School of Electrical Engineering and Computing
Department of Computing

Face Hallucination with Application in Far Distance Face Recognition

Xiang Xu

This thesis is presented for the Degree of
Doctor of Philosophy
at
Curtin University

March 2014

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

XIANG XU

Date

Abstract

Face recognition becomes challenging as the images have lower resolutions. Face hallucination techniques are proposed to enhance the face resolutions, which is supposed to be helpful to improve the recognition performance. Current face hallucination approaches can be divided into two types: holistic models and patch-based models. The advantages and disadvantages of both holistic and patch-based hallucinating models are analyzed respectively in this thesis. Holistic models preserve facial features while they introduce noises in the learning process. Patch-based models can generate smooth results. However, some facial features may be lost if low-resolution faces are enhanced to high-resolution ones through these models.

A new holistic face hallucination model is firstly proposed in this thesis. Face features in Eigen-subspace are adopted to enhance the facial resolutions. Recursive holistic residual compensation method is obtained to render the local features and to reduce the noises. A two-stage method is further proposed to render the residues. The proposed holistic hallucination model can increase the hallucination performance in terms of Peak Signal Noise Ration (PSNR) and Root Mean Square Error (RMSE). A patch-based model is further proposed in order to solve the noise problem. Training sample selection method is proposed based on Curvelet features. After patch-based face hallucination, a holistic based residue compensation method is proposed, which renders the lost global facial features in patch-based enhancement. The proposed approach generates smooth facial images and in the meanwhile, compensates global facial features. Extensive experiments show the improvement in terms of PSNR and RMSE when compared with other popular face hallucination algorithms.

Face recognition is one of the motivations behind face hallucination. As a result,

face recognition performance is proposed to be the evaluation method in this thesis. Recognizing faces in low resolutions and in hallucinated high resolutions are both studied. Experiments show that the recognition improvements of hallucinated faces are not so evident in many of the current face databases. We found that these databases are obtained from high definition cameras in controlled environments. The low-resolution face images are derived by down-sampling method. In this situation, facial images with the size of 32×32 can have higher recognition performance after hallucination. However, faces with the size of 8×8 can hardly be improved in the same way. Traditional evaluating methods PSNR and RMSE are also analyzed and compared with recognition performance. According to the experiments, traditional PSNR and RMSE measurements can not exactly represent the hallucination quality in terms of recognition performance.

A Practical face recognition scenario is proposed and analyzed, where the low-resolution face images are obtained from directly captured images in far distances instead of down-sampled images from high-resolution images. Three factors that influence face recognition performance are proposed and analyzed. And experiments demonstrate that resolutions play a key role in face recognition of surveillance systems. Thus for those low resolution images captured in far distance with surveillance cameras, face hallucination can be very useful. A new approach which combines the advantages of holistic hallucination model and patch-based face hallucination model is proposed. Experiments demonstrate that the proposed approach can improve the recognition performance in surveillance environment.

Acknowledgements

There are many people I would like to say thanks for their help and support during my four-year's PhD study.

Firstly, I would like to thank my supervisor, Associate Professor Wanquan Liu, for his tremendous help during my study. He is always happy to share ideas and gives me many valuable suggestions, which inspires me to discover the scientific laws in my research field. His professionalism in teaching and mentoring wins my respect. Also, his encouragement and patience support me to overcome a lot of difficulties during my research. I feel I am so lucky having the opportunities to carry out research under the supervision of Associate Professor Liu, which is definitely my precious experience during my lifetime.

I would also like to thank my co-supervisor, Professor Ling Li, who spent a lot of her valuable time on my research discussion. And I also appreciate her great efforts to help me improve my academic writing, from which I have benefitted a lot.

In addition, many staff in Curtin University are on my appreciation list. They have provided excellent service during my research study in the University. In particular, my special thanks goes to Mrs Mary Mulligan, who teaches me how to be a professional person in my future career.

Last but not least, I want to thank my family for their understanding and support throughout my research pursuit. Due to their firm support, I am able to devote all myself to my research dream in the past four years: my wife, Ms. Sha Liu, my parents, my parents-in-law and my uncle Mr. Han Qin.

Publications

This thesis is based upon several works that have been published (or submitted) over the course of the authors PhD, listed as follows in chronological order:

- Xiang Xu, Wanquan Liu, Svetha Venkatesh (2012). An Innovative Face Image Enhancement Based on Principal Component Analysis. *International Journal of Machine Learning and Cybernetics*, **3**(4), 259-267.

- Xiang Xu, Wanquan Liu, Ling Li (2013). Hallucinating Face In Curvelet. *IADIS International Conference Computer Graphics, Visualization, Computer Vision and Image Processing*, 35-43, Prague, 2013.

- Xiang Xu, Wanquan Liu, Ling Li (2013). Face Hallucination: How Much It Can Improve Face Recognition. *Australian Control Conference (AUCC)*, 93-98, Perth, 2013.

- Xiang Xu, Wanquan Liu, Ling Li (2014). Low Resolution Face Recognition in Surveillance Systems. *Journal of Computer and Communications*, **2**, 70, Shenzhen, 2014.

- Xiang Xu, Wanquan Liu, Ling Li (2014). The Impacts of Holistic And Patch Models in Face Hallucination. *Pattern Recognition Letters*. (Submitted)

- Xiang Xu, Wanquan Liu, Ling Li (2014). The Extensive Study of Face Hallucination and Face Recognition In Surveillance Environment. *Journal of the Computer Vision and Image Understanding*. (Preparing)

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

When we are using social media e.g. Facebook and some online forums, we often met the problem that we could not recognize people's profile images because of their small sizes. In many cases, the profile pictures are people's faces or include people's faces. It is really difficult to recognize who these people are due to the lower resolution of the photos. There are also other situations that we are bothered by small sized face images, for example, graduation photos where the pictures are taken from far distances and one specific face inside the group is very small. The potential application also includes face recognition in surveillance systems, e.g., CCTV. So how to recognize these small size faces is an open problem in computer community.

1.1 Face Recognition Overview

Face recognition has been a popular research topic in recent decades. Face recognition can be divided into human perception and machine recognition (Zhao *et al.*, 2003). From Zhao *et al.* (2003), we can see that face recognition in human perception is a psychological problem. Human generally recognize faces in holistic. If some of the facial features are obvious, local feature based recognition is also applied. In some situations, contextual knowledge is also used, which makes human perception recognition be a sophisticated system. In most cases human perception performs better than machine recognition. However, machine recognition has its ad-

vantages. For example, machine recognition can store and deal with a large amount of data. Learning from human recognition, machine based face recognition system consists of three steps (Zhao *et al.*, 2003): face detection, feature extraction and face recognition. Face detection techniques have been widely applied in our daily lives, such as the face detection function in digital cameras. In terms of feature extraction and face recognition, many approaches have been proposed such as Principal Component Analysis (PCA) (Turk and Pentland, 1991), Linear Discriminant Analysis (LDA) (Belhumeur *et al.*, 1997), Locality Preserving Projections (LPP) (He and Niyogi, 2004) and Face Recognition via Sparse Representation (SRC) (Wright *et al.*, 2009). Face recognition algorithms are also widely implemented nowadays, for example, the face recognition applications in *facebook* and *iphoto*.

In the situation when face images are very small, low-resolution face recognition approaches are proposed. There are several possibilities that face images are very small. The first possibility is this small face image are resized from a high-resolution face image. Another reason is because that the low-resolution face image was captured from a far distance. And it was also probably captured by CCD(Charge-Coupled Device)/CMOS(Complementary Metal Oxide Semiconductor) image sensors with small sizes. As a special issue in face recognition field, recognizing small size faces is difficult through the existing face recognition algorithms. Because there are very few facial features in these small face images which can not provide enough information for recognition. In general there are two directions to recognize small size face images. One is recognizing these captured faces directly in small sizes. As faces in gallery set are often in high resolutions. The usual way is to down-sample these high-resolution gallery faces to low-resolution face images which have the same resolution as the captured low-resolution face images. Then face recognition algorithms can perform directly on these low-resolution testing faces and gallery faces. However, the other popular direction for low-resolution face recognition is to enhance the image resolutions. Captured low-resolution face images are firstly enhanced to high

resolutions which are the same as gallery faces. And then they are recognized by machine perceptions algorithms.

1.2 Enhancing Face Resolutions

In most cases, the main reason that faces can not be recognized is their low resolutions, especially when facial images are captured in far distances. As a result, enhancing low-resolution face images can be re-directed to enhancing their resolutions. As a special case of image, facial image resolution enhancement can be traced back to image super-resolution.

1.2.1 Generic Image Super-Resolution Overview

The earliest generic image super-resolution was proposed by Tsai and Huang (1984) toward multi-frame image super-resolution. They built a linear relationship between a set of shifted low-resolution images and the high-resolution image in frequency domain. This relationship is based on the shifting properties of Discrete Fourier Transformation and Continuous Fourier Transformation. Based on their work, many frequency based approaches were proposed (Kim *et al.*, 1990; Tom and Katsaggelos, 1995; Bose *et al.*, 1993). However, this frequency based transformation is difficult to deal with noise problem. Since it is not able to add prior information. Such techniques can only deal with multi-frame image super-resolution problem. There are also other multi-frame based super-resolution techniques, e.g., Project Onto Convex Sets (POCS) (Patti and Altunbasak, 2001), Maximum a Posteriori (Chantas *et al.*, 2008, 2007).

For single image super-resolution, it can be simply separated into two categories: direct interpolation and learning based super-resolution. Similarly, single image interpolation is hard to add prior information. Interpolation technique increases the image resolution directly from the low-resolution input images (Wolberg, 1990; Chen and Defigueiredo, 1985; Karayiannis and Venetsanopoulos, 1991; Xue *et al.*, 1992; Schultz and Stevenson, 1994), but could not achieve smooth accurate results since it utilizes information only from low resolution images. Learning-based super-resolution is popular in the past decade. Many of the approaches first divide low-resolution image into patches, and then learn the patch relationship between testing and training examples through machine learning algorithms. The high-resolution image is achieved through reconstructed high-resolution patches. This framework includes two data sets: the low resolution and corresponding high resolution training samples. Since the purpose of image enhancement is to obtain a high resolution image for a corresponding low resolution image, learning algorithms aim to explore the relationship between high resolution and low solution images. Freeman *et al.* (2000) developed a Markov Network to learn the relationship. Hertzmann *et al.* (2001) applied the 'Image Analogies' method to obtain high resolution images using local feature transforms.

1.2.2 Face Hallucination Overview

As a special type of digital images, facial images have the common properties of generic images. Thus their resolutions can be enhanced through image super-resolution techniques. However, facial images also have their special properties. Approaches specifically for enhancing facial images' resolutions were firstly proposed in 2000 by Baker and Kanade (2000). The super-resolution on facial images is then named as 'Face Hallucination'. After that, Wang and Tang (2005) introduced the Eigen-transformation algorithm at a relatively small computational cost.

They project the low resolution face image into the eigen-faces of the low-resolution training set and obtain the coefficients which can be used to construct a high resolution image. This method is computationally efficient for face image resolution enhancement. However only the global face is derived and thus some local features are not well characterized. For example, hair and glasses in some face images are not well represented in the reconstructed high resolution images. Liu *et al.* (2007, 2001) proposed a two-step face hallucination algorithm, who introduced the residue concept for face hallucination. In the first step, a global face is generated. After that, local features are derived by minimizing the energy of a Markov Network. In order to reduce the computational cost, Zhuang *et al.* (2007) used locality preserving projection (LPP) and radial basis function (RBF) to generate the global faces and rendered the residue part by applying the Nearest Neighbor algorithm. Based on Zhuang *et al.* (2007), Huang *et al.* (2010a) assumed the global face hallucinating process to be a black box. Then they estimated the global face with a linear transformation on the basis of PCA. Recently, Yang *et al.* (2008a) proposed a method by combining the Non-negative Matrix Factorization with sparse representation algorithms. Jia and Gong (2008) used trained tensor to construct a high-resolution face and compensate the residue by a nonparametric patch learn process. Huang *et al.* (2010b) used canonical correlation analysis in both global face and residue generation. Liang *et al.* (2010) proposed the use of morphological component analysis (MCA) in global hallucination faces and neighbor reconstruction for local features. Zhang and Cham (2011) proposed a face hallucinating algorithm in frequency domain instead of the conventional spatial domain. They transferred the low-resolution face images into Discrete Cosine Transformation coefficients and inferred the high-resolution coefficients through utilizing Markov Random Field (MRF). Then, the expected high-resolution face images can be acquired by adopting the inverse Discrete Cosine Transformation.

