Face Recognition at a Distance:
a study of super resolution

M.Sc. Thesis

Yuxi Peng

University of Twente
Department of Electrical Engineering,
Mathematics & Computer Science (EEMCS)
Signals & Systems Group (SAS)
P.O. Box 217
7500 AE Enschede
The Netherlands

Report Number: SAS2011-014
Report Date: August 28, 2011
Period of Work: 01/02/2011 – 31/8/2011
Thesis Committee: Dr. ir. R.N.J. Veldhuis
Dr. ir. B. Gokberk
Dr. ir. L.J. Spreeuwers
Prof. Dr. ir. C.H. Slump
Dr. M. Poel
Abstract

We evaluate the performance of face recognition using images with different resolution. The experiments are conducted on Face Recognition Grand Challenge version one (FRGC v1.0) database and Surveillance Cameras Face (SCface) Database. Three recognition methods are used, namely Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Local Binary Pattern (LBP). To improve the performance of face images with low-resolution (LR), two state-of-art super-resolution (SR) methods are applied. One is called Discriminative Super-resolution (DSR). It finds the relationship from low-resolution images to their corresponding high-resolution (HR) images so that the reconstructed super-resolution images would be close to the HR images which belongs to the same subject with them and far away from others. The other SR method uses Nonlinear Mappings on Coherent Features (NMCF). Canonical Correlation analysis is applied to compute the coherent features between the PCA features of HR and LR images. Then Radial Basis Functions (RBFs) is used to find the mapping from LR features to HR features in the coherent feature space. The two SR methods are compared on both FRGC and SCface databases as well.
Acknowledgements

This thesis presents the work I have done in the Signals and Systems group in University of Twente. During the seven-month work, I have worked with a lot of people who contributed their time and efforts to my research. It is a pleasure for me to express my gratitude to them all.

First of all, I owe my deepest gratitude to Dr. ir. R.N.J. Veldhuis for supervision, advice, and guidance from the initial to the final level enabled me to develop an understanding of the project. His truly scientist intuition has made him as a constant oasis of ideas and passions in science, which exceptionally inspire and enrich my growth as a student wants to be. Special thanks are also given to Dr. ir. B. Gokberk for the advice, direct supervision and useful discussions throughout the project. I am also grateful to Dr. ir. L.J. Spreeuwers for giving me useful advice.

I gratefully thank Chanjuan Liu, who was always willing to offer help during my project. I would like to thank our technician Geert Jan Laanstra and our secretary Sandra Westhoff for administration support.

I would like to express my gratitude towards my parents and my friends for their encouragement and support during my study in the Netherlands.
Contents

Abstract i
Acknowledgements iii
Table of Contents vi
List of Figures viii
List of tables ix

1 Introduction 1
   1.1 An introduction to face recognition . . . . . . . . . . . . . . . 1
   1.2 Purpose of our research . . . . . . . . . . . . . . . . . . . 2
   1.3 Outline of the report . . . . . . . . . . . . . . . . . . . . . 2

2 Literature review 3
   2.1 Face recognition at a distance . . . . . . . . . . . . . . . . . . . . 3
   2.2 Super resolution . . . . . . . . . . . . . . . . . . . . . . . 8
      2.2.1 SR for visual enhancement . . . . . . . . . . . . . . 8
      2.2.2 SR for recognition . . . . . . . . . . . . . . . . . . . 9

3 Algorithm introduction 13
   3.1 Face recognition methods . . . . . . . . . . . . . . . . . . . . . . . . 13
      3.1.1 Eigenface . . . . . . . . . . . . . . . . . . . . . . . . 13
      3.1.2 Fisherface . . . . . . . . . . . . . . . . . . . . . . . . 14
      3.1.3 LBP . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
   3.2 Super resolution methods . . . . . . . . . . . . . . . . . . . . . . 16
      3.2.1 RL and DSR . . . . . . . . . . . . . . . . . . . . . . . 16
      3.2.2 NMCF . . . . . . . . . . . . . . . . . . . . . . . . . . 17
## 4 Experiments

4.1 Database description ........................................... 19  
   4.1.1 ORL ...................................................... 19  
   4.1.2 FRGC v1.0 .............................................. 20  
   4.1.3 SCface .................................................. 20  
4.2 The influence of different image resolution ................. 21  
   4.2.1 Configuration ........................................... 21  
   4.2.2 Experimental results ................................... 23  
   4.2.3 Discussion ............................................... 25  
4.3 Surveillance camera results .................................. 25  
   4.3.1 Configuration ........................................... 26  
   4.3.2 Experimental results ................................... 26  
   4.3.3 Discussion ............................................... 30  
4.4 Super resolution .............................................. 31  
   4.4.1 Configuration ........................................... 31  
   4.4.2 Experimental results ................................... 32  
   4.4.3 Discussion ............................................... 44  

## 5 Conclusion ................................................... 45

A Appendix ....................................................... 47

Bibliography ..................................................... 51
List of Figures

1.1 Face recognition procedures. ......................................... 1

3.1 Examples of eigenfaces. ........................................... 14

4.1 Images from the first subject in ORL. ............................ 19
4.2 FRGC sample images. .................................................. 20
4.3 SCface sample images. First row: distance1; second row: distance2; third row: distance3; last row: mug shot frontal images. ................................................................. 21
4.4 FRGC sample images with different resolution. First row: 64 × 64; second row: 32 × 32; third row: 16 × 16; fourth row: 12 × 12; last row: 6 × 6. ......................................................... 22
4.5 All combination of PCA and LDA dimensionalities. Left: 32 × 32; right: 64 × 64. ............................................................ 23
4.6 LDA results on three configurations. ................................. 24
4.7 Compare LBP with LDA .................................................... 24
4.8 LBP extra experiments. Left: with weight vs. without weight; right: all patterns vs. only uniform patterns. ............... 25
4.9 SCface cropped images. First row: distance1; second row: distance2; third row: distance3; last row: mug shot frontal images. ................................................................. 27
4.10 Results of LDA tests of different PCA dimensionality. ......... 28
4.11 FRGC sample images. First row: 56×48, second row: 42×36, third row: 14×12, last row: 7×6. ................................. 31
4.12 SCface cropped images, experiment 3. First row: distance1; second row: distance2; third row: distance3; last row: mug shot frontal images. ................................................................. 32
4.13 FRGC SR image construction using RL ........................... 33
4.14 FRGC SR image construction using DSR. .......................... 34
4.15 DSR results on FRGC. PCA results. Left: input resolution 7 × 6; right: input resolution 14 × 12. ............................. 35
4.16 DSR results on FRGC. LBP results. Left: input resolution $7 \times 6$; right: input resolution $14 \times 12$. ........................................ 36
4.17 DSR results on FRGC. LDA results. Left: input resolution $7 \times 6$; right: input resolution $14 \times 12$. ........................................ 36
4.18 SCface SR image construction using FRGC for training. (1st and 3rd row: distance1 LR images; 2nd and last row: corresponding SR images.) ....................................................... 37
4.19 SCface SR image construction 2. 1st, 4th and 7th row: distance1 LR images; 2nd row: distance2 SR; 3rd row: distance2 HR images; 5th row: distance3 SR; 6th row: distance3 HR images; 8th row: distance2+3 SR. ........................................ 38
4.20 Face recognition results of SCface using DSR ..................... 39
4.21 ORL sample images. First row: $32 \times 32$ train; second row: $8 \times 8$ train; last row: $8 \times 8$ test. .................................................... 39
4.22 NMCF results of different PCA dimensions on ORL ............ 40
4.23 NMCF results compared with [9]. ........................................ 40
4.24 NMCF results of different PCA dimensions on FRGC. Left: LR $7 \times 6$, HR $54 \times 48$; right: LR $14 \times 12$, HR $54 \times 48$. ........ 41
4.25 Face recognition results of FRGC using NMCF. ................. 41
4.26 NMCF results of different PCA dimensions on SCface. (a) distance2 SR. ................................................................. 42
4.27 NMCF results of different PCA dimensions on SCface. (b) distance3 SR. ................................................................. 42
4.28 NMCF results of different PCA dimensions on SCface. (c) distance2+3 SR. ................................................................. 43
4.29 Comparison of NMCF and DSR on SCface. ......................... 43
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Information about selected cropped images</td>
<td>27</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparison of PCA[6], LDA and LBP. Identification rate[%].</td>
<td>28</td>
</tr>
<tr>
<td>4.3</td>
<td>2nd configuration. Identification rate[%]. CAM1t stands for using images from cam1 for training.</td>
<td>29</td>
</tr>
<tr>
<td>4.4</td>
<td>3rd configuration</td>
<td>30</td>
</tr>
<tr>
<td>4.5</td>
<td>Details of LBP and LDA parameters.</td>
<td>35</td>
</tr>
<tr>
<td>4.6</td>
<td>DSR results on SCface</td>
<td>38</td>
</tr>
<tr>
<td>4.7</td>
<td>Number of PCA vectors chosen for FRGC</td>
<td>41</td>
</tr>
<tr>
<td>4.8</td>
<td>NMCF results on SCface</td>
<td>43</td>
</tr>
</tbody>
</table>
Introduction

1.1 An introduction to face recognition

Face recognition has become an attractive research field in recent decades. Given a face image, face recognition system automatically identify or verify a certain individual [26]. Usually, face recognition follows the procedure in Figure 1.1.

Image acquisition: the first stage is to obtain images containing facial region. Different acquisition devices are used for different situation. The most popular device for capturing two-dimensional (2D) images is the charge-coupled device (CCD) camera. Three-dimensional (3D) images are usually acquired by stereo of the images captured by two cameras placed on the left and right. Infrared camera is also used in some situation such as night time.

Face detection/tracking: when images are available, we need to locate
the faces in them. In videos, faces may have to be tracked for a certain time to obtain better quality images.

Feature extraction: at this stage, useful data is extracted for recognition. Some pre-processing approaches might be needed before extracting features, for instance, using image segmentation method to crop the face region. The images may need alignment and normalization as well. Then the global or local features will be extracted from the face images.

Classification: when the image features are extracted, they are fed into a classifier for identification or verification. The system will declare a decision about the identity of the subject.

