicubic	HR	36.1	81.2	31.0
$14 \times 12$	DSR	HR	62.4	85.6	54.2

## 4 Conclusion

In this paper, the performance of several face recognition algorithms, namely PCA, LDA and LBP, is evaluated for low-resolution face images. Our results show that LDA outperforms the others when down-sampled images are used. LBP is found to be not suitable for very low resolution images. The overall recognition accuracy is improved if super-resolution methods are applied. DSR method, as a reconstruction-based super-resolution approach, can be easily paired with standard face recognition algorithms and outperforms NMCF approach. However, the improvement of super-resolution methods

on surveillance images is limited. Finally, using downsampled gallery and training images instead of using super-resolved probe images obtains better results for LDA classifier.

## 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 facial images. In *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010*, 2010.
- [4] M. Grgic, K. Delac, and S. Grgic. Sface — 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] P. H. Hennings-Yeomans, B. V. K. Vijaya Kumar, and S. Baker. Robust low-resolution face identification and verification using high-resolution features. In *Proceedings - International Conference on Image Processing, ICIP*, pages 33–36, 2009. Cited By (since 1996): 1.
- [7] H. Huang and H. He. Super-resolution method for face recognition using non-linear mappings on coherent features. *IEEE Transactions on Neural Networks*, 22(1):121–130, 2011.
- [8] 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, jan. 2010.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face Recognition: A Literature Survey. *ACM Computing Surveys*, pages 399–458, 2003.
- [13] 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.