Regardless of various fields where face hallucination approaches come from, those

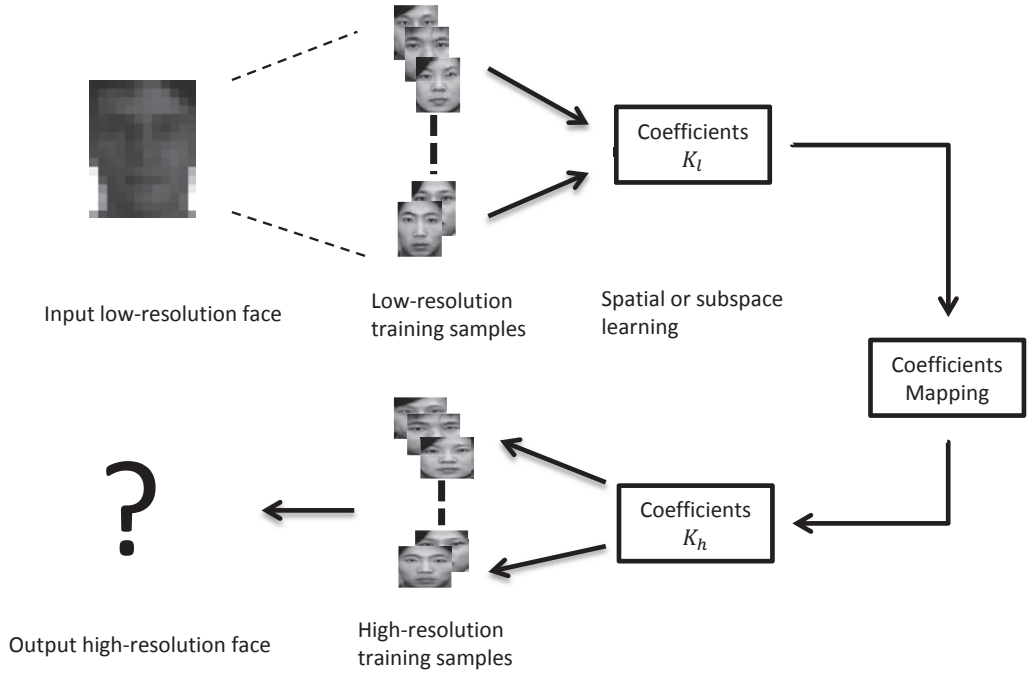


Figure 1.1: Face Hallucination Diagram.

proposed methods can be divided into two classes according to how to deal with the testing faces. One kind of approaches treat facial images as a vector, and make use of face properties in subspaces to enhance the resolutions. The other method is to divided facial images into overlapped patches, and enhance each patch separately. Then by combining overlapped patches, one can construct the higher resolution images.

1.2.2.1 Hallucinating Faces in Holistic Model

The obvious benefit of hallucinating faces in holistic images is that it makes use of facial properties. As seen from Figure 1.1, low-resolution faces are trained in full images, which keeps the face structure and the facial properties. During the learning process, these structures and properties are kept and passed to hallucinated high-resolution face images. In order to strengthen these facial properties, Chapter 2

proposed a method which purely treat both the testing and training faces as holistic images. One advantage of hallucinating faces in holistic model is its global features are well kept. However, many of the local features may be lost and noises are generated during the hallucination.

1.2.2.2 Hallucinating Faces in Patch Model

Most of patch based hallucinating approaches come from generic image super-resolution approaches. Through dividing facial image into small and overlapped patches, each patch is considered as a generic image. Each of them is enhanced separately to high resolutions. After combing all the patches back, a high-resolution facial image is generated. Patch based face hallucination can generate very smooth high-resolution facial images and achieve good performance in terms of Peak Signal Noise Ratio and Root Mean Square Error. However, in some cases the hallucinated faces are too smooth to keep the global facial features.

1.3 Challenges in Face Enhancement

1.3.1 Holistic vs. Patches

The visual results of hallucinated faces can be divided into two types according to two kinds of hallucinating methods. The holistic approaches treat facial images as a whole data, thus face structures and properties are well kept and transferred as many as possible in the learning and reconstruction steps. However, errors are introduced in learning and mapping procedures. The reconstructed facial images have a good representation of global facial features while the local features may be

lost. The vision of hallucinated faces in holistic methods is noisy and not smooth. On the contrary, patch based hallucinating methods divide face images into small overlapped patches. These patches are treated as generic images and one learns the relationship between patch dictionaries. In the reconstruction step, the learned high-resolution patches are overlapped with each other. This step makes the hallucinated faces very smooth. Especially when patch size is small and overlapped size is large, the hallucinated facial images are too smooth to represent human facial structures and properties. In other words, the local features are kept while the global features of faces may be lost in compromise. Figure 1.2 shows the comparison between hallucinating faces in holistic model and in patch based model. The testing and training data are exactly the same. The method of the holistic model is by Liu *et al.* (2007) and the method of the patch-based model is by Wright *et al.* (2009); Yang *et al.* (2010). As we can see from Figure 1.2, face hallucinated by holistic model (Figure 1.2 (b)) represents global face features well. However, the local features are not well constructed. Noises appear around mouth, chin and neck area. Face hallucinated by patch-based model has a smooth appearance. But some of the global face features are lost. Especially in the area around the eyes and eyebrows, hallucinated face can not be as clear as the face hallucinated by holistic model.

1.3.2 Face Hallucination Evaluations

In previously proposed approaches, most of them adopt Root Mean Square Error (RMSE) and Peak Signal Noise Ratio (PSNR) as evaluation metrics to evaluate hallucinating results. These evaluation methods are applicable for grey scale images. For color face images, they will firstly be transferred into grey scale images. For example, Yang *et al.* (2010) first transfer RGB images to YCbCr color space, where Y is the luma component and CB and CR are the blue-difference and red-difference chroma components. Y component is adopted as the grey scale image for RMSE

(a) (b) (c) (d)

Figure 1.2: Examples Display of Comparison between Holistic Model and Patch-based Model. (a) Low-resolution Face Image. (b) Hallucinated High-resolution Face by Holistic Model (Liu *et al.*, 2007). (C) Hallucinated High-resolution Face by Patch-based Model (Yang *et al.*, 2010). (d) Original High-resolution Face Image

and PSNR evaluatoin. Both calculation equations are listed as below:

$$RMSE = \sqrt{\frac{\sum_{1}^{m}\sum_{1}^{n}(I - \hat{I})^2}{m \times n}} \tag{1.1}$$

$$PSNR = 20 \cdot \log_{10} \frac{255}{RMSE} \tag{1.2}$$

where m and n are the numbers of rows and columns of the high-resolution images. I and $\hat{I}$ represent the original high-resolution testing images and hallucinated high-resolution images respectively.

However, RMSE and PSNR can only evaluate the average differences for the whole image. RMSE calculates the average square differences between hallucinated faces and original high-resolution face images. Similarly, PSNR calculates the average signal noise ratio of the whole image. However, as we know that some important facial features are actually the key points for face recognition instead of the whole face images. For example, eyes, eyebrows, mouth, nose and etc. RMSE and PSNR

can not specifically evaluate these features of hallucinated face images. In practice, smoothly enhanced face images often have high RMSE and PSNR values, which are actually too smooth to recognize. For example, face images enhanced through cubic interpolation (Hou and Andrews, 1978) which is a super-resolution method for generic images can achieve good RMSE and PSNR performance. But faces hallucinated by interpolation are usually blurred and difficult to be recognized (Hou and Andrews, 1978; Baker and Kanade, 2000, 2002). Therefore, how to exactly measure those face hallucination approaches is in fact a challenging task, especially for color or 3D images.

1.3.3 Down-sampled Low Resolution Faces vs. Directly Captured Low Resolution Faces

Since hallucinating faces is to enhance low-resolution facial images to high-resolution images, the objects of hallucinating (low-resolution facial images) are important. In many previous approaches, these low-resolution faces are derived from down-sampling method. The down-sampling equation is given below:

$$x_l(m,n) = \frac{1}{k^2} \sum_{p=0}^{k-1} \sum_{q=0}^{k-1} x_h(m*k-p, n*k-q) \tag{1.3}$$

where x_l and x_h is low and high resolution face pair, m, n are image pixel position and k is down-sampling rate.

Down-sampling is a good way to achieve low-resolution images with good quality. However, in many circumstances, low-resolution resolution facial images are directly captured from cameras. The difference between them are quite obvious. Figure 1.3 lists both down-sampled low-resolution faces and directly captured low-resolution faces. How much the hallucination methods can improve the directly captured low-resolution faces and how much these hallucinated faces can improve face recognition

(a) (b) (c)

Figure 1.3: Examples Display of Faces Captured through Different Environments. All the Faces are in Similar Resolutions. (a) Down-sampled Faces Captured by HD Camera in controlled Environment. (b) Directly Captured Faces by HD Camera in Far Distance. (c) Directly Captured Faces by Surveillance Camera.

performance are open issues and we will investigate these problems in this thesis.

1.4 Contributions of this Thesis

Considering the above challenges in face hallucination field, this thesis gradually explores these issues in separate chapters. Hallucinating faces in holistic model and patch based model are explored individually in Chapter 2 and 3. The conditions and limitation of face hallucinating evaluation are discussed in Chapter 4. Down-sampled faces and directly captured faces in real world are compared and discussed in Chapter 5.

1.4.1 Hallucinating Faces in Holistic Images

The advantage of holistic model is keeping the global features of facial images which makes hallucinated faces to be 'closer' to human faces instead of generic images. Chapter 2 proposed an approach which purely hallucinates facial images in holistic model. In order to remove the mapping errors which are the disadvantages taken by holistic model, two methods are proposed. The first is a holistic compensation step. After mapping and reconstructing a high-resolution facial image, this hallucinated image is then down-sampled to low-resolution. There is a residual difference between low-resolution testing and this down-sampled hallucinated image. This residual image is then mapped to a high-resolution residual image. This step adopts holistic models to render a residue image. It can be performed iteratively, thus reduce errors. Another method named 'two-stage' is also proposed which hallucinating low-resolution faces in separate resolution stages. The proposed approach purely utilizes holistic model which keeps the global features. The holistic and iterative residual method and two-stage method can help to compensate local features in holistic way which makes the hallucinated faces more smooth.

1.4.2 Hallucinating Faces in Patches

Although the proposed holistic approach mentioned in previous section can both keep global features and compensate local features, the hallucinated facial image are still not smooth enough for vision perception. Chapter 3 proposes a method through combing holistic model and patch model. This approach has three components: training sample selection, patch based hallucination and holistic residual compensation. The main hallucinated face is generated by patch based model, which makes the hallucinated faces very smooth. However, in order to compensate global face features, two more rendering parts are added before and after the hallucination.

Before hallucination, a training sampling pre-selection part is performed. Based on Curvelet features of facial images, training samples of hallucinating faces are first selected and this forces testing faces can only learn the relationship from those selected training samples with similar global features as the testing face. After hallucination part, a residual compensation part that renders residues in Curvelet frequency domain is adopted. Both the above methods can improve the global face features in the hallucinated faces.

1.4.3 Face Hallucination for Recognition Performance Improvement

Chapter 4 analyzes the current popular face hallucination evaluation methods RMSE and PSNR and proposes to use face recognition performance to replace RMSE and PSNR as an evaluation method. Further more, Chapter 4 indicates that hallucinated high-resolution facial images do not guarantee high face recognition performances. If the low-resolution facial images are derived from down-sampling method, the hallucinated facial images can perform even worse in recognition test when the low-resolution face images have the size of 8×8 before hallucination. On the contrary, if the low-resolution testing faces have the resolution of 32×32, the recognition performance of hallucinated faces can be improved. In this situation, only hallucinating faces from 'higher' low resolutions (32×32) can improve the face recognition performance. When the testing facial images are in a very small resolution, for example 8×8, the recognition performances of hallucinated faces would decrease. Besides, we also find that recognition performances on down-sampled low-resolution faces in many face databases do not decrease magnificently along with the decrease of resolutions.

1.4.4 Face Recognition in Surveillance Scenarios

Chapter 5 explores face recognition performances with real life cameras. It reveals that directly captured facial images in surveillance environment are different from those face databases taken in laboratories. When the image resolutions of captured faces drop, the face recognition performances would drop dramatically. Three factors which influence recognition performance in surveillance systems are analyzed: Capture distance, types of cameras and face resolutions. Through extensive experimental analysis, resolution is regarded as the most important factor for face recognition in surveillance environments. Regardless of cameras and capturing distances in indoor scenarios, the higher captured resolutions prove to have higher recognition performances. Thus, face hallucination techniques are more useful in this situation. A new hallucinating scheme is also proposed for face recognition in this case. Low-resolution facial images are hallucinated separately both in holistic model and patch based model. A decision maker process is proposed to determine each pixel value of the output faces, which can improve the recognition performance as validated in experiments.

1.5 Face Databases used in this Thesis

In this thesis several face databases are used for different purposes. As for face hallucination restrictions, hallucinating faces with large poses is too difficult to carry out. All the testing faces only have a small pose variation with the angles which are less than thirty degrees in this thesis. However, training face selection does not have this restriction in order to derive more facial features in the learning process.

All the experiments follow leave-one-out principle. In face databases, each subject

means one person's ID. All the face images of this person inside a face database belong to this subject. This means when one human subject is selected as testing face, all facial images for this subject will be excluded in the training data. Furthermore, in Chapter 4 and Chapter 5 we extend this leave-one-out principle to database level. We did not hallucinate faces inside the same face database. We adopted an independent database Face Recognition Grand Challenge (FRGC) Phillips *et al.* (2006) as the training data to hallucinate faces. This principle makes our experiments closer to real applications for face recognition scenarios.

1.5.1 The Facial Recognition Technology (FERET) Database

The Facial Recognition Technology (FERET) Database (Phillips *et al.*, 2000) is sponsored by the Department of Defense's Counterdrug Technology Development Program, which aimed to develop automatic face recognition capabilities and was supposed to assist security, intelligence and law enforcement. Totally 14051 images were collected through a 35 mm camera and converted to eight-bit gray scale images of human heads. The original image size is 256×384. The FERET database consists of 24 sets for each person with various situations. For example, Fa includes regular face expressions while Fb consists of alternative frontal faces with facial expressions. Other sets also include illuminations and poses.