Face recognition systems can obtain promising results when using high-resolution (HR) frontal images, but face recognition at a distance (FRAD) is still challenging. The face regions of images acquired at a distance are usually small and have low quality. Face recognition task can be more difficult in surveillance situation.

To deal with the low-resolution (LR) problem of images acquired from long distance, super-resolution (SR) methods can be applied. Super-resolution methods try to enhance the resolution of an image. In this report, face recognition performance of low-resolution images is evaluated and two state-of-art super-resolution methods are compared.

1.2 Purpose of our research

The objective of our research is to find out the influence of low-resolution images to face recognition system and how the state-of-art super-resolution approaches would improve the recognition performance.

1.3 Outline of the report

The content of this report is arranged as follows:

Chapter 2: literature review. In this chapter, we review previous work on face recognition at a distance and super resolution.

Chapter 3: algorithms. Three important face recognition methods and two state-of-art super-resolution algorithms are introduced.

Chapter 4: experiments. This chapter includes database description and all the experiments we have conducted. For each experiment, the experiment configuration is first introduced, then the experimental results and the discussions are presented.

Chapter 5: Conclusion. In this chapter, an overall conclusion is given for this master thesis project and future work to improve the performance of face recognition at a distance is also discussed.
In this chapter, firstly the main problems of face recognition at a distance are described and several recent studies on face recognition at a distance are introduced. Then super-resolution methods for both visual enhancement and face recognition are discussed.

2.1 Face recognition at a distance

Although face recognition algorithms obtain high recognition rates in close proximity, great challenge still exists in long distance which is also referred to as face recognition at a distance. Why face recognition at a distance different? First, it’s hard to ensure the quality of images. Second, long-distance face recognition is usually used for surveillance without cooperation of users. In addition, it may need to work 24 hours and in outdoor condition. Issues like lighting and weather condition are significant problems. It may also deal with multi-faces.

Problems of FRAD

The main problems and some relevant solutions of FRAD are stated in Chapter 6 of [19]. Firstly, the images captured at a distance are usually with low resolution which will lead to low recognition accuracy. Using high-definition camera is a possible solution but it can result in decreasing of speed of detection. Secondly, interlace is a problem in video images. It happens when the faces are moving fast enough that each video frame is captured at a different position. A progressive scan video system can reduce the chance of interlace. Thirdly, in many cases of FRAD the face is blurred because it is out of focus of the lens. The degree of blur can be decreased by small aperture lens. Fourthly, motion blur also happens frequently in FRAD when the face is moving fast or the camera is shaking. Rapid exposures can be used to avoid this problem, but the aperture stop would be increased which con-
fects to the out-of-focus problem. Furthermore, in FRAD especially under surveillance situation, it is also important to cover most users’ heights and capture frontal faces. Besides, weather and atmospherics, such as thermal waves, have a significant impact on recognition accuracy [19].

A FRAD system for pose variation

In [14], some solutions to different aspects of face recognition at a distance especially for pose variation problem are proposed. To obtain good quality images, two different cameras are used. A wild field camera is placed in a fixed position and a foveal camera is placed on a pan/tilt platform. The output of the wild-field camera is used to orient the foveal camera at faces.

They developed a robust face detection approach combining three modalities: 2-frame motion differencing, background subtraction and skin color detection. These three modalities can make up for the weakness of each other. For each pixel in an image, the possibility of it being a face pixel is modeled as a mixture of Gaussian distributions of the results of the three modalities. In order to track an individual, a fixation prior is introduced which causes the camera to dwell on the person. When a face is successfully detected, a feedback map is generated to inhibit this position temporarily so that the system would move to a different position for next detection.

In the pre-processing phase, a cascade-based face detector is employed. Firstly, twenty landmarks on the face is found by combining a local likelihood model learnt from a training dataset and a Gaussian model of the covariance of the twenty feature points. Then a warp is performed by triangulating the images based on the landmarks and mapping the vertices of the triangle to the standard shape. Finally, the features are extracted by placing a regular grid of 25 squares centered on each landmark and computing the average gradient in eight directions and the mean intensity level. These extracted features are correlated across different poses.

A one-to-many mapping from an identity space without pose variation to the observed data space is proposed to deal with various poses for recognition. This mapping is learnt from a training set using expectation-maximization (EM) algorithm. For input images, the mapping can be applied to obtain the corresponding features in the identity space. Thus, a simple way to perform recognition is to use nearest neighbor classifier on these features.

This method has promising results on pose problem but can not deal with illumination problem. Another weakness is that the system may fail to recognize subjects who only enter the area for a certain time period.
3D model for pose problem in FRAD

In [11], a prototype system to locate, track and identify people going through a predefined zone is proposed. The system focus on detection and tracking and pre-processing. A single ultra-high resolution camera is used to capture images. This system locates a human head by locating the entire body of a person. Moving subjects are detected using background subtraction while static subjects are detected using model-based detection module which employs Edgelet features. Elliptic shape mask and an appearance map are applied when tracking subjects under difficult crowd conditions.

Because this long-distance face recognition is without user cooperation, frontal images are not always obtained. 3D models are used to deal with pose problem. The pose of the acquired images is initially estimated using a tracker which combines a recursive and key-frame based approach to minimize tracking drift and jitter. Then a two-stage 3D-reconstruction is applied. First, optimal poses are selected and corresponding points are back-projected and intersected to determine the 3D position. Second, the individual reconstructions are integrated into a single 3D point cloud. A mesh description of the dense cloud of points is inferred based on the approximation of the head geometry.

The reconstructed 3D models are nice for both indoor and outdoor situations. However, the 3D recognition has not been tested.

Magnifications and recognition rate

A database acquired from long distances and with high magnifications and proposed an algorithm to deal with magnification blur is presented in [25]. Indoor images and video sequences are taken at distance 10 to 16 meters by a commercially available pan tilt zoom (PTZ) camera. The images are captured at different distances. System magnification and illumination changes are present in the images. The gallery images are taken at 0.5 m distance. The outdoor images and video sequences are captured by a Meade ETX-90 telescope coupled with a JVC MG-37U camcorder. The subjects have two motions: standing still and walking. The gallery images are taken at 1 m at indoor situation.

According to the quality assessment of images from the database described above, the recognition rate drops as system magnification increases and image sharpness decreases as system magnification/distance increases. A multi-scale processing based on wavelet transforms is used to restore and enhance data with high magnification values. At first, the mean sharpness of 1× and 10× image sets is computed. If the sharpness of an input image is larger than the mean sharpness, the image is decomposed via Haar wavelet transform of level one. Then deblurring algorithms to the approximation coefficients and denoising algorithms to the vertical/horizontal/diagonal detail
coefficients are applied. Adaptive grey level contrast stretching is applied then and output image is reconstructed by the corresponding inverse wavelet transform. Lasso regularization is used for recognition. This wavelet based algorithm increases recognition rate. But as magnification increases, improvement decreases.

Close, medium and far distance scenarios

Close, medium and far distance face recognition in segmented faces are analyzed in [13]. Two recognition system is used. One uses Principal Component Analysis (PCA) with Support Vector Machine (SVM) classifier. The other one applies Discrete Cosine Transform (DCT) with Gaussian Mixture Model (GMM) classifier. The images are manually marked with “close”, “medium” and “far” distance based on whether the image contains only head and should or the upper body or the full body. The faces are segmented using the VeriLook SDK and the errors are corrected by manually marking the eyes. The quality assessment shows that, for full images, the farther the distance the higher the entropy while the opposite occurs for segmented images. Then the recognition performance of PCA-SVM system and DCT-GMM system are evaluated. Half of the close distance images are used for training. The other half of close distance images as well as medium and far distance images are used for testing. The results indicate that the recognition accuracy decreases as the distance increases. And GMM-based system works better in far distance conditions.

This work is extended in [20] by the same authors. They make the fusion of the previous two face verification methods to obtain better results. A “distance index” is defined to estimate distance:

\[
D = -\log(A_s/A_f)
\]

where \(A_s\) stands for segmented face area and \(A_f\) is full image area. Two ways of fusing the recognition methods PCA-SVM and DCT-GMM are as follows:

Fixed fusion: \(s = (s_{DG} + s_{PS})/2\);
Adaptive fusion: \(s = (D/2)s_{DG} + (1 - D/2)s_{PS}\)

where \(s_{DG}\) and \(s_{PS}\) are the scores obtained by PCA-SVM and DCT-GMM respectively, and \(s\) is the score of fusion.

In the recognition phase, images from different distances are used for training in different experiments respectively. According to the results, adaptive fusion works better that fixed fusion. Moreover, using medium distance for training has better performance than other two scenarios. Thus, training with medium distance images is a good way to control the performance degradation due to varying distance.
2.1 Face recognition at a distance

A surveillance system

In [23], a complete system for surveillance is proposed. The image acquisition setup is a co-located pair of Wide Field Of View (WFOV) and Narrow Field Of View (NFOV) cameras with a high-end but standard workstation containing Matrox frame grabbers to accept the video streams. The aim of the systems is to detect and track people moving in the field of view of the fixed WFOV camera, so background subtraction is used for detection. And the NFOV camera view is calibrated with respect to the WFOV camera view. WFOV camera detects multiple subjects and uses a priority mechanism to select a target for NFOV camera. Every target has a score in his target record. The score is determined by the distance, direction, speed, etc. The subject that is more likely to leave the region will be selected first. Once a subject is selected, the filter will predict the location of his/her face at about 0.5 to 1.0 sec. The NFOV camera will zoom in and wait for the target to show up. And the resolution will increase by 20% each time the same target’s image is captured.

The faces are detected by Pittsburgh Pattern Recognition FT-SDK and cropped for next procedure. Face recognition is operated asynchronously. And the result will be passed to the target scheduler to update target record. Several commercial recognition methods are available in this system. In the detection experiment, they report one failed person detection and eight failed face captures out of 466 trials, usually due to the subject or face being obscured in some manner. The recognition result of Cognitec FaceVACS shows all successful recognition and most are recognized at 16 to 20 meters.