In this thesis, The Facial Recognition Technology (FERET) Database is utilized in Chapter 2. As the most frequently used face database in face hallucination, only frontal faces are collected. In this thesis, totally 839 subjects are collected. Each subject includes 2 to 10 frontal face images. Leave-one-out algorithm is adopted in all the hallucinating face experiments in this thesis. This means one subject is selected as testing samples and the left are used as training samples. The faces samples are shown in Figure 1.4.

(a) (b) (c) (d)

Figure 1.4: Examples Display of FERET Database.

Figure 1.5: Examples Display of YaleB Database.

1.5.2 The Extended Yale Face Database B (YaleB)

The Extended Yale Face Database B (Georghiades *et al.*, 2001; Lee *et al.*, 2005) contains 28 human subjects. Each subject includes 9 poses sessions and each pose session has 64 images with different illumination conditions. It means each human subject has 576 images.

The Extended Yale Face Database B is used in Chapter 2, 3, 4 and 5. Only frontal faces are collected for both hallucination and recognition. Thus in this thesis 28 subjects are used and each subject includes 64 frontal faces in various lighting conditions. The examples faces are illustrated in Figure 1.5.

(a) (b) (c) (d)

Figure 1.6: Examples Display of CAS-PEAL Database.

1.5.3 CAS-PEAL

CAS-PEAL face database (Gao *et al.*, 2008) is an newly published Chinese face database for face recognition. CAS-PEAL database is sponsored by National Hi-Tech Program and ISVISION by the Face Recognition Group of Joint Research & Development Laboratory for Advanced Computer and Communication Technologies (JDL), Institute of Computing Technology Chinese Academy of Sciences (ICT), Chinese Academy of Sciences (CAS). This database consists of 1040 human subjects including 595 males and 445 females. As a large scale human face database, CAS-PEAL provides 9 poses, 5 expressions, 6 accessaries (3 sunglasses and 3 caps) and 15 lighting conditions for each subject.

This database is used in Chapter 3 in order to verify the proposed hallucinating approaches are applicable in a wide range. The examples of CAS-PEAL are shown in Figure 1.6.

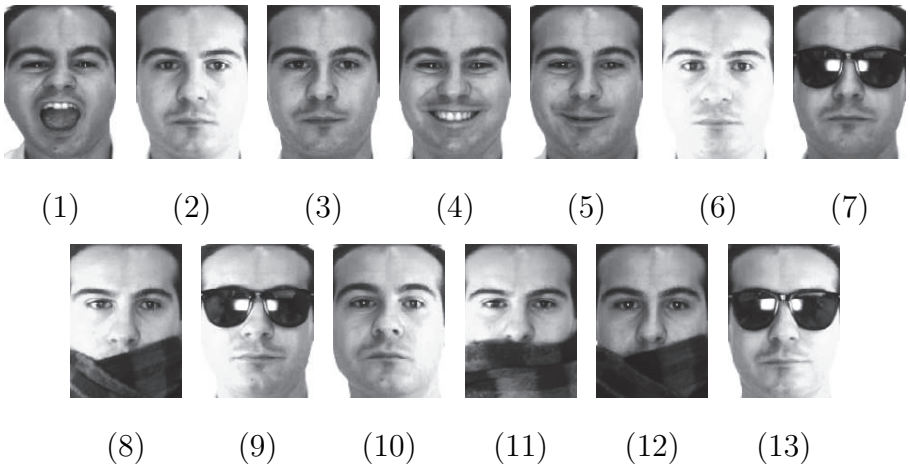


Figure 1.7: Examples Display of AR Database.

1.5.4 AR Face Database

AR face database (Martinez and Benavente, 1998) was created in the Computer Vision Center (CVC) in 1998. This database includes 126 human subjects with 70 males and 56 females. Each subject has two sessions which were collected at difference time (14 day away). Each session contains 13 images in various conditions including expressions, illuminations and occlusions. In this thesis AR database is performed in Chapter 4 and 5. The 13 examples of AR database are listed in Figure 1.7.

1.5.5 Face Recognition Grand Challenge (FRGC) Face Database

Face Recognition Grand Challenge face database (Phillips *et al.*, 2006)was collected from 2004 to 2006. This database contains a huge number of face data both in 2D and 3D. The primary goal of the FRGC was to promote and advance face recognition technology designed to support existing face recognition efforts in the U.S. Government. There are six experiments inside FRGC database. Experiment 1

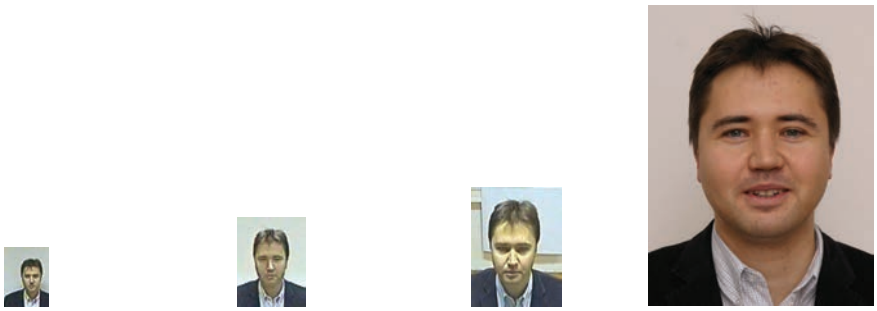
Figure 1.8: Examples Display of FRGC Database.

and 2 contain single and multi controlled face images. Experiment 3, 5 and 6 consider 3D face recognition. Experiment 4 consists of both controlled and uncontrolled 2D face images. The controlled facial images in FRGC database is defined as images taken in a studio setting and facial images taken under two illumination conditions and with two facial expressions (smiling and neutral). The uncontrolled images are defined as images which were taken in varying circumstances, e.g., hallways, atriums, or outside. Similarly, the uncontrolled facial images also include two expressions with smiling and neutral.

In this thesis, FRGC face database is used in Chapter 4 and 5. Most of the previous face recognition databases only contain still facial images under controlled environments. However, this thesis deals with far face enhancement which apparently can not be performed purely under controlled circumstances. Therefore, FRGC database is a good data source as face image training data since it contains both controlled and uncontrolled facial images. The facial data contained in Experiment 4 of FRGC is used in this thesis as facial training data. The example faces in FRGC are listed in Figure 1.8.

1.5.6 Surveillance Cameras Face Database (SCface)

Surveillance Cameras Face Database Grgic *et al.* (2011) is a newlypublished face database dealing with surveillance cameras. SCface database is collected through six cameras including five different surveillance cameras and one high definition digital camera. It contains 4160 static images in 130 subjects with 115 males and 15



(a) (b) (c) (d)

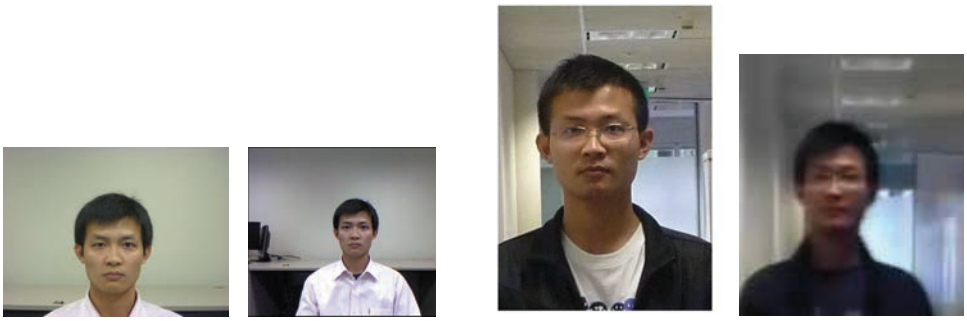
Figure 1.9: Examples Display of SCface Database.

females aging from 20 to 75. This database were captured in Croatia and all the 150 participants were Caucasians. The high definition camera captured human faces in an controlled environment. The other five surveillance cameras acquired face data in the same time in a room with natural lighting condition. Three distances were adopted for face capturing. As a result, three images were taken for each person in one camera in terms of three capturing distances.

As both down-sampled low-resolution facial images and directly captured low-resolution facial images are considered in this thesis, SCface face database meets the experimental demanding. This database is used in Chapter 5, where performances between directly captured faces and down-sampled faces are compared. The example faces in SCface database is shown in Figure 1.9.

1.5.7 CurtinFace Database

CurtinFace Database was collected in 2011 in Curtin University Australia. This this face database include 52 human subjects both males and females. A various conditions including lighting, poses and glasses are included. This database contains three cameras and four sections. Section one consists of facial images captured by



(a) (b) (c) (d)

Figure 1.10: Examples Display of CurtinFace Database.

a Microsoft Kinect camera with 3D information and section two contains facial images captured by a high definition camera. Both these two sections are taken under controlled environment. Section three and section four are produced under uncontrolled surveillance conditions in a corridor. Face images in section three are collected by the same high definition camera as section two, but in different circumstances and far distances. Section four contains facial images captured by a commercial surveillance camera in far distances.

This face database is adopted in Chapter 5, where surveillance scenarios are analyzed and compared with controlled face capturing environment. The examples of CurtinFace database is shown in Figure 1.10.

Chapter 2

Hallucinating Faces in Holistic

2.1 Introduction

The research in learning based face hallucination can be divided into two categories. One is to treat face images as generic images and utilize image super-resolution methods to enhance the resolutions. The other way is to utilize the special property of face images, e.g., eigen-subspace and to learn the relationship between testing face and training faces. After mapping this relationship from low-resolution field to high-resolution field, a high-resolution result can be derived. This holistic method treats each facial image in one entity and learns the relationship in a holistic model.

In order to keep face features when mapping from low dimension to high dimension, subspace methods are frequently used in face hallucination. In this chapter a new face image enhancement method is proposed based on holistic model. First, the relationship of projection coefficients between high-resolution and low resolution images on the respective eigen-faces is investigated. Based on this investigation, a high-resolution global face is constructed. Then we propose a residue technique to render the global face for detailed parts. As such, residue image is constructed by eigen-subspace projections derived from high and low resolution residue training samples. The reconstructed face image is the combination of the global face and the residue image. Further we propose a recursive residue rendering method with an aim to compensate more details of local features. Finally a two-stage learning

framework is proposed by adding a middle-stage resolution learning set and enhancing resolution in two stages. This two-stage method has proven to be able to obtain more information from the middle-stage learning set, and thus can improve the face hallucination results. The proposed method can improve the approach proposed by Wang and Tang (2005) in terms of Peak Signal Noise Ratios. Meanwhile it has lower computational cost compared with methods by Liu *et al.* (2007) and Yang *et al.* (2008a).

The rest of this chapter is organized as follows. Background and related algorithms are introduced in Section 2.2, followed by details of the proposed algorithms in Section 2.3. Experiments are conducted in Section 2.4. In Section 2.5 summary of this chapter is represented.

2.2 Background and Related Works

2.2.1 Down-sampling from High-resolution to Low-resolution

2.2.1.1 Down-sampling Function

Since most of previous research on face hallucination utilizes face recognition databases to perform experiments, in this chapter we also use general face databases to hallucinate faces. The provided faces in those databases are usually in high resolutions. Just like previous work, we adopt down-sampling method to produce low-resolution face images in this chapter.

Let x_l denote a low-resolution face image, and x_h denote the corresponding high-

resolution version. x_l can be derived by the down-sampling method from x_h as follows:

$$x_l(m,n) = \frac{1}{k^2} \sum_{p=0}^{k-1} \sum_{q=0}^{k-1} x_h(m*k-p, n*k-q) \tag{2.1}$$

where k is an integer and represents the down-sample rate, m and n are the image pixel position.

Down-sampling method is adopted for two kinds of data. The first is the generation of low-resolution testing data. As we know, in current face databases, facial images are generally in high resolution. In order to generate a proper low resolution testing data which matches hallucinating experiments, these picked testing images need to be down-sampled. The other data is low-resolution training data. In the learning process, low-resolution testing image is used to learn the relationship between itself and training data. As the same reason for the low-resolution testing data's generation, selected training data is usually in high resolution. These high-resolution training data need to be down-sampled to the same resolution as testing data.

In fact, this down-sampling method calculates the average value of each non-overlapped block with the size of $k \times k$. For example, let $k = 4$. The high-resolution image is then separated into a set of 4×4 blocks. If the high-resolution image is a 128×128 facial image, there will be 1024 blocks in this image. After down-sampling, these 1024 blocks will be 1024 pixel as the low-resolution image. The value of each pixel is the average value of each block.