Stereo reconstruction for FRAD

In [16], a face recognition method based on sparse-stereo reconstruction is proposed. The image acquisition system contains stereo pair of high resolution cameras with adjustable baseline. User can remotely pan, tilt, zoom and focus on the subject’s nose tip. The images of 30 subjects are taken from three distances: 3, 15, and 33 meter ranges. The logRGB space is used to detect face region and extract landmark points. Then an Active Appearance Model (AAM) mesh is warped to detected landmarks. The stereo-based reconstruction is then used to infer the 3D information of the mesh vertices. PCA and Procrustes methods are used for recognition on both 3D vertices and their x-y plane projection (2D). Their experimental results show that both 2D experiments outperform 3D PCA and Procrustes. Procrustes performs slightly better than PCA. Moreover, the recognition accuracy decreases as distance increases. Additionally, the use of proper numbers of AAM points are better than using all AAM points for 15 and 33 meters because the face outline points and points inside the mouth contour lead to the error.
The authors extended their work in [15]. A combination of Viola-Jones detector and skin detector is used to identify possible facial regions and find the landmarks. To reconstruct 3D features, a dense, global stereo matching is first performed using global optimization algorithms. Then AAM is used to find sparse correspondences of the left and right images of the stereo pair. The experimental results have slightly improvement comparing to their work in [16]. However, these methods can not deal with expression variations. In their later research, texture is also used to obtain better results [17][18].

2.2 Super resolution

In the field of face recognition at a distance, most of the previous works focus on better image acquisition systems such as using high-quality cameras to improve the quality of captured images. However, when low-quality images are obtained, an intuitive approach of improving the quality of images is super-resolution. Super-resolution was designed at first to construct high-resolution images for visual enhancement. Some of the SR methods attempt to accumulate information of several low-quality images to reconstruct one super-resolved image. Since enough images would not always be available, researchers focus on reconstructing high-resolution image from only one LR input in recent years.

2.2.1 SR for visual enhancement

Three SR methods are proposed in Chapter 7 of [4]. The first one is called closed-loop super-resolution. It uses Generic 3D face model to improve the resolution of 3D texture. Bilinear basis images are obtained for further estimation of the pose and illumination coefficients. The SR method proposed in this paper is extended from an iterative back-projection (IBP) method to 3D. The SR texture is constructed separately based on six facial regions: eyes, eye brows, mouth and other parts to handle expression changes. And the super-resolution 3D facial texture is fed back to generate low-resolution bilinear basis images. Because the method has feedback mechanism, the quality of SR images becomes better as the number of frames used grows. This method can deal with illumination and expression, but the images are almost frontal images and their expression can’t change too much.

The second method can deal with significant expression changes. Face regions are tracked by a plane tracking algorithm based on minimizing the sum of squared difference between stored image of the reference facial region and the current image of it. Next an incremental free-form deformation formulation integrated with LR image formation model is used to model the local deformation. Then a super-resolution algorithm such as IBP is used to reconstruct high-quality images. To deal with warping errors, facial images are partitioned into six regions: eyes, eyebrows, mouth and other
2.2 Super resolution

parts. A match statistics is designed to measure the alignment. The SR results demonstrated that the proposed deformation method is much better than global registration.

The third one deals with side view face images in video. A face image is acquired by cutting the upper 16% of the segmented human body. Elastic registration method is used for motion estimation and a match statistic is introduced to detect and discard images that are poorly aligned. Then an iterative method is proposed to construct high-resolution images. In this approach, multiple low-resolution images are needed to reconstruct a high-resolution image.

2.2.2 SR for recognition

Though the super-resolution methods focus on visual enhancement have achieved great success, the objective of most SR methods is to construct high frequency details which is not sufficient for recognition of low-resolution images. Recently, some methods were developed for face recognition perspective. Here we briefly explain some of them.

$S^2R^2$

In [7], an algorithm called simultaneous super-resolution and recognition ($S^2R^2$) is proposed. A high-resolution training set is needed and let $F$ be the super-resolution feature which is provided by the training set. For an input low-resolution probe image $I_p$ and the class $k$ is claimed, first task is to solve the problem

$$X_p = \arg \min_x \|Bx - I_p\|^2 + \alpha^2 \|Lx\|^2 + \beta^2 \|Fx - FI_g\|^2 \quad (2.1)$$

where $I_g$ is the gallery image with class $k$. Matrix $B$ is the linear model for downsampling the HR image $x$ to its corresponding LR image $I_p$. $Lx$ is a vector of edge values in which $L$ is first or second derivative approximation. $\alpha$ is a regularization parameter. $\beta$ is an additional regularization parameter. Then the distance is measured by a combination of the residual norm on each set of model assumptions in Equation (2.1).

According to their experiment results, the performance of this algorithm is better than the original LR, bilinear interpolation, hallucinating faces[2] and Tikhonov regulation. This method is also extended for multiple resolution problems that the training images and gallery images are with different resolution and can be applied to non-linear features by using an appropriate non-linear optimization algorithm to minimize Equation (2.1). Furthermore, this method is improved in [8]. Nonetheless, this approach has to do optimization for every images in the gallery which requires long time especially for large database.
MDS

In [5], a novel approach for improving the matching performance of LR images using multidimensional scaling (MDS) is proposed. Their goal is to find a transformation matrix that the distance between transformed features of LR images can be as close as possible to their corresponding HR images. Let $I^h$ and $I^l$ represent for HR and LR images, and $W$ is the transformation matrix. The objective function is described as following

$$J(W) = \lambda \sum_{i=1}^{N} \sum_{j=1}^{N} \left( W^T (\phi(I^l_i) - \phi(I^l_j)) \right)^2 \quad \text{(2.2)}$$

where $\phi(x)$ can be a linear or non-linear function of the input feature vectors. $\lambda$ controls the relative effect of the distance preserving and the class separability on the total optimization. A simple way of make use of class information is that $\delta(\omega_i, \omega_j)$ is set to one when $I_i$ and $I_j$ are from the same class and otherwise it is set to zero. This specific form makes sure the distance between data from the same class to be small. The iterative majorization algorithm is used to minimize the objective function.

In testing phase, for input image $I_{input}$, the transformed feature $\hat{I}_{input}$ is obtained by

$$\hat{I}_{input} = W^T \phi(I_{input}) \quad \text{(2.3)}$$

Euclidean distance is then computed for classification.

Their experimental evaluation shows that this method performs better than bicubic interpolation and a super-resolution method using sparse representation [24].

CLPM

In [10], a SR method using coupled locality preserving mappings (CLPM) is proposed. This method requires HR and LR image pairs for training. The main idea is to obtain coupled mappings which can project both HR and LR image features to a unified feature space so that direct comparison of HR and LR in that feature space could be possible.

The objective is to make sure the projections of LR images and their corresponding HR images would be as close as possible in the new unified feature space. Let $I^h$ and $I^l$ be the HR and LR images, this can be formulated as

$$J(f_L, f_H) = \sum_{i=1}^{N} \left\| P^T_L I^l_i - P^T_H I^h_i \right\|^2 \quad \text{(2.4)}$$
where $P_L$ and $P_H$ are the transformation matrix from LR and HR image space to the unified feature space. We can obtain the mappings by minimizing this formula. The gallery and probe images then will be mapped to a unified feature space where we can use nearest neighbor classifier for recognition.

This method has good performance on face recognition. Besides, it is more suitable for real-time systems because the test process can be very fast once the training part is done offline.

**Key-frames for long video sequences**

A face recognition system for long video sequences is presented in [12]. It selects key-frames and then applies hybrid super-resolution. Key-frames are chosen as follows. First, the most frontal images are selected using auto-associative memories. Then, a quality score is assigned to each frontal image by a face quality assessment (FQA) and the images with higher score are kept for next procedure. Lastly the selected images are aligned with the best one. The next step is hybrid super-resolution. Multiple images from the key-frames are used to construct a HR image. Multi layer perceptron (MLP) is applied then to further improve the constructed HR image. Finally, a face recognition algorithm that is sensitive to degradation in its input is used for testing the performance. This system has better results than Schultz[12] and Baker[15] and can be a good way to improve the recognition rate when long video sequences are available. Nonetheless, the performance of this system on images with lower solution is unknown since the LR images used in this paper are $46 \times 56$ pixels.

**DSR and NMCF**

In [27], a super-resolution method named discriminative super-resolution (DSR) for very low resolution situation that face region is smaller than $16 \times 16$ pixels is proposed. A new data constraint is developed to make sure that the constructed super-resolution image would be as close as possible to the corresponding high resolution image. Identity information of the subject is also used to improve the recognition accuracy. In [9], another SR method using nonlinear mappings on coherent features (NMCF) is proposed. Canonical correlation analysis is used to project HR and LR image pairs to a coherent feature space and then radial based functions is applied to find the mapping between them. These two methods will be discussed in detail in Section 3.2.
Algorithm introduction

3.1 Face recognition methods

3.1.1 Eigenface

Principal Component Analysis (PCA), or the eigenface method, is a widely used basic face recognition algorithm [21]. It provides a simple approach to extract the information contained in a face image to capture the variation of a collection of face images, and the information is used to encode and compare individual images.

Given a training set of $N$ images ($\{I\} = \{I_i\}_{i=1}^N$), first compute the mean image $\bar{I} = \frac{1}{N} \sum_{i=1}^N I_i$. Then subtract the mean from each face $\hat{I}_i = I_i - \bar{I}$. The covariance matrix is computed as following

$$ C = \frac{1}{N} \sum_{i=1}^N \hat{I}_i \hat{I}_i^T = AA^T \quad (3.1) $$

where the matrix $A = \{\hat{I}_i\}_{i=1}^N$. Then solve the eigenvalue problem

$$ \Lambda = V^T C V \quad (3.2) $$

where $V$ is the eigenvector matrix of $C$ and $\Lambda$ is the diagonal matrix containing eigenvalues of $C$ on its main diagonal. The eigenvectors are then sorted so that their corresponding eigenvalues are in descending order. These vectors are called eigenfaces, see Figure 3.1. Usually, the number of pixels in the images ($M$) would be relatively larger than the number of images in the training set ($N$). Thus, the complexity of calculation is reduced.