There are also other down-sampling methods to down-sample a high-resolution image to a low-resolution one. From the Nyquist-Shannon sampling theorem (Shannon, 1949) of signal processing, a high-resolution image can be down-sampled through sampling its row and column. For example, for a facial image with size of 128×128, the first column of every four columns is sampled and kept as the columns of the low-resolution image. After that the first row of every four rows is sampled and

kept as the rows of the low-resolution image. Then the 32×32 low-resolution image is derived. The low-resolution image can also be derived by selecting the second, the third or the fourth column and row of high-resolution image. Nyquist-Shannon sampling theorem (Shannon, 1949) is generally used to transfer a analog signal to discrete one, which indicates the sampling rate should be twice equal or bigger than the bandwidth of signal in order to perfectly reconstruct the signal. However, when the sampling rate is smaller than than twice of the signal bandwidth, some high frequency parts of the signal will be lost and the signal can not be reconstructed completely. The generation of low-resolution face images has the same problem. As a kind of digital signal, the recording process of digital images contains the converting from analog signal to digital signal. The deriving of low-resolution face images is due to the limited size of camera sensors. For instance, if the captured faces are far away from the camera, the captured faces can only be represented by a few pixels. When converted from analog to digital, the sampling rate is thus limited compared with bandwidth of face images. As there is not any analog face signals in the face databases, we can only further sample the digital images to low-resolutions, which have being sampled when recorded in cameras, to simulate the naturally capture process of low-resolution face images. The high-resolution face images are assumed as being sampled under the Nyquist-Shannon sampling theorem (Shannon, 1949), which can be reconstructed. The low-resolution face image are assumed as being sampled out of Nyquist-Shannon sampling theorem and high frequency components are lost in the low-resolution face images. In this chapter, the down-sampling method based on sampling theory follows this method (Shannon, 1949).

In frequency domain, the low-resolution image can also be derived by filtering the high-frequency components of high-resolution image. High-resolution image is first transferred to frequency domain, the high frequency components are then abandoned and low frequency components are kept. When these low frequency components are transferred back to spatial domain, the low-resolution image can be derived. Figure

(a) (b) (c) (d)

Figure 2.1: Examples Display of Faces through Different Down-sampling Methods. (a) Original High-resolution Facial Image (with the resolution of 192×128). (b) Low-resolution Facial Image Down-sampled by Equation 2.1 (with the resolution of 48×32). (c) Low-resolution Facial Image Down-sampled by Sampling Theory (Shannon, 1949) (with the resolution of 48×32). (d) Low-resolution Facial Image Down-sampled in Frequency Domain (with the resolution of 48×32).

2.1 shows the down-sampling results through different methods. From the Figure 2.1 it can be concluded that low-resolution faces down-sampled by Equation 2.1 (Figure 2.1 (b)) and by frequency domain (Figure 2.1 (d)) have similar quality. However low-resolution face down-sampled by sampling method (Figure 2.1 (c)) has a poor quality. The influences brought by different down-sampling methods in hallucinating faces will be investigated in our experiments in Section 2.4.

2.2.1.2 Zero-mean Face Matrix

Let $F^h = [F_1^h, F_2^h, \ldots, F_N^h]$ denote the high-resolution training face images set, each F_i^h $(i = 1, \ldots, N)$ represents one face in the database. F_i^h is a one-column matrix by reshaping the face matrix. Similarly, $F^l = [F_1^l, F_2^l, \ldots, F_N^l]$ denotes the low-resolution training face images set, m^h and m^l represent the mean face images respectively. m^h and m^l are also reshaped as one-column matrix. In order to reduce

calculation errors, we first minus the mean face from F^h and F^l. The mean face m^h and m^l are the average values at each point of the face matrix. The zero-mean training face images can be obtained as:

$$A^h = [F_1^h - m^h, F_2^h - m^h, \ldots, F_N^h - m^h] = [A_1^h, A_2^h, \ldots, A_N^h],$$

$$A^l = [F_1^l - m^l, F_2^l - m^l, \ldots, F_N^l - m^l] = [A_1^l, A_2^l, \ldots, A_N^l],$$

2.2.1.3 Up-sampling Faces

In terms of different down-sampling methods, there are different equations for down-sampling methods. However, all the image down-sampling equations including E-quation2.1 can be denoted in a symbolic form as:

$$A^l = K \downarrow \times A^h \tag{2.2}$$

where $K \downarrow$ is the down-sample operator. A^l and A^h represent low-resolution and high-resolution samples respectively.

For image super-resolution or face hallucination fields, we are interested in investigating the $K \uparrow$, where $K \uparrow$ can satisfy the following equation:

$$A^h = K \uparrow \times A^l \tag{2.3}$$

In the proposed method, $K \uparrow$ is derived from learning the relationship between low-resolution training set A^l and high-resolution training set A^h through Principal Component Analysis.

2.2.2 Eigen-transformation

Eigen-transformation approach by Wang and Tang (2005) is based on Principal Component Analysis. And it can be outlined as follows:

First, the eigen-subspace for high-resolution training images is constructed. Let $E^h = [E_1^h, E_2^h, \ldots, E_N^h]$ and Λ^h represent eigen-vector and eigen-value matrices respectively, which are obtained from a covariance matrix C:

$$C = \sum_{i=1}^{N} (F_i^h - m^h)(F_i^h - m^h)^T = A^h (A^h)^T \tag{2.4}$$

The weight vector w^h for an input high-resolution image x^h can be computed as follows.

$$w^h = (E^h)^T (x^h - m^h) \tag{2.5}$$

In Principal Component Analysis based face recognition (Turk and Pentland, 1991), the weight vector w^h is used as face features. The nearest neighbor of w^h in the weight vectors of training faces is the recognized face.

With above defined w^h, a high resolution image y^h can be reconstructed based on the high resolution eigen-faces as:

$$y^h = E^h w^h + m^h \tag{2.6}$$

However, the dimension of the covariance matrix C is high. In order to reduce computation, eigen-vectors $V^h = [V_1^h, V_2^h, \ldots, V_N^h]$ of covariance matrix $\hat{C} = {A^h}^T A^h$ are adopted (Turk and Pentland, 1991). Then E^h can be computed from V^h as:

$$E^h = A^h V^h \frac{1}{\sqrt{\Lambda^h}} \tag{2.7}$$

Therefore, Equation 2.6 can be rewritten as:

$$y^h = (A^h V^h \frac{1}{\sqrt{\Lambda^h}})w^h + m^h = A^h K^h + m^h \tag{2.8}$$

where $K^h = [K_1^h, K_2^h, \ldots, K_N^h] = V^h \frac{1}{\sqrt{\Lambda^h}} w^h$ are coefficients when input image x^h is projected to the training data basis.

Similarly, in the low-resolution version face database, a low resolution face image y^l can be reconstructed as:

$$y^l = A^l K^l + m^l \tag{2.9}$$

In Wang and Tang (2005), the high-resolution reconstructed face image can be calculated by the following equation:

$$y^w = A^h K^l + m^h \tag{2.10}$$

In fact Wang and Tang (2005) replaced K^h in Equation 2.8 by K^l, which indicates that K^h and K^l are approximately the same. This will be explained in the next section.

2.3 Proposed Approach

2.3.1 Global Face Hallucination

In order to establish the relationship between the low-resolution domain and high-resolution face images, we project a low-resolution image into the low-resolution eigen-subspace using the low-resolution training set, and its corresponding high-resolution image into the high-resolution eigen-subspace using the high-resolution

training set separately. Then the coefficients (K^l and K^h) are calculated respectively before comparison.

An experiment is carried out in order to test whether K^l equals to K^h. We design the experiment as follows. We randomly select one facial image from FERET database (Phillips *et al.*, 2000), and then randomly select other 200 facial images from 200 human subjects. Each person provides one image. The first selected one person x^h is not included in the 200 images, neither in the 200 human subjects. These selected 200 face images are set as training data A^h. The eigen-subspace is then calculated as Equation 2.7. x^h is then projected into the eigen-subspace and K^h is derived through the following equation:

$$K^h = [K_1^h, K_2^h, \ldots, K_N^h] = V^h \frac{1}{\sqrt{\Lambda^h}} w^h \tag{2.11}$$

Also, K^l can be derived exactly the same as K^h.

Figure 2.2 shows the value of coefficients for a pair of test images. The black line with stars represents the values of K^h and the green one with circles denotes K^l.

It can be seen from the figure that the values of coefficients (K_i^l and K_i^h) are very similar, which means if we replace K_i^h with K_i^l in Equation 2.8, we can obtain a high-resolution output as follows:

$$y^h = A^h K^h + m^h \approx A^h K^l + m^h \tag{2.12}$$

The coefficients K^l in equation Equation 2.12 can be derived when a test image x^l is provided. We can then construct a high-resolution global image y^g according to Equation 2.12 as follows:

$$y^g = A^h K^l + m^h \tag{2.13}$$

Though the global face can be constructed based on K^l in Equation 2.13, the global

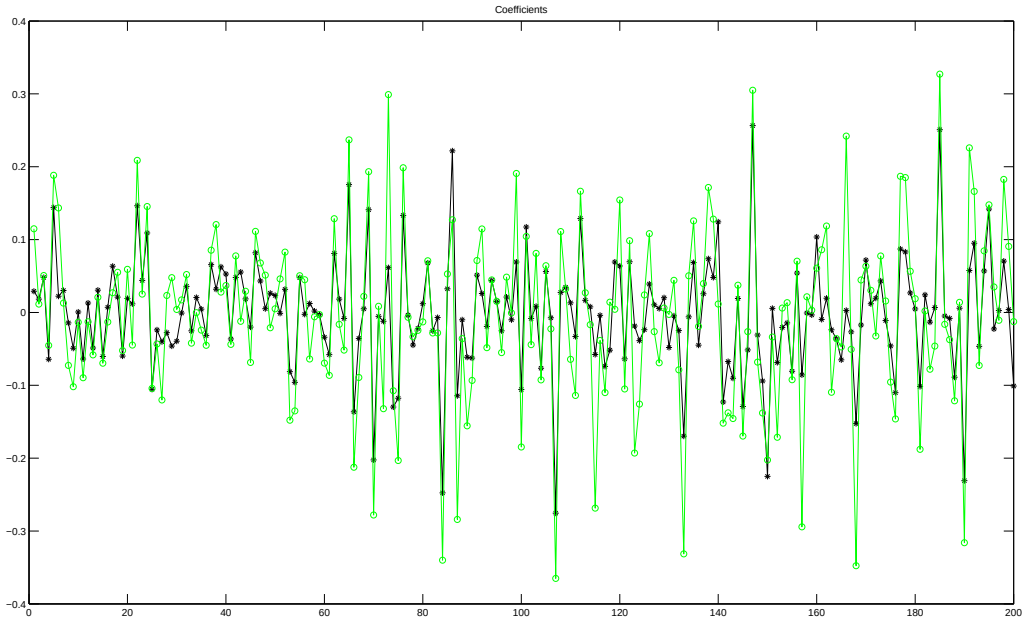


Figure 2.2: Comparison of projection coefficients K^h and K^l

face can not represent the details quite well partially because K^l is only an approximation of K^h. As can be noticed these two parameters are not exactly the same as observed in Figure 2.2. The difference between K^l and K^h will lead reconstruction errors. Thus residue technique is adopted to adjust the hallucinated faces as introduced in next subsection.

2.3.2 Residue Computation

A high resolution global face image is reconstructed based on Eigen-transformation method Equation 2.12. However, it is only a generic approximation of high-resolution human face image and thus may lack some details. There are differences between hallucinated high-resolution image and the original high-resolution image. Figure 2.3 shows the residue between a hallucinated facial image by Zhuang *et al.* (2007) and the original high-resolution facial image.

(a) (b) (c)

Figure 2.3: Examples Display of Residue between Hallucinated Face and Original Face Image. (a) Original High-resolution Face Image. (b) Hallucinated High-resolution Face by Zhuang *et al.* (2007). (C) Residue between Hallucinated Face Image and Original High-resolution Face Image.

In order to obtain the lost residue part of an hallucinated face, a compensation method is proposed as follows: The high-resolution hallucinated face y^g derived from Equation 2.13 is down-sampled to the same resolution as x^l, which is denoted by y^d. Then its residue s^l is calculated as:

$$s^l = x^l - y^d \tag{2.14}$$

This residue is a low-resolution image that represents local features for the test image. This low-resolution residue s^l needs to be enhanced to a high-resolution image which can be thought of as the local features of the test image in high resolution. Learning-based PCA is used to achieve this aim as explained below.

The high-resolution residue and the corresponding low-resolution residue training sets are constructed as follows. Each low-resolution training image is considered as a test image which is then enhanced to a high-resolution image R_i^h, using Equation 2.13. The high-resolution residue training samples S_i^h are the difference between

R_i^h and original high-resolution training sample F_i^h. Accordingly, the low-resolution training set S_i^l is derived by down-sampling S_i^h to the low-resolution.

$$S_i^h = F_i^h - R_i^h \tag{2.15}$$

$$S_i^l = S_i^h \downarrow \tag{2.16}$$

When a low resolution residue s^l is obtained from Equation 2.14, the corresponding high resolution residue s^h can be derived by projecting s^l onto the low resolution residue training images S_i^l. Then s^h is obtained as follows:

$$s^h = S^h K^{sl} + m^{sh} \tag{2.17}$$

where K^{sl} are the contributing coefficients of s^l projecting on the low resolution residue training images S_i^l and m^{sh} is the mean of the high resolution residue training images. Then the final enhanced image is obtained by:

$$y^s = y^g + s^h \tag{2.18}$$

This high resolution image y^s improves the quality of reconstruction results compared with Wang's work. The local features of the human face such as hair and a mustache can be well represented by this residue method. Figure 2.4 indicates the framework of this residue based PCA method (PCAR).

2.3.3 Recursive Residue Computation

Although the quality of the reconstructed face image in the last section is generally better than global faces hallucinated from Equation 2.13, it may be improved further by using a recursive method proposed in this section, which compensates more local

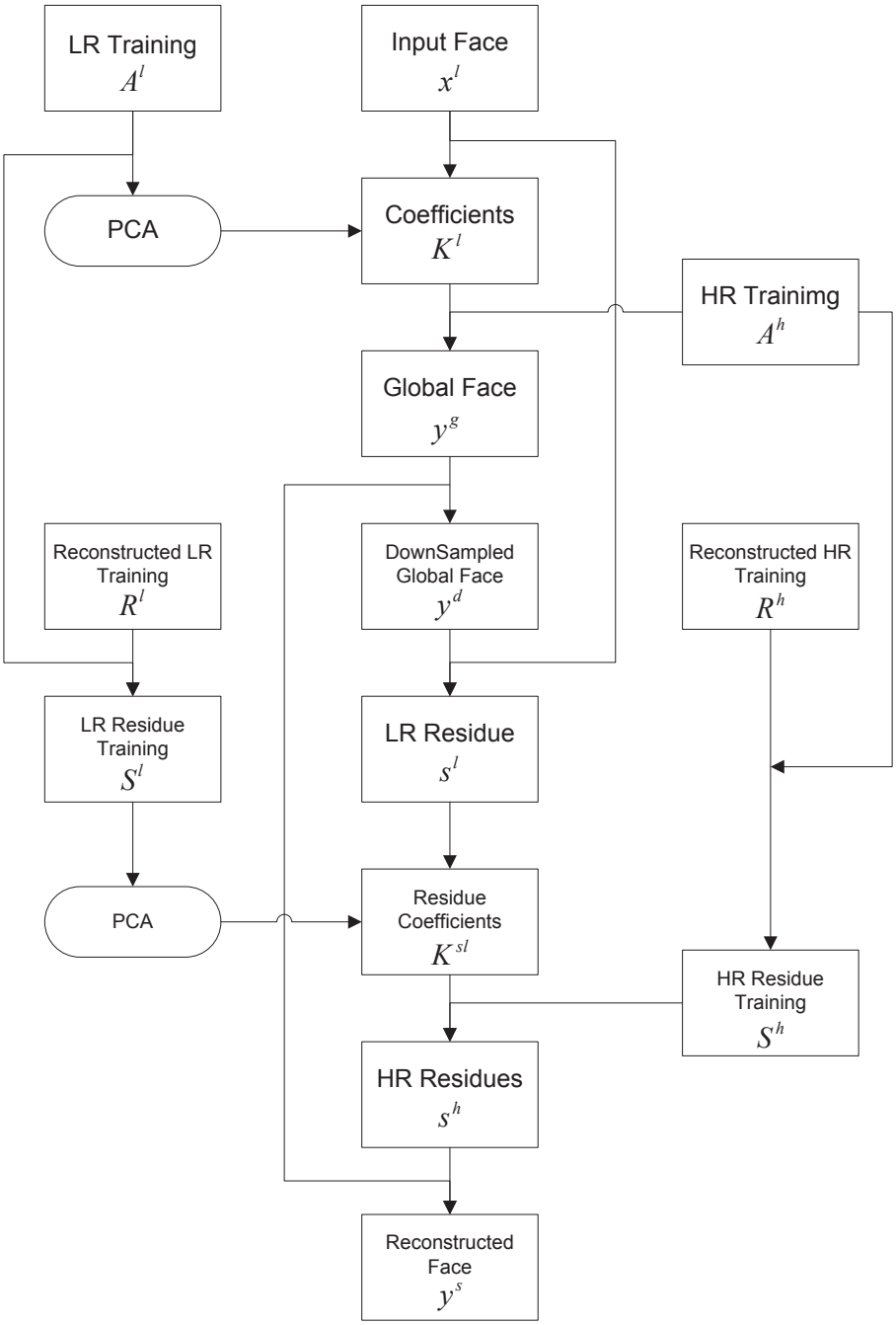


Figure 2.4: Process chart of PCA based residue (PCAR) method

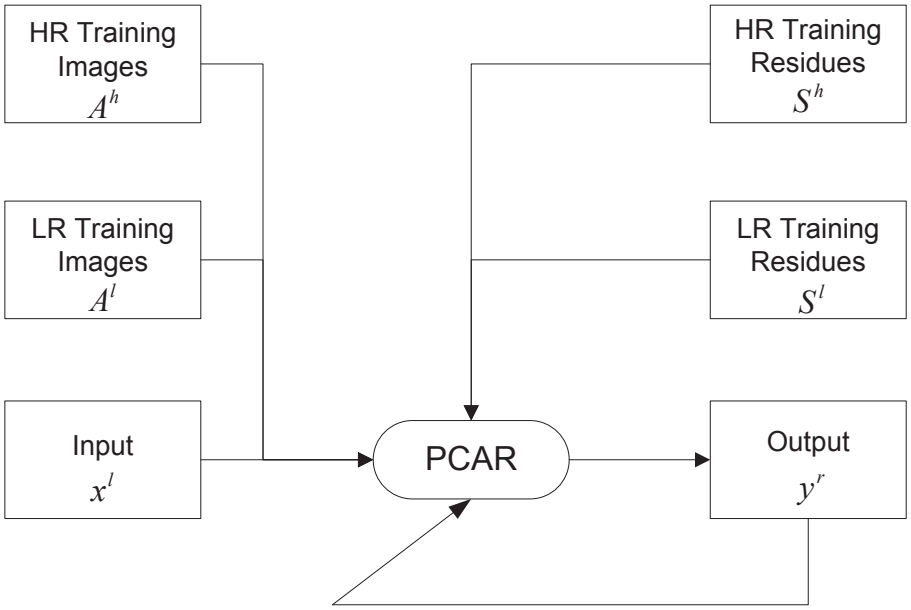


Figure 2.5: Recursive method

features. When an output image y^s is produced from Equation 2.18, it can be down-sampled to the low-resolution image y^{d2} and used to calculate the residue s^{l2} between it and the original low-resolution input image x^l again.

$$s^{l2} = x^l - y^{d2} \tag{2.19}$$

Similarly, the residue is projected onto the corresponding residue training sets and one can obtain a high resolution residue:

$$s^{h2} = S^h K^{sl2} + m^{sh2} \tag{2.20}$$

Finally a high-resolution image can be obtained by the proposed recursive method (Fig. 2.5).

$$y^r = y^s + s^{h2} \tag{2.21}$$

Although this recursive residue framework can be used in face image enhancement for infinite rounds, fewer local features will be obtained when the number of rounds

increases. The experiments show that one or two rounds will be sufficient.

2.3.4 Two-stage PCAR Computation

When the resolution difference between the high and low resolution images is significantly large, due to loss of information, it may be difficult to enhance a low resolution image directly with the proposed technique in last section. In order to obtain more information from training sets, a "medium" resolution training set is introduced when the low-resolution and high-resolution images have large deviations. We use three sets of training samples with three resolutions here respectively: (32×24), (64×48) and (128×96). We first enhance a low-resolution input image (32×24) to a middle stage (64×48) by using PCAR. A middle-stage output y^{mr} with the resolution (64×48) is produced. In the second stage y^{mr} is assumed to be an input image. By learning from the training data, the middle-stage image y^{mr} can be enhanced to high resolution by using the same algorithm as the first stage (see Fig. 2.6).

In the first stage we derive the mid-stage output y^{mr} according to Eq. 2.21:

$$y^{mr} = y^{ms} + s^{mh2} \tag{2.22}$$

In the second stage y^{mr} is used as the input of this method. We first enhance y^{mr} to a high-resolution global face by learning from two sets of training samples: A^h and A^m. Then this global face is down-sampled to a low-resolution face image. The residue between this low-resolution face image and original low-resolution test image is enhanced to a high-resolution residue image by residue learning process. In this stage, *We down-sample the global image to the low-resolution instead of mid-resolution. This is because the mid-resolution face* y^{mr} *is an approximate image*

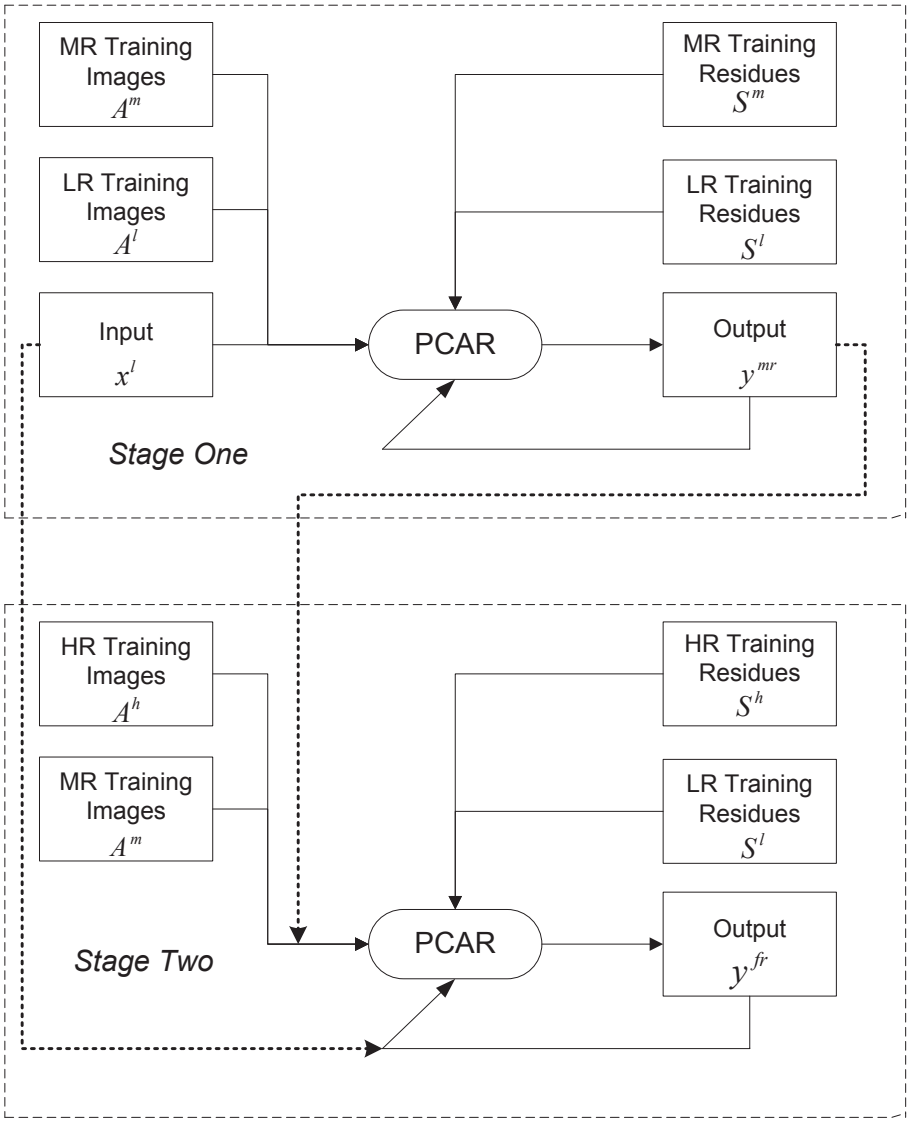


Figure 2.6: Process chart of two-stage method

which cannot be used to determine the true deviation. Thus, this two-stage program can benefit from the two-stage learning processes while compensating the residues correctly. The final output of the two-stage method is:

$$y^{fr} = y^{fs} + s^{fh2} \tag{2.23}$$

Similar to the recursive method, this two-stage framework can be further expanded to a three-stage, four-stage or more, depending on the difference between the low resolution of input images and the high resolution of target images. In this chapter we only use the two-stage method for experiments to validate the proposed idea.

Now we can present a detailed algorithm as below:
Step 1: Given a set of high-resolution and low-resolution training pairs A_i^h, A_i^l and an input low-resolution face image x^l, compute the eigen-subspace coefficients K^l according to Equation 2.4, Equation 2.8 and Equation 2.9.
Step 2: Setting each low-resolution training sample A_i^l as an input image, a set of high-resolution images R_i^h are generated from Equation 2.13. The residue training samples are obtained from Equation 2.15.
Step 3: Given an low-resolution input face image x^l, a global face is obtained from Equation 2.13.
Step 4: A residue image is derived from Equation 2.14. Then it is enhanced to a high-resolution residue according to Equation 2.17 and Equation 2.18.
Step 5: A high-resolution image y^s is generated by PCAR method from Equation 2.18.
Step 6: Repeat steps 4-5, a high-resolution face image y^r based on recursive residue frame work, can be obtained from Equation 2.19, Equation 2.20 and Equation 2.21.
Step 7: Assume that there are three training sets: A_i^h, A_i^m, A_i^l, which represent three resolutions. We first execute step 1-6, enhancing a low-resolution input face image x^l to a middle-resolution one y^{mr} from Equation 2.22. Then steps 1 to 6 are executed to obtain a final high-resolution image y^{fr} with a middle-resolution input image y^{mr}

from Equation 2.23.