In testing phase, an input image $I_{input}$ is transformed to its eigenface feature $X_{input}$ by

$$ X_{input} = V^T (I_{input} - \bar{I}). \quad (3.3) $$

Then eigenface features $X_i$ and $X_j$ are compared using Euclidean distance or cosine distance. Three methods for distance measure is introduced in Ap...
Algorithm introduction

3.1.2 Fisherface

Fisher’s Linear Discriminant (FLD) method, or Linear Discriminant Analysis (LDA) selects projection vectors in such a way that the ratio of the between-class scatter and within-class scatter is maximized [3].

Given a training set of $N$ images ($\{I\} = \{I_i\}_{i=1}^N$), assume they belong to $m$ different classes $\{C_i\}_{i=1}^m$ that each class has $N_i$ images. Define the between-class scatter as

$$S_B = \sum_{i=1}^m N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

and the within-class scatter as

$$S_W = \sum_{i=1}^m \sum_{I_k \in C_i} (I_k - \mu_i)(I_k - \mu_i)^T$$

where $\mu$ is the mean image of training set, and $\mu_i$ is the mean of $i$th class. Then the optimal projection is selected to maximize the ratio of the between-class scatter and within-class scatter as following

$$W_{opt} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|}$$

(3.4)

where $W$ can be solved as the eigenvectors of

$$S_B W = \Lambda S_W W.$$  

(3.5)

However, $S_W$ is always singular because the rank of $S_W$ is at most $N - m$ which is much smaller than the number of pixels in an image in most cases. One way to solve this problem is Fisherface method, or PCA-LDA. This method firstly use PCA to reduce the feature dimensionality to $N - m$, and
then apply the standard FLD as discussed above to reduce the dimension to \( m - 1 \). Thus, the projection is computed as follows

\[
W_{\text{opt}} = W_{\text{pca}}^T W_{\text{fld}}^T.
\]

In testing phase, an input image \( I_{\text{input}} \) is projected to Fisherface feature space by

\[
X_{\text{input}} = W_{\text{opt}}(I_{\text{input}} - \mu).
\]

Different methods of distance measure can be used to compare the features while cosine distance which has the best performance is used in our experiments.

Because the Fisherface method is the only LDA method used in our experiments, we will call it LDA for the rest of this report.

### 3.1.3 LBP

Local Binary Pattern (LBP) operator was originally designed for texture description [1]. For a \( 3 \times 3 \) neighborhood, LBP assigns a label to each pixel except the center one according to their values and the value of the center pixel. More specifically, if the value of one pixel is smaller than the value of the center pixel, its label is assigned to 0; otherwise, it is assigned to 1. Thus, an 8 bit binary value is generated for this neighborhood. Then the histogram of the labels can be used as a texture descriptor. This \( 3 \times 3 \) neighborhood is also referred as \((8,1)\) neighborhood which means eight sampling points on a circle of radius of one. Besides, \((8,2)\) and \((16,2)\) neighborhoods are also commonly used.

A local binary pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or 1 to 0 when the bit pattern is considered circular. For example, 11111111 (0 transitions), 00000011 (2 transitions) and 11100001 (2 transitions) are uniform patterns while 11001001 (4 transitions) and 01100101 (6 transitions) are not. Using only uniform patterns for LBP means that each uniform pattern has a separate bin and all nonuniform patterns are assigned to one bin when computing the histogram. It has been demonstrated in [1] that uniform patterns account for 90 percent of all patterns in the \((8,1)\) neighborhood using images from the Facial Recognition Technology (FERET) database.

In face recognition process, the basic histogram can be extended to a spatially enhanced histogram. Each image is divided into several facial regions and the LBP histogram is computed for each region. Then the histograms are concatenated to build a global description of a face.

The common method for distance measure of LBP is the weighted Chi square distance which is defined as

\[
\chi^2_w(x, y) = \sum_{j,i} w_j \frac{(x_{i,j} - y_{i,j})^2}{x_{i,j} + y_{i,j}}
\]
where $i$ refers to $i$th bin in histogram corresponding to the $j$th region of the image, and $w_j$ is the weight of region $j$.

### 3.2 Super resolution methods

Two state-of-art super-resolution methods are used in our experiments.

#### 3.2.1 RL and DSR

A super-resolution method for very low resolution is proposed in [27]. Very low resolution (VLR) face recognition problem is defined as the situation that face region is smaller than $16 \times 16$ pixels in this paper. A new data constraint is proposed for relationship learning based super-resolution. Given a set of HR and VLR image pairs $\{I^h_i, I^l_i\}_{i=1}^N$, let $R$ be the relationship between HR and VLR space. The idea is to minimize the distance between HR and the space of VLR projected by $R$, see Equation (3.8).

$$ R = \arg \min_{R'} \sum_{i=1}^N \left\| I^h_i - R'I^l_i \right\|^2 \quad (3.8) $$

Then a SR image can be constructed from a given LR image $I_{input}$ by $I_{SR} = R I_{input}$. This method is called RL, which is used for reconstructing high-resolution images.

Furthermore, discriminative super-resolution (DSR) is proposed to acquire better results in face recognition. This method makes use of class information of the training data to make sure the reconstructed HR images are clustered with the images from the same class and far away from those from other classes. A distance function (Equation (3.9)) is introduced to the minimization problem.

$$ d(R) = \text{mean}\left( \left\| I^h_i - R'I^l_j \right\|^2 | \text{class}(I^h_i) = \text{class}(I^l_j) \right) - \text{mean}\left( \left\| I^h_i - R'I^l_j \right\|^2 | \text{class}(I^h_i) \neq \text{class}(I^l_j) \right) \quad (3.9) $$

Then the DSR formula integrates Equation (3.8) and (3.9)

$$ \hat{R} = \arg \min_{R'} \frac{1}{N} \sum_{i=1}^N \left\| I^h_i - R'I^l_i \right\|^2 + \gamma d(R') \quad (3.10) $$

where $\gamma$ is a constant to balance the above two terms. In our experiments, $\gamma$ is set to 1.

Thus, for a given image $I_{input}$, we first apply $I_{SR'} = \hat{R} I_{input}$, and then feed the SR image to recognition process.
3.2 Super resolution methods

3.2.2 NMCF

The super-resolution approach using nonlinear mappings on coherent features (NMCF) is proposed by Huang et al. in [9]. First, PCA features are extracted to reduce computational costs. Given a training set of HR and LR image pairs \( \{I^H, I^L\} = \{I^H_i, I^L_i\}_{i=1}^N \), PCA features of them are defined as \( X^H = \{x^H_i\}_{i=1}^N \) and \( X^L = \{x^L_i\}_{i=1}^N \) and extracted by

\[
x^H_i = (V^H)^T (I^H_i - \bar{I}^H),
\]

\[
x^L_i = (V^L)^T (I^L_i - \bar{I}^L)
\]

(3.11)

where \( V^H \) and \( V^L \) are the corresponding eigenvectors of the covariance matrices \((I^H - \bar{I}^H)(I^H - \bar{I}^H)^T\) and \((I^L - \bar{I}^L)(I^L - \bar{I}^L)^T\), and \( \bar{I}^H \) and \( \bar{I}^L \) are the mean of the HR and LR images respectively. Then Canonical Correlation Analysis (CCA) is used to project PCA features to a new coherent feature space. First we subtract the mean \( \bar{X} \) from the PCA features:

\[
\hat{X}^H = X^H - \bar{X}^H,
\]

\[
\hat{X}^L = X^L - \bar{X}^L
\]

(3.12)

To maximize the correlation between HR and LR coherent features \( C^H \) and \( C^L \), the correlation coefficient is defined as

\[
\rho = \frac{E[C^H C^L]}{\sqrt{E[(C^H)^2]E[(C^L)^2]}} = \frac{E[(W^H)^T \hat{X}^H (\hat{X}^L)^T W^L]}{\sqrt{E[(W^H)^T \hat{X}^H (\hat{X}^L)^T W^L]E[(W^L)^T \hat{X}^L (\hat{X}^L)^T W^L]}}
\]

(3.13)

where \( E[\cdot] \) stands for mathematical expectation. 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. Compute

\[
R_1 = C_{11}^{-1}C_{12}C_{22}^{-1}C_{21},
\]

\[
R_2 = C_{22}^{-1}C_{21}C_{11}^{-1}C_{12}.
\]

(3.14)

Then the base vector \( W^H \) is made of eigenvectors of \( R_1 \) and \( W^L \) is made of eigenvectors of \( R_2 \) when their corresponding eigenvalues are sorted in descending order. The last step of training is to use Radial Basis Functions (RBFs) to learn the relationship between the coherent features of HR and LR. RBFs are used to build up function approximations of the form

\[
f = \sum_{j=1}^N \omega_j \varphi(||t - t_j||)
\]

(3.15)
where the approximating function $f$ is represented as a sum of $N$ radial basis functions, each associated with a different center $t_j$ with weighting coefficient $\omega_j$. Multiquadric basis function

$$
\varphi(\|t - t_j\|) = \sqrt{\|t - t_j\|^2 + 1}
$$

is chosen for the implementation. The goal is to build the mapping between HR and LR coherent features $C^H$ and $C^L$. Thus, by setting $f = c_i^h$ and $t = c_i^l$, Equation (3.15) is represented as $C^H = W\Phi$, specifically

$$
[c_1^h, \ldots, c_n^h] = [\omega_1, \ldots, \omega_n] \begin{bmatrix}
\varphi(\|c_1^h - c_1^l\|) \cdots \varphi(\|c_n^h - c_1^l\|) \\
\varphi(\|c_1^h - c_2^l\|) \cdots \varphi(\|c_n^h - c_2^l\|) \\
\vdots \quad \vdots \quad \ddots \\
\varphi(\|c_1^h - c_n^l\|) \cdots \varphi(\|c_n^h - c_n^l\|)
\end{bmatrix}.
$$

So the weight matrix $W$ can be solved by

$$
W = C^H(\Phi + \tau I_{id})^{-1}
$$

where $\tau I_{id}$ is added because $\Phi$ is not always invertible. $\tau$ is a small positive value such as $10^{-3}$, and $I_{id}$ is identity matrix.