2.4 Experiments and Discussion

2.4.1 Data and Evaluation

Before the experiments, all the face images are aligned manually by fixing the centers of the eyes and mouth with the resolution (128×96). The cropping procedure is as follows:
Step 1: For a face image with the resolution of 128×96, the centers of left eye and right eye are first clicked manually.
Step 2: The face image along the horizontal axis is interpolated till the distance between left eye center and right eye is 48, which takes account half of the target resolution in horizontal axis.
Step 3: The center of mouth is then click manually.
Step 4: The face image is interpolated along vertical axis. The vertical distance between eyes and mouth is set as 5/12 of the whole face image. In other words, the vertical distance between eyes and forehead accounts for one third of the whole length and the vertical distance between the mouth to the bottom of the face is set one fourth of the whole image.

All the cropped images are then down-sampled to the low resolution of (32×24) according to Equation 2.1.

The results of the experiment are compared using the Peak Signal Noise Ratio

(PSNR) defined as below.

$$PSNR = 10 \log \left[\frac{255^2}{\frac{1}{M \times N} \sum \sum (I_O - I_R)^2} \right] \tag{2.24}$$

where M and N are the numbers of rows and columns of the high-resolution image. I_O and I_R represent the original high-resolution testing image and reconstructed high-resolution image respectively.

2.4.2 Proposed methods

2.4.2.1 PCAR Method.

According to Equation 2.13 and 2.17 the global faces and residues are constructed. They are then combined to generate the final output Equation 2.18. Figure 2.7 shows the images of low-resolution face, global face, our three algorithms and original high-resolution face, It can be seen that the our output images combined with residue have more local features than those global face images which were obtained from the eigen-transformation method. For example the hair of the third person and the moustache of the fifth person are reconstructed better by using our method when compared with global face images.

2.4.2.2 Recursive and Two-stage methods.

After obtaining a high-resolution face image from the residue based PCA method, we can increase its quality by recursively using the residue algorithm from Eq. 2.21. However, this recursive method can not be used infinitely since no more useful information can be learned and more errors may occur from after several rounds. Figure 2.8 indicates that the average PSNR values of testing images reach the peak

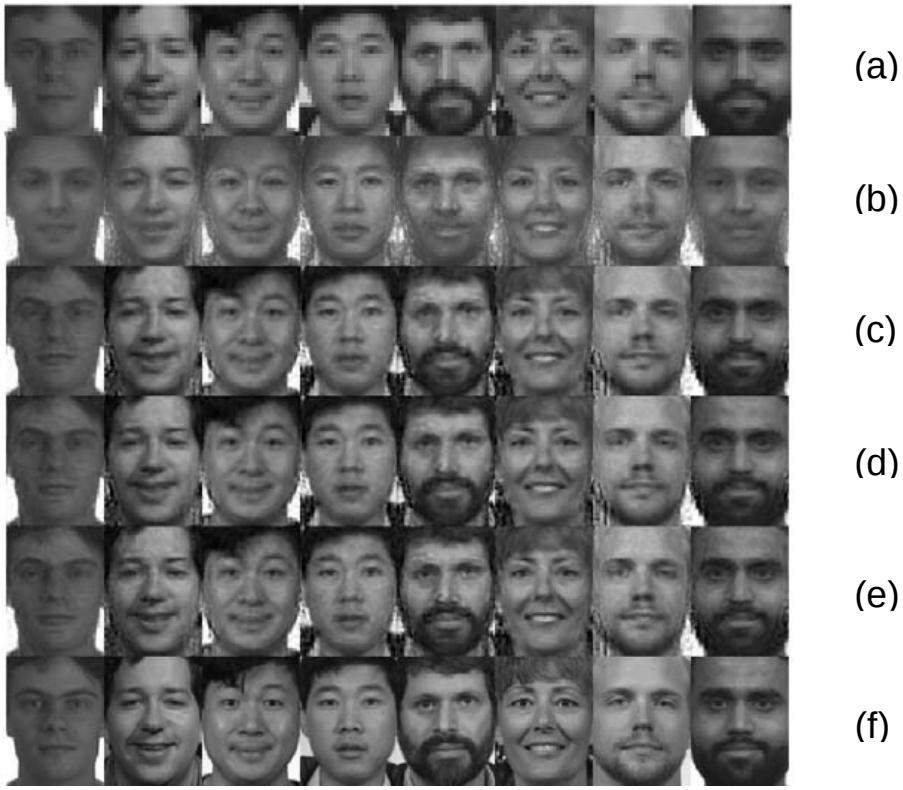


Figure 2.7: Experimental results using PCAR. (a) Input 32×24 low-resolution images. (b) Global face. (c) Reconstructed images using PCAR method. (d) Reconstructed images using recursive residue compensation. (e) Reconstructed images using two-stage compensation. (f) Original 128×96 high-resolution images

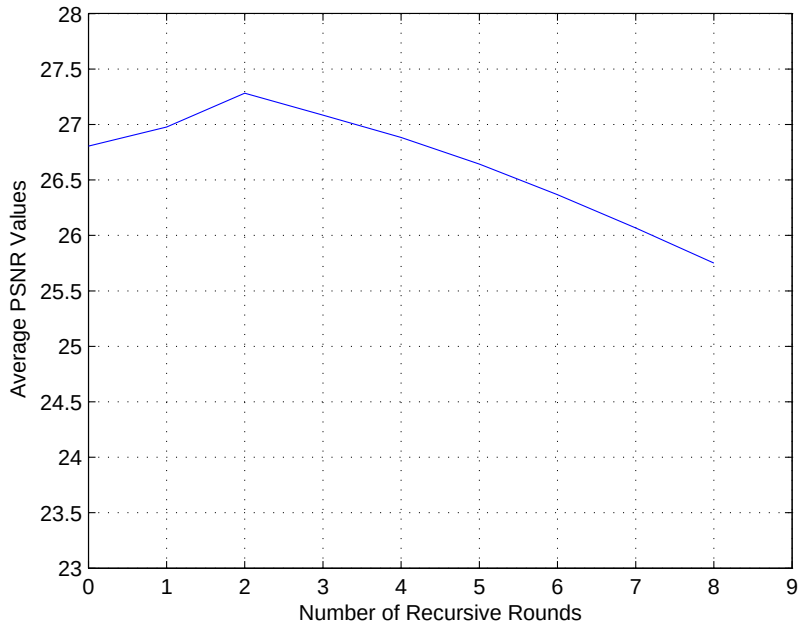


Figure 2.8: PSNR values in terms of different recursive rounds

when the number of recursive rounds is two and it will drop when more rounds of residue compensations are performed.

In order to obtain more information from training data, we also set up a two-stage experiment. Firstly, we enhance the low-resolution testing face to an image with mid-stage resolution and then enhance it to a high-resolution face image using Eq. 2.23. In this experiment we can improve the result since more information is added due to the mid-stage training data. Figure 2.9 shows the values of PSNR of the our three methods (Residue, Recursive and Two-stage) when the number of training samples is 200. It can be observed that the two-stage approach can improve the performance of the recursive method.

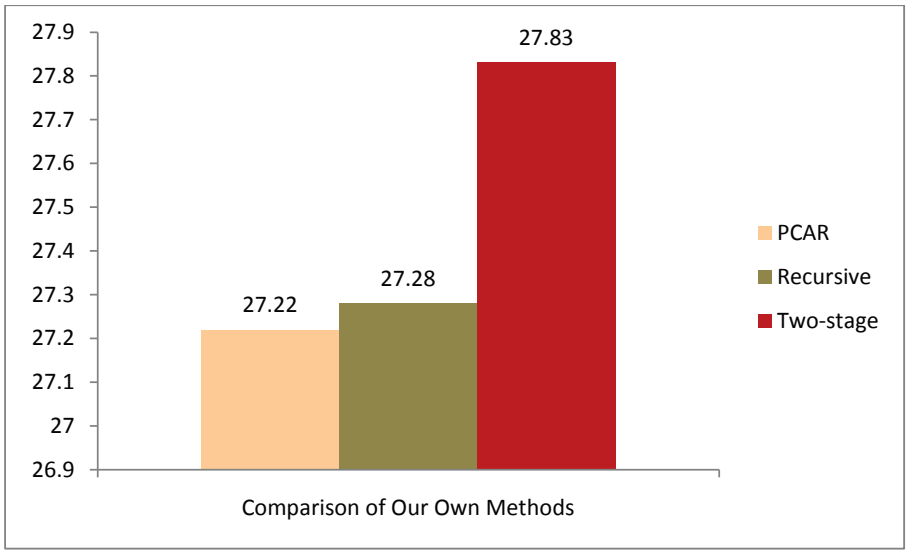


Figure 2.9: Comparison of our own methods in terms of PSNR

2.4.3 Comparison of Different Down-sampling Methods

As discussed in 2.2.1.1, there are different methods to down-sample high-resolution image to low-resolution. Experiments are designed to compare the influence in terms of different down-sampling methods. We can see from Figure 2.1 that general average based down-sampling method (Figure 2.1 (b)) and frequency based down-sampling method (Figure 2.1 (d)) perform better than sampling method (Figure 2.1 (c)) in visual quality. An experiment is designed to compare the hallucinating results between different down-sampling methods. Figure 2.10 displays the comparison in terms of three different down-sampling methods.

Both Figure 2.1 and Figure 2.10 show the visual effects of different down-sampling methods. Figure 2.1 demonstrates the low-resolution images in terms of different down-sampling methods. Figure 2.10 illustrates the hallucination effects of different down-sampling methods. Table 2.1 also shows effects of the hallucination results in PSNR and RMSE.

(a) (b) (c) (d) (e)

Figure 2.10: Examples Display of Hallucinated Faces through Different Down-sampling Methods. (a) Original Low-resolution Face Image (32×24). (b) Hallucinated High-resolution Image from Low-resolution Facial Image Down-sampled by Equation 2.1 (32×24). (c) Hallucinated High-resolution Image from Low-resolution Facial Image Down-sampled by Sampling Theory (32×24). (d) Hallucinated High-resolution Image from Low-resolution Facial Image Down-sampled in Frequency Domain (32×24). (e) Original High-resolution Facial Image (128×96).

Table 2.1: PSNR and RMSE of Hallucinated Faces in terms of Different Down-sampling methods.

	by Average	by Sampling	in Frequency Domain
$PSNR$	27.73	21.47	23.04
$RMSE$	10.53	21.84	18.92

2.4.4 Comparison with Other Methods

In this section the results of the experiment of the proposed two-stage approach are compared with approaches by Wang and Tang (2005), Liu *et al.* (2007) and Yang *et al.* (2008a), when the number of training samples is 200, which are shown in Fig. 2.11. The first row is the low-resolution testing images. The second, third, fourth and fifth rows are the results of Wang and Tang (2005), Liu *et al.* (2007) and Yang *et al.* (2008a) respectively. The sixth row lists images of our two-stage method and the original high-resolution images are in the last row. It can be seen that our method obviously performs better than Wang's method. The images obtained by our two-step method have more detailed information and local features than Wang and Tang (2005) such as hair and moustache. We mainly focus on comparing our method with Liu *et al.* (2007) and Yang *et al.* (2008a), as they are good representations of holistic hallucination model and patch-based hallucination model respectively. Figure 2.12 shows a line chart that indicates the trend of average PSNR values between a reconstructed high-resolution image and original high-resolution image when an input low-resolution image is enhanced using Wang and Tang (2005), Liu *et al.* (2007), Yang *et al.* (2008a) and our methods, separately. The horizontal axis represents the number of training samples and the vertical axis shows the average PSNR values. It can be seen that our method performs better in the case of small number of training samples. More importantly, Liu *et al.* (2007) that based on probabilistic model has a higher computational cost and Yang *et al.* (2008a) in-

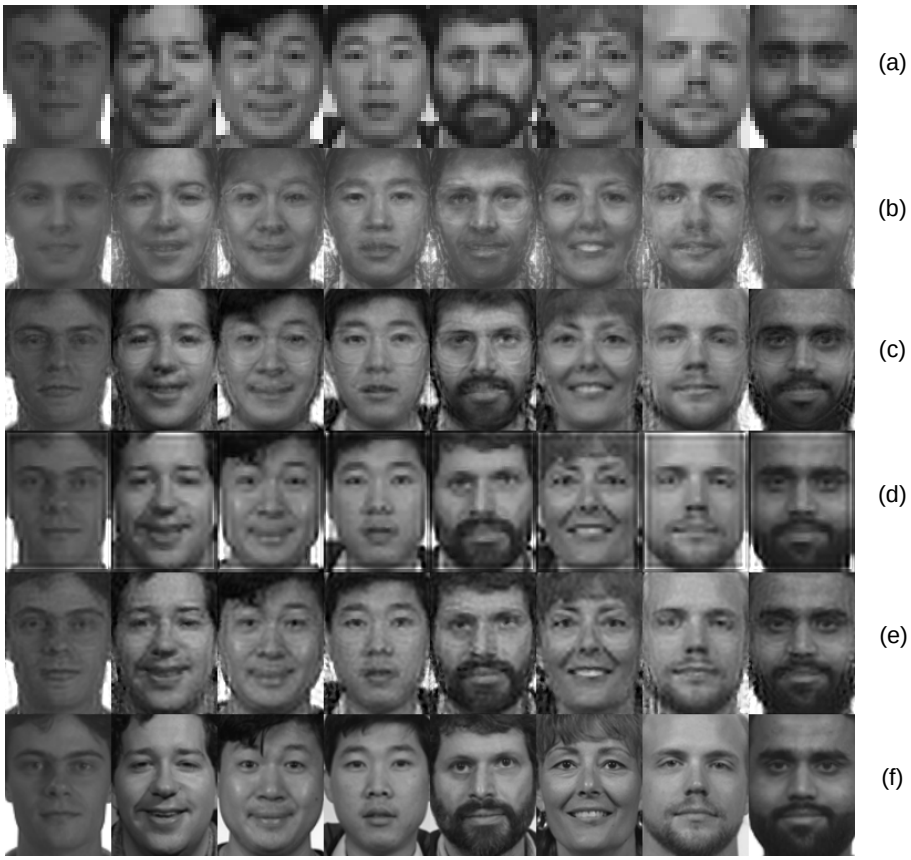


Figure 2.11: Comparison of different methods. Experimental results using PCAR. (a) Input 32×24 low-resolution images. (b) Wang's eigen-transformation approach. (c)Liu's two-step approach. (d) Yang's method. (e) Reconstructed images using our two-stage method. (f) Original 128×96 high-resolution images

cludes optimization approaches both in generating global faces and residues. As the proposed approach in this chapter is only a linear combination of training matrices, its execution speed is much faster than Liu *et al.* (2007) and Yang *et al.* (2008a).

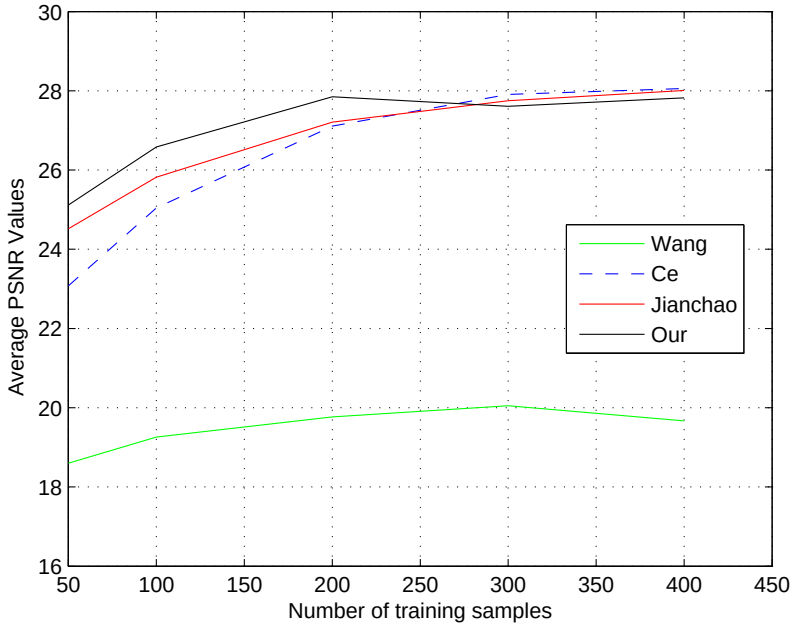


Figure 2.12: Average PSNR of different methods on different training samples

2.5 Summary

In this chapter an innovative framework for image enhancement is proposed. In the proposed framework, the input low-resolution image is first enhanced to a high-resolution face based on Principal Component Analysis (Wang and Tang, 2005). Then, a residue image is computed to derive the local features. A reconstructed image is achieved by using recursive and two-stage methods in order to improve the quality. This framework improves the enhancement performance compared with general PCA eigen-transformation method and is also computationally efficient. However, with large training samples, the approaches proposed in Liu *et al.* (2007) and Yang *et al.* (2008a) are supposed to perform slightly better. This phenomenon may be due to the requirement of large training samples in their approaches.

Different Down-sampling methods lead to different visual effects of low-resolution facial images. They also further affect the hallucination results. However all the

down-sampling methods can only be performed by computers in labs. In real face capturing systems, low-resolution faces are not obtained from these down-sampling methods. Instead small faces are often captured in far distances. The far distance causes the low resolution of captured face images. How the distance between camera and object affects the face hallucination and face recognition will be discussed in other chapters of this thesis.

Chapter 3

Hallucinating Faces in Patches

3.1 Introduction

3.1.1 Holistic and Patch based Face Hallucination

As discussed in Chapter 1.3.1, hallucinating faces in holistic images may introduce noises. These hallucinating approaches can achieve good performances in global face features. However, they do not perform well for local features. Especially in hallucinating the chin area of faces, these holistic models generate noises. Though the proposed approach in Section 2 has good performances in terms of PSNR and RMSE, the effectiveness is not so significant.

In order to solve this problem, a patch based face hallucination approach is proposed in this chapter. Patch based super-resolution approaches can produce smooth high-resolution images due to overlapped patches. However, these methods may ignore the specific properties of faces and only enhance them as generic images. Many face features are lost during the super-resolution process. In order to add more face features into hallucinated images, a residue compensation step is proposed in this chapter. After hallucinating faces in overlapped patches, a frequency based compensation is conducted to render the facial global features. This frequency based compensation process adopts Curvelet frequency as face features. We use this step to render lost face features during the previous patch based hallucinating

step. This rendering will increase the holistic features in the hallucinated faces as demonstrated in experiments.

3.1.2 Assumptions of the Proposed Method

Most current face hallucination techniques are proposed in spatial domain, which often require a large amount of computation. This is due to the large number of training data. In terms of frequency domain based super-resolution techniques, though they are efficient, they can hardly represent detailed facial features without the learning process (Milanfar, 2010). In order to synthesize the advantages of the methods in both spatial and frequency domains, we propose a face hallucination method combined with pre-selection processes based on Curvelet features. As we know from Liu *et al.* (2007), in face hallucination, face images include two types of features: global features and local features. Global features describe the common human features like eyes, mouths and noses. The local features represent the specific features of an individual face image. However, in traditional two step approach, global features and local features are described in spatial domain. Instead, we adopt the Curvelet frequency features to describe those two types of features in this chapter.

Feature 1 *Global face features which include most of the low-frequencies of human faces;*

Feature 2 *Local face features which consist of the high-frequencies in face images.*

The previously proposed global features and local features (Liu *et al.*, 2007) are the separated parts of digital facial images. This separation divides face images into two parts in spatial domain. However, these two parts could only be approached

approximately by learning algorithms and residual methods. Thus there is not any algorithms which specifically define which part of an image is global features and which part is local features. In this chapter, we define the global features and local features in frequency domain. For any digital image including facial image, the high frequency components and low frequency components can be defined accurately through digital image processing theory. Therefore, both these two parts could be approached easily and accurately.

3.1.3 Contributions of this Chapter

With these two features, we design the learning based face hallucination method in two steps. One is the low-frequency face image hallucination and the other is the high frequency based face image hallucination. In order to reduce computational cost and reconstruction errors, Curvelet features of a testing image are used to select the associated training samples in both two steps. For given full data samples both in high and low resolution used for learning, we first decompose the pairs into Curvelet frequency domain. In fact, the fine Curvelet coefficients describe the high frequency components of face images, and the coarse Curvelet coefficients represent the low frequency part of face images. In order to reduce computational complexity, we only use two layers of Curvelet coefficients in this chapter. Now for each image, we have both the fine and the coarse coefficients. Then we use K_{th} Nearest Neighbors algorithm to find K_1 images, which have the best matched coarse coefficients with the coarse coefficients of testing face image. Similarly, we also can find K_2 images which have the best matched fine coefficients compared with the fine coefficients of testing image. In this chapter, we use the selected K_1 images as the training samples in first step and the selected K_2 images as the training set in second step.

In the first step, we estimate the high resolution global features for a low resolution

testing face image using the sparse representation learning method. The examples of low resolution images are shown in the first column of Figure 3.8. In the second step, we produce a residue training data, and estimate the high resolution residue which compensates the missing local features for the global face in the first step. By learning the Curvelet features of the residue training pairs, we estimate the Curvelet features of the high-resolution residue face for a testing image and construct the high resolution image by using the Inverse Curvelet transformation. Figure 3.1 shows the implementation steps of our method.

The main contributions of chapter have three parts. 1). We extract two types of features based on Curvelet frequency domain: low-frequency part, which represents the global features of human faces; high-frequency part, which demonstrates the local features of human faces.
2). We use the Curvelet features to select training samples for a testing image in both global and local hallucination algorithms, which reduced the computational cost significantly due to the selected smaller training data.
3). In high frequency feature estimation, we hallucinate this residue image through the inverse Curvelet transformation.

This chapter is organized as follows: the proposed algorithms are illustrated in Section 3.2. More specifically, in Section 3.2.1 we extract image features in Curvelet frequency domain and select training samples based on global and local features respectively. Then we hallucinate the low-resolution faces to global high-resolution faces by employing the sparse representation algorithm in Sec. 3.2.2. In Sec. 3.2.3, the residual faces are derived from the Inverse Discrete Curvelet Transformation. Experimental results are illustrated in Sec. 3.3 with comparison with other approaches. Conclusion and summary are stated in Sec. 3.4.

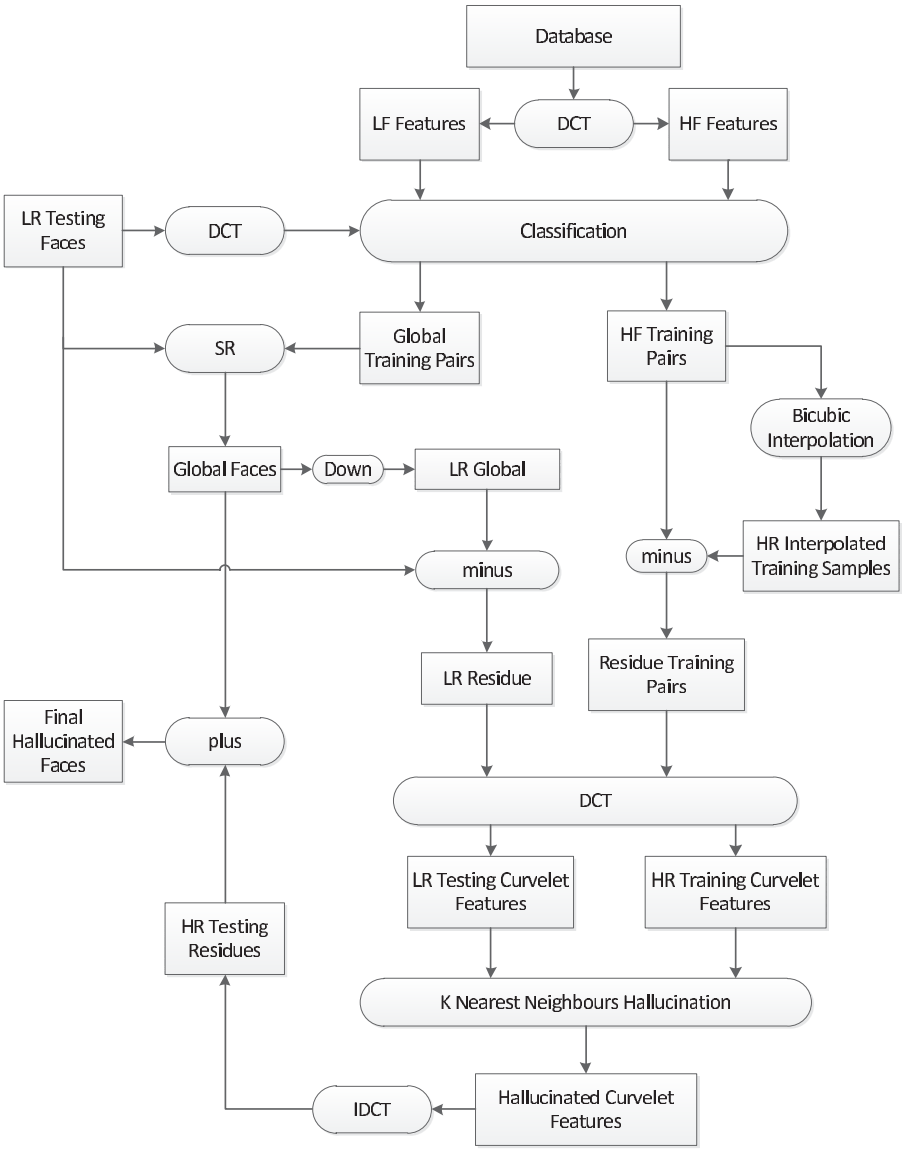


Figure 3.1: Process Diagram. DCT:Discrete Curvelet Transformation. LF:Low Frequency. HF:High Frequency. LR:Low Resolution. HR:High Resolution. SR:Sparse Representation Hallucination. Down:Down Sample. IDCT:Inverse Discrete Curvelet Transformation.

3.2 Proposed Algorithm

3.2.1 Curvelet Based Training Sample Selection

3.2.1.1 Curvelet Overview

Curvelet was first proposed by Candès and Donoho (1999), and then developed to the second generation in 2006. Both of them are fast and accurate (Candes *et al.*, 2006). As a multi-scale representation, Curvelet has been designed to overcome the shortages of wavelet when dealing with singularities in signal processing, data compression and image denoising. For example, Curvelet transformation provides far more sparse representations than wavelet for objects with edges. Human faces in gray scale images are good examples of those edged images. Previous work (Mandal *et al.*, 2009) has proved that Curvelet can ideally extract the human face features and address the face recognition problem. There are two Discrete Curvelet Transformation versions, namely USFFT and Wrapping, we adopt the latter one. For a 2D image, Curvelet transformation is performed as follows (Candes *et al.*, 2006):

As from Candes *et al.* (2006), Curvelet coefficients can be derived as follows:

1. Apply the 2D FFT and obtain Fourier samples $\hat{f}[n_1, n_2]$ of $f[t_1, t_2]$, where $0 \leq t_1, t_2 \leq n, -n/2 \leq n_1, n_2 < n/2$.