In testing stage, for HR gallery images $I^G = \{I_i^g\}_{i=1}^M$, coherent features are computed as following

$$
x_i^g = (V_H)^T(I_i^g - \bar{I}^H), \\
c_i^g = (W_H)^T(x_i^g - \bar{X}^H).
$$

For a LR probe image $I^p$ first we compute the coherent feature

$$
x_p^p = (V_L)^T(I_p^p - \bar{I}^L), \\
c_p^p = (W_L)^T(x_p^p - \bar{X}^L).
$$

Then apply the learnt mapping to obtain the corresponding SR features by

$$
c_{SR} = W : [\varphi(\|c_1^l - c_p^p\|) \cdots \varphi(\|c_N^l - c_p^p\|)]^T.
$$

Then the above features are feed to the nearest neighbor classifier for recognition

$$
\text{classify}(c_{SR}) = \min(\|c_{SR} - c_i^g\|^2).
$$
In this chapter, first the databases used in our experiments are introduced. Then we show the face recognition results of images with different resolution and evaluate the performance of two state-of-art super-resolution methods. In all the experiments, image identification is performed other than verification. That is, for an input probe image, the distances between it and each gallery images are computed in the recognition phase and it is assumed to have the same class as the image which has the smallest distance with it.

4.1 Database description

4.1.1 ORL

Olivetti Research Laboratory (ORL) database contains ten images of each of 40 distinct subjects. The images have slightly various poses, expressions and illumination conditions. The size of each image is $92 \times 112$ pixels. The ten images from the first subject is shown in Figure 4.1.

Figure 4.1: Images from the first subject in ORL.
4.1.2 FRGC v1.0

Face Recognition Grand Challenge version one (FRGC v1.0) database contains controlled images which were taken in a studio setting and uncontrolled images which were taken in varying illumination conditions. In our experiment, only 2820 controlled high-quality images are used. These images are all frontal images with two expressions (smiling and neutral). The images in Figure 4.2 are cropped according to the upper-left and lower-right coordinates provided in the database and turned into gray-scale.

![FRGC Sample Images](image)

Figure 4.2: FRGC sample images.

4.1.3 SCface

Surveillance Cameras Face (SCface) Database contains 4,160 static images of 130 subjects. They have five surveillance cameras namely cam1, cam2, cam3, cam4 and cam5. Cam1 and cam5 were also used for infrared (IR) night vision mode which named cam6 and cam7. These cameras each took one photos of each subject at three distances, 4.20 meters (distance1), 2.60 meters (distance2), and 1.00 meter (distance3). IR mug shots were taken by cam8. Another set of images were taken by a mug shot camera at day time. Nine images per subject provide nine discrete ranging from left to right profile in equal steps of 22.5 degrees including a mug shot frontal image at 0 degree [6]. Sample images from SCface are in Figure 4.3. Note that although images from distance3 are with higher resolution than distance1 and distance2, the pose variation is much more serious.

In [6], an experiment was conducted to evaluate the performance of SCface. They normalized images following standard procedure for face recognition. First, the color images were turned into grey-scale images. Then the positions of eyes were used for alignment. Finally, all images were cropped.
4.2 The influence of different image resolution

This experiment evaluates the influence of images with different resolution on FRGC v1.0 database.

4.2.1 Configuration

2820 high-quality images from FRGC v1.0 database were used in this experiment. The MATLAB imresize function with bicubic interpolation parameter was used to change resolution. Images with following resolution were created: $6 \times 6$, $8 \times 8$, $12 \times 12$, $16 \times 16$, $24 \times 24$, $32 \times 32$, $40 \times 40$, $48 \times 48$, $56 \times 56$, $64 \times 64$ and masked with an elliptical mask. 501 images were randomly chosen as training images from the FERET database, one image per person. The frontal images taken by mug-shot camera in the SCface were put into the gallery set. For experiments at night, cam8 was also used as gallery. Images from each camera were used as probe each time. The performance of baseline PCA was evaluated using 40% eigenvectors, which is a 200 dimensional subspace. The rank-1 recognition rate varies from 1% to about 8%.

In our experiments, the mug-shot frontal images and the images from cam1 to cam5 taken at day time are used.
Experiments

The 2820 images were divided into training, gallery and probe sets in different ways to find the best configuration:

1st: training set contains images from the first 80 persons and all images from the persons who have less than four images (including four). This training set contains 1228 images of 139 persons. Two images per person from the rest images are used as gallery images (242 images) and the others as probe images (1350 images).

2nd: training set contains 812 images of 109 persons from 1st training set. Gallery and probe sets are the same as the 1st configuration.

3rd: training set contains 400 images of 76 persons from 1st training set. Gallery and probe sets are the same as the 1st configuration.

Two face recognition methods LDA and LBP were used. Cosine angle was used to measure distance for LDA method while weighted Chi square was for LBP. Moreover, the $LBP_{u^2}^{(8,1)}$ operator was selected when using local binary pattern. $u^2$ stands for using only uniform patterns. And $(8, 1)$ neighborhood was used because $(8, 2)$ and $(16, 2)$ neighborhood are not applicable when the dimensionality of the smallest images is only $6 \times 6$. Images were divided into different number of windows: $2 \times 2$, $3 \times 3$, $4 \times 4$, $5 \times 5$, $6 \times 6$, and $7 \times 7$. All the images can be divided into $2 \times 2$ windows. However, $(8, 1)$ neighborhood requires that the size of each window is at least $3 \times 3$ pixels.
4.2 The influence of different image resolution

So $3 \times 3$ windows cannot be used for images with resolution lower than $9 \times 9$. And $4 \times 4, 5 \times 5, 6 \times 6$, and $7 \times 7$ windows cannot be used for images with resolution lower than $12 \times 12, 15 \times 15, 18 \times 18$, and $21 \times 21$ respectively. The weight of each window was assigned based on the recognition rate of performing LBP on training sets using only one window each time.

4.2.2 Experimental results

LDA

Proper number of eigenvectors for both PCA and LDA part must be chosen to obtain better results of LDA method. All combination of PCA and LDA dimensionalities were tried on resolution $32 \times 32$ and $64 \times 64$ for 1st configuration (see Figure 4.5). The recognition accuracy reach a maximum number 85.11% when PCA/LDA feature dimensionalities are 350/120 or 430/130 for resolution $32 \times 32$. And the highest rate of $64 \times 64$ is 84% when PCA dimensionality is 350 and LDA dimensionality is 130.

![Figure 4.5: All combination of PCA and LDA dimensionalities. Left: 32×32; right: 64×64.](image)

According to the results shown above, the number of PCA vectors were chosen as the minimum of 350 and the image dimensionality and the number of LDA vectors were chosen as the minimum of the number of classes minus one and the image dimensionality. Then we applied LDA on the three different dataset configurations.

As we can see in Figure 4.6, more images for training improves the system performance. Besides, the highest recognition rates are obtained when the image resolution is $12 \times 12$ or $16 \times 16$ for all three configurations. The accuracy drops a bit as resolution become higher while a sharp decrease appears from $12 \times 12$ to $6 \times 6$. 
LBP

Then we performed LBP using different number of windows and compared the results with LDA method on 1st configuration which gives the best results according to the previous experiment, see Figure 4.7.

The performance of LBP method is largely influenced by how many windows an image is divided into. Dividing images into more regions would lead to better results, but this improvement become smaller as the number of windows increases. For instance, the results of $3 \times 3$ windows are more than 10% better than the results of $2 \times 2$ windows while the recognition rate only improves for about 2% from $6 \times 6$ windows to $7 \times 7$ windows. Furthermore, the performances of all of them are not good at lower resolution. Another finding is that LDA is less sensitive with resolution changes than LBP. Though the LBP results remains stable and LDA is slightly worse than LBP $7 \times 7$ when image resolution is larger than $32 \times 32$, LDA has much better performance for lower resolution. What’s more, LBP is slower than
LDA especially when more windows are divided because the local binary features must be computed for each window and for each image.

Two more experiments were performed using LBP to find out the influence of weight in distance measure and the difference between using uniform patterns and using all patterns (Figure 4.8). The Chi square distance with weight has 1% to 3% better results than these without weight. And using all patterns instead of only uniform patterns slightly improves the results (except for $7 \times 7$), but it results in much longer computation time.

![Figure 4.8: LBP extra experiments. Left: with weight vs. without weight; right: all patterns vs. only uniform patterns.](image)

### 4.2.3 Discussion

In general, as the resolution become lower, the recognition rate decreases. LDA performance decreases slightly but still remains the same level when image resolution is higher than $16 \times 16$ and LBP performance becomes stable when image resolution is higher than $32 \times 32$. A sharp decrease happening when images are smaller than $12 \times 12$ pixels indicates that some important information for LDA recognition is missing when a image is less than $12 \times 12$ pixels.

### 4.3 Surveillance camera results

In Section 4.2, the low-resolution images were generated by downsampling and smoothing the high-resolution images. But there are some other factors that may influence the quality of images captured at a distance. In this section, we use images taken by surveillance cameras and evaluate the recognition performance.
4.3.1 Configuration

The images of surveillance from cam1 to cam5 and the frontal images of mug-shot camera from SCface database were used. There are 130 subjects in SCface database. The mug-shot camera took one frontal image of each subject. And the five surveillance cameras took one image per person per distance. So we have 650 images for each distance and 130 mug-shot frontal images. 2820 high-quality cropped images of 260 persons from FRGC v1.0 database were also used.

We have three dataset configurations in this section as well described as follows.

1st: 2820 high-quality images from FRGC are used as training images. Gallery images are the frontal images from SCface database and the images taken by camera 1 to 5 are probe images.

2nd: only images from SCface database are used for this configuration. Images taken by one camera are used as probe images at a time; images from the other four cameras are used as training set both separately and all together for each distance. Gallery images are the mug shot frontal images.

3rd: only images from SCface database are used for this configuration. Images are divided into four datasets: distance1, distance2, distance3 and mug shot frontal images. The first three sets are used for training and testing in turn. And the fourth set is gallery.

Two face recognition methods LDA and LBP were used. Cosine angle was used to measure distance for LDA the same as in the previous section. As demonstrated in Figure 4.8, the Chi square distance provides similar results with weight or without weight. Thus, Chi square (or weighted Chi square with weights set to one) was used for LBP in the following experiments. The other parameters of LBP were the same as in Section 4.2.

4.3.2 Experimental results

Face cropping

The first step was to crop faces from the original images of SCface database. By using eye coordinates information provided by the database, all images were scaled and rotated so that the distance between the eyes is always the defined number, and the eyes are lined up in a straight line. In order to find out the proper way of cropping faces, four different cropping configurations were tested, each of them cropped a different range of face region. The performance of each cropping configuration was tested by both LDA (FRGC as training) and LBP. Information regarding the most effective cropping configuration is provided in Table 4.1. The cropped images are shown in Figure 4.9.
4.3 Surveillance camera results

<table>
<thead>
<tr>
<th>distance</th>
<th>image size</th>
<th>left eye coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance1</td>
<td>12 pixels</td>
<td>$27 \times 31$</td>
</tr>
<tr>
<td>distance2</td>
<td>22 pixels</td>
<td>$45 \times 45$</td>
</tr>
<tr>
<td>distance3</td>
<td>36 pixels</td>
<td>$81 \times 82$</td>
</tr>
</tbody>
</table>

Table 4.1: Information about selected cropped images

Figure 4.9: SCface cropped images. First row: distance1; second row: distance2; third row: distance3; last row: mug shot frontal images.

FRGC as training set

The proper number of PCA/LDA vectors still needs to be found for LDA method. We only tested the influence of PCA vectors because using maximum number of LDA vectors can be good enough according to the results of previous experiments. The FRGC images are used for training, and mug-shot frontal images are used as gallery. The images from each distance are probe images.

We can see from Figure 4.10 (a) that the identification rate of distance3 reaches the highest number of 8.15% when PCA dimensionality is 50 and then keeps dropping after 200. This trend is quite similar to both distance2 and distance1. So only the results between 10 and 150 for these two distances are shown in Figure 4.10 (b) and (c). Then we chose to use PCA/LDA pair 50/50 for the following experiments.

Then the experiments were done using images from each camera as probe set. This configuration is similar to the configuration in [6]. The difference is that, in [6], they use images from the FERET database for training and PCA for recognition. We compared the performance of our experiment with the results in [6] in Table 4.2. The PCA results are from the experiments in [6] and LDA and LBP results are from our experiment. Note that LBP
Experiments 4

Figure 4.10: Results of LDA tests of different PCA dimensionality.

<table>
<thead>
<tr>
<th></th>
<th>distance1</th>
<th>distance2</th>
<th>distance3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>2.3</td>
<td>3.1</td>
<td>1.5</td>
</tr>
<tr>
<td>LDA</td>
<td>10.0</td>
<td>7.0</td>
<td>6.2</td>
</tr>
<tr>
<td>LBP</td>
<td>7.0</td>
<td>8.5</td>
<td>5.4</td>
</tr>
</tbody>
</table>

Table 4.2: Comparison of PCA[6], LDA and LBP. Identification rate[%].

do not need a training set, so the LBP results are not affected by which
4.3 Surveillance camera results

training set was used in this section. The performance of LDA and LBP is better than PCA, but the conclusion is similar: the distance influence the results, but results from distance 3 are not as good as distance 2. This is because images taken at distance 3 are more of the top part of the head which can be seen in Figure 4.9. Furthermore, LBP outperforms LDA at distance 2 and distance 3.

2nd configuration

The experiments discussed above were using FRGC as training set. The performance was also evaluated when images from SCface are used as training images as described in the 2nd configuration. For example, if cam 5 is used as testing set, the training sets are cam 1, cam 2, cam 3, cam 4, and all images from cam 1 to cam 4. Both training and testing images are from the same distance, except gallery images are always the mug-shot frontal images. This experiment is only applied on LDA because the difference between the two configurations is the training set used and LBP does not need a training set.

<table>
<thead>
<tr>
<th></th>
<th>CAM1t</th>
<th>CAM2t</th>
<th>CAM3t</th>
<th>CAM4t</th>
<th>CAM5t</th>
<th>Allt</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM1</td>
<td>–</td>
<td>10.00</td>
<td>10.00</td>
<td>7.69</td>
<td>7.69</td>
<td>19.23</td>
</tr>
<tr>
<td>CAM2</td>
<td>3.08</td>
<td>–</td>
<td>2.31</td>
<td>3.85</td>
<td>3.85</td>
<td>22.31</td>
</tr>
<tr>
<td>dist1</td>
<td>CAM3</td>
<td>4.62</td>
<td>4.62</td>
<td>–</td>
<td>3.85</td>
<td>4.62</td>
</tr>
<tr>
<td></td>
<td>CAM4</td>
<td>10.00</td>
<td>8.46</td>
<td>7.69</td>
<td>–</td>
<td>9.23</td>
</tr>
<tr>
<td></td>
<td>CAM5</td>
<td>7.69</td>
<td>7.69</td>
<td>6.92</td>
<td>9.23</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14.62</td>
</tr>
<tr>
<td></td>
<td>CAM1</td>
<td>–</td>
<td>6.15</td>
<td>8.46</td>
<td>9.23</td>
<td>4.62</td>
</tr>
<tr>
<td></td>
<td>CAM2</td>
<td>3.08</td>
<td>–</td>
<td>4.62</td>
<td>6.15</td>
<td>5.38</td>
</tr>
<tr>
<td>dist2</td>
<td>CAM3</td>
<td>5.38</td>
<td>4.62</td>
<td>–</td>
<td>5.38</td>
<td>3.08</td>
</tr>
<tr>
<td></td>
<td>CAM4</td>
<td>7.69</td>
<td>10.77</td>
<td>10.00</td>
<td>–</td>
<td>5.38</td>
</tr>
<tr>
<td></td>
<td>CAM5</td>
<td>6.92</td>
<td>3.85</td>
<td>4.62</td>
<td>6.92</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.92</td>
</tr>
<tr>
<td></td>
<td>CAM1</td>
<td>–</td>
<td>3.08</td>
<td>6.15</td>
<td>6.92</td>
<td>6.15</td>
</tr>
<tr>
<td></td>
<td>CAM2</td>
<td>6.92</td>
<td>–</td>
<td>3.08</td>
<td>4.62</td>
<td>6.15</td>
</tr>
<tr>
<td>dist3</td>
<td>CAM3</td>
<td>4.62</td>
<td>3.85</td>
<td>–</td>
<td>2.31</td>
<td>4.62</td>
</tr>
<tr>
<td></td>
<td>CAM4</td>
<td>9.23</td>
<td>10.77</td>
<td>10.00</td>
<td>–</td>
<td>6.15</td>
</tr>
<tr>
<td></td>
<td>CAM5</td>
<td>6.92</td>
<td>7.69</td>
<td>5.38</td>
<td>6.15</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 4.3: 2nd configuration. Identification rate[%]. CAM1t stands for using images from cam 1 for training.

As can be seen in Table 4.3, images taken by only one camera are not enough for training, the results of using images from all the other cameras for training are much better than using images from only one camera. The results of distance 2 are slightly better than others, but they do not deviate a
lot. Besides, comparing Table 4.3 with Table 4.2, using images from SCface as training set gives much better results than using FRGC.

3rd configuration

In [27], 10 images (5 LR images and 5 HR images) per subject from SCface are used and the results of eigenface, kernel PCA and SVM are all around 15% without super-resolution and 20% with the proposed super-resolution method. In order to compare to their results, we scaled the images of distance 3 so that the distance between eyes is the same as distance 1 (12 pixels). Then the same cropping configuration is applied to obtain images of the same size. The images from distance 3 are used both as training images and gallery images, and images from distance 1 are used for testing. The identification rate is better: 23.5%. Then this experiment is extended on all three distances and also use mug shot frontal images as gallery, see Table 4.4.

<table>
<thead>
<tr>
<th>train</th>
<th>test</th>
<th>accuracy1 [%] (^a)</th>
<th>accuracy2 [%] (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance 3</td>
<td>distance 1</td>
<td>23.54</td>
<td>16.15</td>
</tr>
<tr>
<td>distance 3</td>
<td>distance 2</td>
<td>30.62</td>
<td>18.46</td>
</tr>
<tr>
<td>distance 2</td>
<td>distance 1</td>
<td>26.00</td>
<td>14.62</td>
</tr>
<tr>
<td>distance 2</td>
<td>distance 3</td>
<td>26.31</td>
<td>15.38</td>
</tr>
<tr>
<td>distance 1</td>
<td>distance 2</td>
<td>27.54</td>
<td>22.00</td>
</tr>
<tr>
<td>distance 1</td>
<td>distance 3</td>
<td>23.69</td>
<td>11.38</td>
</tr>
</tbody>
</table>

\(^a\) gallery = training  
\(^b\) gallery = mug shot frontal images

Table 4.4: 3rd configuration

Obviously, the results of using same dataset for training and gallery is better than using mug-shot frontal images for gallery. Moreover, distance 2 as probe set outperforms using the other two as probe. And the worst results are obtained when using images from distance 1 and distance 3 in one experiment.

4.3.3 Discussion

SCface is a very difficult database, from all the experiments described above, the identification rates are no more than 30%. Though images from distance 3 have higher resolution than the other two, they don’t have better results than distance 2 due to the pose variation problem. And as the result of the difference between the image resolution and pose variation of distance 1 and distance 3, the results are disappointing when using one of them for training and the other one for testing. Moreover, using images from the
same database for training and testing has better performance than using other databases for training when enough training images are available.

4.4 Super resolution

After evaluating the recognition performance of low resolution images in the previous sections, effort is needed to improve these results. An intuitive way is to apply super-resolution on the low-resolution images to enhance the image quality. In this section, we applied two state-of-art super-resolution approaches and compared their performance.

4.4.1 Configuration

2820 images from FRGC v1.0 database were used in this experiment. Because the background and non-face part of the head in the images would affect the result of super-resolution, the images used in Section 4.2 were cropped again to obtain the face region only. Then the MATLAB imresize function with bicubic interpolation parameter is used to rescale these cropped images to $70 \times 60$, $56 \times 48$, $42 \times 36$, $28 \times 24$, $14 \times 12$ and $7 \times 6$ pixels, see Figure 4.11. The images were divided into training, gallery and probe sets the same way as the 1st configuration in Section 4.2: training set contained 1228 images of 139 persons. Two images per person from the rest images were used as gallery images (242 images) and the others were probe images (1350 images).

![Figure 4.11: FRGC sample images. First row: 56 x 48, second row: 42 x 36, third row: 14 x 12, last row: 7 x 6.](image)

Images from SCface dataset were also used. 5 images taken by 5 surveillance cameras for 130 persons at distances of 4.20, 2.60 and 1.00 meters
(named distance1, distance2 and distance3) were used, that’s $5 \times 130 \times 3 = 1950$ images. The mug-shot frontal images, one image per person, are used as gallery. The distances between eyes are 12, 24 and 36 pixels for cropped images of distance1, distance2 and distance3, and the image sizes are $31 \times 27$, $62 \times 54$ and $93 \times 81$ pixels respectively. Some samples of the cropped images are in Figure 4.12. The probe set is always the images from distance1.

Two super-resolution algorithms RL/DSR and NMCF described in Section 3.2 were evaluated in this section. Three recognition methods, namely PCA, LDA and LBP were used. Cosine angle was used to measure distance for PCA and LDA and Chi square was used for LBP.

### 4.4.2 Experimental results

#### RL and DSR

First, the reconstruction of super-resolution images using RL (Equation (3.8)) was performed on FRGC dataset. The SR reconstruction requires HR and LR image pairs for training. In this experiment, the FRGC training set which contains 1228 images were used to train the SR system, and the SR images were reconstructed from the images from the gallery set. Images with resolution $70 \times 60$ were used as HR images. LR images were the images with resolution $42 \times 36$, $28 \times 24$, $14 \times 12$ and $7 \times 6$. SR images of these LR images were generated respectively. The results can be seen in Figure 4.13. The SR images of $42 \times 36$ and $28 \times 24$ are very similar to original HR images. Then the SR images become blurred as the resolution of LR decreases. SR images of $7 \times 6$ do not look like the original HR because the
LR images can provide little information.

SR images were also reconstructed using DSR (Equation (3.10)). As we can see in Figure 4.14, the SR images are darker than those in Figure 4.13. But the SR images become similar to those in Figure 4.13 when the resolution of LR decreases.

Then we evaluated the performance for recognition of DSR using face recognition methods PCA, LDA and LBP. Images with resolution $7 \times 6$ and $14 \times 12$ are used as LR images. For LR $7 \times 6$, the performance was evaluated using $70 \times 60$, $56 \times 48$, $42 \times 36$, $28 \times 24$ and $14 \times 12$ as HR images; for LR $14 \times 12$, HR images were images with higher resolution: $70 \times 60$, $56 \times 48$, $42 \times 36$ and $28 \times 24$. The training set is used for both training of LDA and PCA and learning the relationship between HR and LR pairs. Two different experiment configurations were applied. For the first configuration which is...
Figure 4.14: FRGC SR image construction using DSR.

named DSR1, the relationship of LR images and the corresponding HR images was computed by HR and LR training set, then the super-resolution features of LR training, gallery and probe images were reconstructed and fed to face recognition system. In real situation, this can be the case only when we want to know whether the images taken by two surveillance cameras belong to the same person. In most cases, the training and gallery images are high-resolution images which are acquired in advance and input LR images are compared with HR gallery images after a SR process. So in the second configuration DSR2, the training and gallery images for recognition were HR images, SR features were only computed for the LR probe set. $7 \times 7$ windows are used for LBP, except $7 \times 6$ and $14 \times 12$ for which $2 \times 2$ windows and $4 \times 4$ windows were applied respectively. The LDA vectors were also
selected to obtain the best results. And the best results of SR with the first configuration were obtained when using the same number of LDA vectors as the corresponding LR while the second configuration is different, see Table 4.5 for detail. The recognition results are shown in Figure 4.15, Figure 4.16 and Figure 4.17.

<table>
<thead>
<tr>
<th>resolution</th>
<th>LDA1(^a)</th>
<th>LBP windows</th>
<th>LDA2(^b)</th>
<th>LDA3(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70 × 60</td>
<td>420</td>
<td>7 × 7</td>
<td>70</td>
<td>280</td>
</tr>
<tr>
<td>56 × 48</td>
<td>430</td>
<td>7 × 7</td>
<td>70</td>
<td>320</td>
</tr>
<tr>
<td>42 × 36</td>
<td>330</td>
<td>7 × 7</td>
<td>70</td>
<td>270</td>
</tr>
<tr>
<td>28 × 24</td>
<td>290</td>
<td>7 × 7</td>
<td>230</td>
<td>200</td>
</tr>
<tr>
<td>14 × 12</td>
<td>150</td>
<td>4 × 4</td>
<td>90</td>
<td>–</td>
</tr>
<tr>
<td>7 × 6</td>
<td>40</td>
<td>2 × 2</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

\(^a\) LDA1: the LDA dimensionality when using original images for recognition.
\(^b\) LDA2: the LDA dimensionality when using images with different resolution as HR and 7 × 6 as LR for the 2nd SR configuration.
\(^c\) LDA3: the LDA dimensionality when using images with different resolution as HR and 14 × 12 as LR for the 2nd SR configuration.

Table 4.5: Details of LBP and LDA parameters.

The identification rates of original images using PCA are lower than 20%. The results of DSR1 are much better and all higher than the original HR images. The identification rate of 7 × 6 using super resolution are around 25% and of 14 × 12 using super resolution are about 30%. Nonetheless, the DSR2 results are worse than original HR results but better than original LR results.

Figure 4.15: DSR results on FRGC. PCA results. Left: input resolution 7 × 6; right: input resolution 14 × 12.
Figure 4.16: DSR results on FRGC. LBP results. Left: input resolution $7 \times 6$; right: input resolution $14 \times 12$.

Figure 4.17: DSR results on FRGC. LDA results. Left: input resolution $7 \times 6$; right: input resolution $14 \times 12$.

The LBP results of original $7 \times 6$ and $14 \times 122$ are relatively low, but the results reach 80% when image resolution is larger than $28 \times 24$. The two DSR performance are both better than the original LR but worse than original HR. The results of DSR1 are close to HR results, but the results of DSR2 are much lower and close to original LR results.

LDA is the only one that can’t be improved by DSR super resolution. The LDA results of original images are above 90% for resolution larger than $14 \times 12$ while they are slightly above 76% for $7 \times 6$. The recognition rates of DSR1 are a bit lower than their original LR images, whereas the results of DSR2 are much worse.

In addition, we also tried to use the SR images reconstructed using RL for recognition. The results are 3% to 10% worse than DSR results.

The performance of RL and DSR was also evaluated using images from SCface database. The super-resolution image reconstruction is shown in Figure 4.18. FRGC training set was used to train the SR system and the SR images were reconstructed from images from distance1. As we can see,
the quality of the SR images are not good, some of them even does not look like a human face.

Figure 4.18: SCface SR image construction using FRGC for training. (1st and 3rd row: distance1 LR images; 2nd and last row: corresponding SR images.)

In case that using training set from SCface may provide better results, three other configurations for SR construction were performed:

‘Distance2 SR’: the images from distance2 were used as HR images and images from distance1 were used as LR images to train the SR system. Then the SR images were reconstructed also from images of distance1.

‘Distance3 SR’: similar to distance2 SR, images from distance3 were used as HR images instead of distance2.

‘Distance2+3 SR’: the images from distance2 and distance3 were used as HR. The LR images were generated by downsampling the HR images. SR images were reconstructed from images of distance1 as well.

The results can be seen in Figure 4.19. Though the image quality of ‘distance2 SR’ and ‘distance3 SR’ are better, the SR images especially their poses are more similar to the HR images instead of LR images. The ‘distance2+3 SR’ images are better than the images in Figure 4.18. However, the images are blurred due to the misalignment that the scale and pose of the input images are different although they have already been aligned using eye coordinates.

Then we conducted recognition experiment using the above three super-resolution configurations and DSR is used for super-resolution. PCA and LDA were used for face recognition. The detailed configuration and results are shown in Table 4.6 and Figure 4.20 is the corresponding histogram. Note that images from FRGC are used to train the recognition system of PCA and LDA because images from SCface are not enough. Gallery images were those frontal images taken by mug-shot camera. In Figure 4.20, the first
Figure 4.19: SCface SR image construction 2. 1st, 4th and 7th row: distance1 LR images; 2nd row: distance2 SR; 3rd row: distance2 HR images; 5th row: distance3 SR; 6th row: distance3 HR images; 8th row: distance2+3 SR.

<table>
<thead>
<tr>
<th>HR&amp;LR</th>
<th>train</th>
<th>gallery</th>
<th>probe</th>
<th>LDA [%]</th>
<th>PCA [%]</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>dist3&amp;1</td>
<td>FRGC</td>
<td>frontal</td>
<td>dist1</td>
<td>8.31</td>
<td>7.23</td>
<td>dependent</td>
</tr>
<tr>
<td>dist2&amp;1</td>
<td>FRGC</td>
<td>frontal</td>
<td>dist1</td>
<td>15.85</td>
<td>12.31</td>
<td>dependent</td>
</tr>
<tr>
<td>2+3 pairs</td>
<td>FRGC</td>
<td>frontal</td>
<td>dist1</td>
<td>10.46</td>
<td>8.31</td>
<td>independent</td>
</tr>
<tr>
<td>no SR</td>
<td>FRGC</td>
<td>frontal</td>
<td>dist2</td>
<td>15.85</td>
<td>10.00</td>
<td>original LR</td>
</tr>
<tr>
<td>no SR</td>
<td>FRGC</td>
<td>frontal</td>
<td>dist3</td>
<td>7.38</td>
<td>5.23</td>
<td>original HR</td>
</tr>
</tbody>
</table>

Table 4.6: DSR results on SCface

three groups of bars are SR results and the last three are original results.

The results of ‘distance2 SR’ are better than others. The results of ‘distance2+3 SR’ are also better than the original results of distance1 and distance3. However, the improved results are still very low, from 10% to 15%.
Another super-resolution method NMCF was also evaluated in our experiments. First it was tested on ORL database using the same configuration as in [9]. There are 10 images of 40 subjects in ORL. 5 images were used for training and the other 5 images were used for testing. All the images were resized to $32 \times 32$ as high-resolution images and $8 \times 8$ pixels as low-resolution images, see Figure 4.21.

![Figure 4.21: ORL sample images. First row: $32 \times 32$ train; second row: $8 \times 8$ train; last row: $8 \times 8$ test.](image)

The number of PCA features of both HR and LR must be selected for NMCF, and the number of CCA features is the same as the smaller one of them. In Figure 4.22 we can see the influence of different PCA dimensionalities. Different color stands for different PCA dimensionalities of LR.
features and x axis is PCA dimensionality of HR. The best result 90% is obtained when the HR dimensionality is 160 and the LR dimensionality is 30. Then the best result was selected and compared with the results in [9] in Figure 4.23. From our results, the identification rate of PCA of 8 × 8 is 83.5% and 32 × 32 is 85%, the corresponding PCA dimensionality are 20 and 50 respectively. Our recognition rates are a bit lower than the results in [9] maybe because our training and testing set contains different images with the configuration in [9]. Nevertheless, our conclusions are the same: the NMCF outperforms PCA of both original HR and LR.

Figure 4.22: NMCF results of different PCA dimensions on ORL.

Figure 4.23: NMCF results compared with [9].

Then NMCF was evaluated on FRGC database. The experiment configuration was the same as previous experiments on FRGC. Images with resolution 7 × 6 and 14 × 12 were used as LR images, the rest were used as HR. The maximum number of PCA vectors of LR was chosen for the
4.4 Super resolution

experiment. And the influence of different numbers of PCA vectors of HR was tested. The results of using images with resolution $54 \times 48$ as HR images and images with resolution $7 \times 6$ and $14 \times 12$ as LR images are shown in Figure 4.24. For each experiment, the number of PCA vectors which leads to best result is selected, see Table 4.7.

![Figure 4.24: NMCF results of different PCA dimensions on FRGC. Left: LR $7 \times 6$, HR $54 \times 48$; right: LR $14 \times 12$, HR $54 \times 48$.](image)

<table>
<thead>
<tr>
<th>Resolution</th>
<th>70 $\times$ 60</th>
<th>56 $\times$ 48</th>
<th>42 $\times$ 36</th>
<th>28 $\times$ 24</th>
<th>14 $\times$ 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR $7 \times 6$</td>
<td>1190</td>
<td>1190</td>
<td>1180</td>
<td>620</td>
<td>160</td>
</tr>
<tr>
<td>LR $14 \times 12$</td>
<td>1060</td>
<td>1060</td>
<td>1060</td>
<td>670</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 4.7: Number of PCA vectors chosen for FRGC

![Figure 4.25: Face recognition results of FRGC using NMCF.](image)

The results on FRGC database were compared with the performance of DSR in Figure 4.25. DSR1 uses LR images for training and gallery while DSR2 uses HR training and gallery sets. As we can see, the SR results of NMCF are better than original LR and HR results. And NMCF outperforms
the DSR method when HR images are larger than \(42 \times 36\).

Figure 4.26: NMCF results of different PCA dimensions on SCface. (a) distance2 SR.

Figure 4.27: NMCF results of different PCA dimensions on SCface. (b) distance3 SR.

This method was also tested on SCface database. The SR configurations were the same as the previous experiments using DSR which were ‘distance2 SR’, ‘distance3 SR’ and ‘distance2+3 SR’. NMCF does not need extra training set for recognition. The gallery was mug-shot frontal image set and LR probe was distance1 as well. Figure 4.26, Figure 4.27 and Figure 4.28 show the NMCF results of different PCA dimensions. The results of the three configurations have similar trend that as both HR and LR dimensionalities increases the identification rate drops. Thus, the best results are selected and the detailed information are in Table 4.8. The NMCF performance is compared with DSR using PCA and original PCA results in Figure 4.29, refer to Table 4.6 for the experiment information of DSR and original PCA. Generally, NMCF outperforms DSR. The first two ‘dist2’ and ‘dist3’ have good results, but they are with dependent configurations, im-
4.4 Super resolution

Figure 4.28: NMCF results of different PCA dimensions on SCface. (c) distance2+3 SR.

Ages from distance1 were used as both LR images for SR and recognition probe set. The identification rate is much lower when using the SR configuration ‘distance2+3 SR’. The result of NMCF is a bit higher than DSR and the original result of distance1 and distance3, but is lower than the original result of distance2.

<table>
<thead>
<tr>
<th>NMCF</th>
<th>HR</th>
<th>LR</th>
<th>gallery</th>
<th>probe</th>
<th>dimension</th>
<th>result[&quot;]</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Dist3’</td>
<td>dist3</td>
<td>dist1</td>
<td>frontal</td>
<td>dist1</td>
<td>80</td>
<td>16.62</td>
</tr>
<tr>
<td>‘Dist2’</td>
<td>dist3</td>
<td>dist1</td>
<td>frontal</td>
<td>dist1</td>
<td>90</td>
<td>20.92</td>
</tr>
<tr>
<td>‘2+3’</td>
<td>dist2+3</td>
<td>dist2+3</td>
<td>frontal</td>
<td>dist1</td>
<td>440</td>
<td>8.92</td>
</tr>
</tbody>
</table>

Table 4.8: NMCF results on SCface

Figure 4.29: Comparison of NMCF and DSR on SCface.
4.4.3 Discussion

From the experiments discussed above, we come to the conclusion that both DSR and NMCF can improve the recognition accuracy of PCA and NMCF outperforms DSR. On the other hand, DSR can be applied to different face recognition methods such as LBP which has better results than PCA on FRGC database. However, DSR can not improve LDA performance on FRGC. DSR has much better performance using LR images for training and gallery than using HR images for all the three recognition methods. In addition, SCface is very challenging. The improvement of both SR methods is limited especially for independent configuration. The performance may be better if more images are available. Experiments with other different configurations including using images from distance2 and their corresponding LR images created by downsampling were also conducted, but there was no improvement at all. And the ‘distance2+3 SR’ does improve the recognition performance, though the SR training set is consisted with images from different distance. So we speculate that the results would be better if we double the number of images in this database. Pose variation might be a big problem as well. One possible approach for future work is to apply better alignment on SCface images.
Conclusion

In this report, the performance of face recognition using images with different resolution is evaluated using PCA, LDA and LBP. For experiments on FRGC v1.0 database in which the LR images are generated by downsampling original images, the recognition rate is steady for images with resolution higher than a certain number (12 × 12 pixels for FRGC) while the accuracy declines suddenly when image resolution is lower. Obviously, the important information for classification is missing when images are so small. For surveillance image database such as SCface, low resolution is not the only problem of the images captured at a distance. Other problems, take pose variation for example, also influence the performance. The recognition rates are all very low (no more than 30%). Images from distance2 (2.60 meters) have better results than images from distance1 (4.20 meters) and distance3 (1.00 meters) because they have higher resolution than images from distance1 and less pose variation than images from distance3. Moreover, using images from SCface for training has better performance than using images from other databases for recognition tests on SCface, although SCface doesn’t contain great amount of images. And the considerable difference between distance1 set and distance3 set leads to disappointing results when using one of them as training set and the other one as testing set. Thus, it is better to have similar datasets for training and testing.

Two state-of-art super-resolution methods DSR and NMCF for improving the low-resolution face recognition performance are compared. Both methods require a training set containing HR and LR image pairs to learn a relationship or mapping from LR feature space to HR feature space, and the SR images or features are reconstructed by this relationship or mapping. NMCF extracts the PCA features from face images and then applies CCA and RBFs to obtain a mapping from LR features to HR features in a coherent feature space. Nearest neighbor classifier is used for NMCF. DSR is more flexible. It solves an objective function for the relationship between LR and HR images to make sure the reconstructed SR image close
Conclusion

to the HR images in the same class and far away from the HR images in
different classes. The relationship is applied on the input images to obtain
the corresponding SR features. Then different recognition methods can be
used for classification. To apply DSR successfully, LR dimensionality must
smaller than the number of images in the training set. When HR training
and gallery sets are used, NMCF results are better than DSR results and
the results of using HR images while DSR results are worse than the results
of using HR images. NMCF performs better when the HR dimensionality
is large enough (larger than $42 \times 36$ pixels in our experiments). DSR re-
results are better when LR images are used for training and gallery. DSR can
improve LBP as well. On the contrary, LDA performance on FRGC can’t
be improved by DSR. One possible reason is that LDA already makes use
of the class information while PCA and LBP do not. So the discrimination
of DSR doesn’t make much sense to LDA. In the experiments on SCface,
the improvement of both DSR and NMCF are limited especially for inde-
pendent configuration. In our experiments, the images were aligned using
the eye coordinates provided in SCface database, but the scale of the faces
were different and nothing was done to deal with the pose problem. Better
methods to align the images is expected to improve the recognition perfor-
ance. Besides, for such a difficult database, more images are needed for
the training of both the SR and recognition systems of existing methods.
Appendix

Distance measure

There are three popular methods to measure the distance between two input vectors $x$ and $y$ [22].

L1: city block distance. The distance is measured as the sum of the absolute difference between the two vectors.

$$d(x, y) = |x - y| = \sum_{i=1}^{k} |x_i - y_i| \quad (A.1)$$

L2: Euclidean distance (squared). The Euclidean distance between two points is the length of the line segment connecting them.

$$d(x, y) = \|x - y\|^2 = \sum_{i=1}^{k} (x_i - y_i)^2 \quad (A.2)$$

Cosine angle: negative angles between image vectors.

$$d(x, y) = -\frac{x \cdot y}{\|x\| \cdot \|y\|} = -\frac{\sum_{i=1}^{k} x_i y_i}{\sqrt{\sum_{i=1}^{k} x_i^2} \sqrt{\sum_{i=1}^{k} y_i^2}} \quad (A.3)$$
Bibliography