2. For each scale j and angle l , compute the product $\tilde{U}_{j,l}[n_1, n_2]\hat{f}[n_1, n_2]$, where $\tilde{U}_{j,l}[n_1, n_2]$ is the discrete localizing window.

3. Wrap this product around the origin and obtain $\tilde{f}_{j,l}[n_1, n_2] = W(\tilde{U}_{j,l}\hat{f}[n_1, n_2])$, where W is the wrapping function.

4. Apply the inverse 2D FFT to each $\tilde{f}_{j,l}$ and collect the discrete Curvelet coefficients $C\{j\}\{l\}(k_1, k_2)$.

where j and l represent the scales and angles and k_1, k_2 denote the position of Curvelet coefficient matrix.

3.2.1.2 Using Curvelet Features to Select Training Data

In this chapter, we use Curvelet features to select training samples for each testing image. First it is used to select the training samples with global features. As we know, training samples play a key role in the single frame face hallucination. As a result, how to carefully select training samples is a key issue for face super-resolution. Here we select training sample from the face data sets through Curvelet coefficients before hallucinating global faces. we transfer the low-resolution testing image and the training images into Curvelet domain, where we calculate the nearest neighbors in the low-resolution coarse coefficients of the testing image. Then, those images whose coarse coefficients are the nearest neighbors of the coefficients of the testing image are selected as the training samples for such testing image.

Specifically, let $A_i^h = [A_1^h, A_2^h, \cdots, A_n^h, \cdots]$ denote the high resolution face database, and $A_i^l = [A_1^l, A_2^l, \cdots, A_n^l, \cdots]$ denote the corresponding low resolution face database. we first decompose the i_{th} low resolution face image into Curvelet features $C_i\{j\}\{t\}(k_1, k_2)$, where $(i = 1, 2, \cdots, n, \cdots)$. Here j and t represent the scales and angles of Curvelet coefficients respectively. k_1, k_2 indicate the coefficient matrix positions. We simplify $C_i\{j\}\{t\}(k_1, k_2)$ to be $C_i\{j\}\{t\}$ in the left part of this chapter. When a test low resolution image x comes, it is decomposed to Curvelet domain to get the coefficients $C_x\{j\}\{t\}$. In order to reduce computational cost, we set scale $j = 2$, angle $l = 8$ in this chapter. Once the Curvelet coefficients $C_x\{j\}\{t\}$ are derived, the coarsest

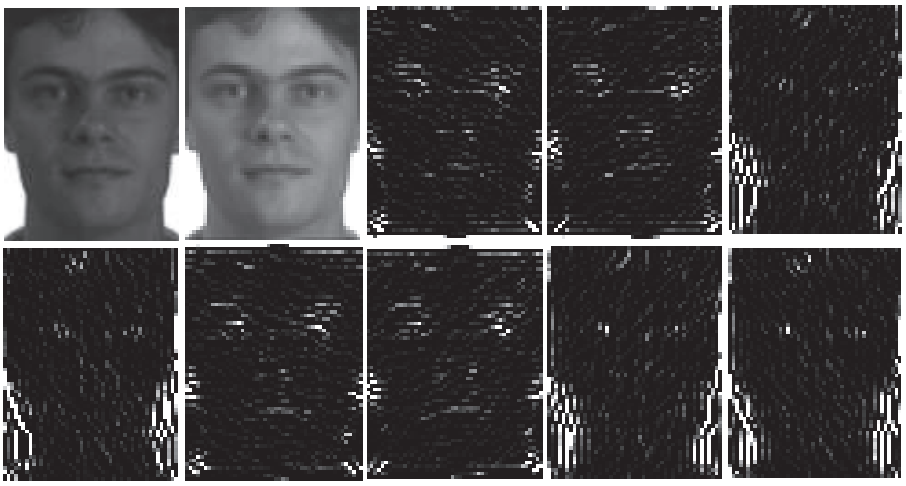


Figure 3.2: Curvelet Coefficients. The top left is the original face image. The left 9 images are the low-frequency image of the first layer and the 8 high-frequency images of the second layer respectively.

coefficient $C_x\{1\}\{t\}$ $(t = 1)$ represent the low frequency feature and the finest coefficients $C_x\{2\}\{t\}$ $(t = 1, 2, \cdots, 8)$ represent the high frequency feature. Figure 3.2 shows the Curvelet coefficients of a testing image x. For the convenience of display, all the feature images are plotted in the same size.

Now we have a set of Curvelet coefficients $C_i\{j\}\{t\}(i = 1, 2, \cdots, n, \cdots)$ for the face database and $C_x\{j\}\{t\}$ for the testing face. we first utilize the low frequency feature $C_x\{1\}\{t\}$ and calculate its K_1 nearest neighbors in $C_i\{1\}\{t\}(i = 1, 2, \cdots, n, \cdots)$. For computational efficiency, we first adopt the Principal Component Analysis (PCA) to reduce the dimension and then apply the nearest neighbor algorithm for K_1 elements selection. Consequently, we have K_1 selected training samples for the global face enhancement, named as $I = [I_1, I_2, \cdots, I_{K_1}]$.

Similarly, for high frequency components, we also use the local features to select another set of training samples. In order to achieve this, we treat the whole second scale of $C_x\{2\}\{t\}$ as one image and resize all the coefficients to form one column. This one column image represents the high frequency features through t different

angles ($t = 8$). Similarly as global case, we also adopt PCA for dimension reduction. We keep first 20 eigenvectors corresponding to 20 largest eigenvalues, which keep the most of the data energy. After that we select K_2 nearest neighbors in $C_i\{2\}\{t\}(i = 1, 2, \cdots, n, \cdots)$ as local feature training samples, namely $F = [F_1, F_2, \cdots, F_{K_2}]$.

The main reasons that we utilize classifier before learning processes is to reduce computational cost and also to minimize errors. We can see that after using classifies, the number of training data does not rely on the whole database. Instead, we reduce it according to the value of K and N in pre-selection step. We further reduce the computation by decompose spatial faces into Curvelet frequency features. Thus we avoid dealing with the huge dimensions of spatial face images. Another advantage for Curvelet based classification is to reduce errors in learning process. In the full face database, some images involve very specific characteristics which may increase the errors in the learning program. According to the global and local constraints of face images, the proposed global classifier classifies the faces which have the close characteristics in global features and local classifier classifies the images which have the similar characteristics in local features. Thus this proposed method avoids learning the unnecessary features which may increase the errors.

3.2.2 Hallucinating Faces via Sparse Representation

In sparse sensing theory, signals can be represented by basis signals through a well-constructed dictionary (Donoho, 2006a; Candès and Wakin, 2008). If we think of face images as a kind of signal, one face image can be represented by a set of face basis when there is a large training data set. A generic image sparse representation algorithm has been proposed, which estimates the high-resolution image from raw image patches (Yang *et al.*, 2008b). As a kind of images, face image can also be recovered by redundant training samples. Follow Yang *et al.* (2008b), we first take

face images as generic images and enhance low-resolution face images in a patch-based model. However, face images have their special properties. In reality the quality of face super-resolution depends on how well the dictionary is designed. In this chapter, we design the training dictionary through the Curvelet features selected in previous section.

When face images are enhanced in overlapped patches, the face images are taken as generic images. Thus the global face features might be lost. In order to compensate the lost global features, we add the pre-selection procedure in previous section. Only those faces who have the similar global features are selected as the training samples for patch-based face image super-resolution. The low-resolution testing patches learn the relationship in high-resolution training patches. Those high-resolution training patches come from the faces who have the similar global features as the testing face image. Thus some of the global features will be kept during the learning process.

The selected training set $I = [I_1, I_2, \cdots, I_{K_1}]$ in Section 3.2.1 are low-resolution training images which have the closest global features as the low-resolution testing face x. The selected low-resolution training images $I = [I_1, I_2, \cdots, I_{K_1}]$ have their high-resolution pairs $I^h = [I_1^h, I_2^h, \cdots, I_{K_1}^h]$. These high-resolution training image are firstly divided into overlapped patches P^h. Correspondingly, there are also overlapped low-resolution patch pairs P^l in the low-resolution training data. The learning dictionary pairs D_h and D_l are selected as follows: The testing face image x is firstly divided into overlapped patches. Similarly, the each of the low-resolution training faces $I = [I_1, I_2, \cdots, I_{K_1}]$ is also divided into overlapped patches. For every patch p inside x, we only consider the patches which have the same positions in training faces $I = [I_1, I_2, \cdots, I_{K_1}]$ to be the dictionary. This can force the testing patches to learn the images features which have the same position in high-resolution space. For example, patches around the eyes can only learn from the high-resolution patches around the eyes. Some special features might be kept in this learning pro-

cess. In the meanwhile, the learning algorithm avoid to learn from a large amount of training samples. Thus the computational costs are reduced.

Suppose the selected training set in Section 3.2.1 to be $D_h(high - resolution)$ and $D_l(low - resolution)$ for a testing patch p, the high-resolution patch image can be reconstructed by solving the following optimization problem:

$$\begin{aligned} \hat{\alpha} &= argmin \parallel \alpha \parallel_0 \\ s.t. \quad & D_l\alpha = p \end{aligned} \tag{3.1}$$

where α is the sparse representation coefficients in 0 norm.

Generally this 0 norm problem is NP hard to be solved. However, according to Donoho (2006b), if α can be sufficiently sparse, p can be recovered by solving the problem in ℓ_1 norm.

$$\begin{aligned} \hat{\alpha} &= argmin \parallel \alpha \parallel_{\ell_1} \\ s.t. \quad & D_\ell\alpha = p \end{aligned} \tag{3.2}$$

where α is the sparse representation coefficients in ℓ_1 norm. The solution of this problem has been presented by solving the equivalent Lagrange multipliers Yang *et al.* (2010, 2008b).

$$min \parallel \alpha \parallel_1 + \frac{1}{2} \parallel D_\ell\alpha - p \parallel_2^2 \tag{3.3}$$

Once being solved, this sparse representation coefficients α in low-resolution face images can then be mapped to the high-resolution. The global high-resolution face can be reconstructed as:

$$y = D_h\hat{\alpha} \tag{3.4}$$

In practice, the sparse representation does not perform well when treating the whole face as one signal. We first divide face images into overlapped patches and enhance

Table 3.1: PSNR and RMSE of Hallucinated Faces in terms of Different Patch Overlapped Sizes.

	Non-Overlap	1-Overlap	2-Overlap	3-Overlap
$PSNR$	30.22	30.96	30.97	31.13
$RMSE$	7.94	7.30	7.16	7.16

those patches respectively. Then by combining those patches, we can derive the high-resolution faces. In this chapter the size of each patch is set as 4×4 in low-resolution images and 16×16 in high-resolution images. The overlapped size is set as 3 and 12 respectively. The visual differences of holistic model and patch-based model are shown in Figure 3.3. The low-resolution face images are hallucinated by both holistic based sparse representation method and patch-based sparse representation method. The training data are exactly the same with 200 people. The visual hallucinated results are demonstrated in Figure 3.3. As can be seen clearly that the hallucinated faces by holistic model have a good representation of global face features. The images around eyes are very clear. However, noises exist around the chin and cheeks area. The hallucinated faces by patch-based model are very smooth. But some global face features are lost.

When merging the patches, the values of overlapped pixels in the high-resolution face are the average values of pixels in the same position. The experiment of different overlapping sizes is shown in Figure 3.4 and Table 3.1. The visual effects in terms of overlap sizes are not obvious, as can be seen from Figure 3.4. However, the hallucinated performances in terms of PSNR and RMSE are enhanced when the overlap sizes increase.

(a) (b) (c) (d)

Figure 3.3: Comparison between Holistic Model and Patch-based Model. (a) Low-resolution Face Images. (b) Hallucinated Face Images by Holistic Model. (c) Hallucinated Face Images by Patch-based Model. (d) Original High-resolution Face Images.

(a) (b) (c) (d)

Figure 3.4: Our approach in terms of different overlapping sizes. (a) Hallucinated Faces without Patch Overlap for each 4×4 Patch. (b) Hallucinated Faces with One Pixel Overlap for Each 4×4 Patch. (c) Hallucinated Faces with Two Pixel Overlap for Each 4×4 Patch. (d) Hallucinated Faces with Three Pixel Overlap for Each 4×4 Patch.

3.2.3 Residue Face Enhancement in Curvelet

A global face constructed in last section only represents the low frequency information of a face image, thus some detailed features might be lost. Residue face enhancement is required for this reason. We treat the residual face to be the special features towards each individual. For each person, this residue should be unique. In this chapter, a learning process in frequency subspace is developed.

More precisely, we have obtained a globally hallucinated face y in previous section. As we only have a testing image x, which is in low-resolution, we first down sample y to low-resolution, and derive its residue image s:

$$s = x - Down(y) \tag{3.5}$$

where $Down$ represents the downsample function. The down sample rate is set to be 4 in this chapter.

This residue image is thought of as the local features of the test image x. However, s only represents the low resolution local features and we need its high resolution local features s^h to render the globally hallucinated face y. In this section we derive this s^h through a frequency domain learning process. Since we have selected a training set $F = [F_1, F_2, \cdots, F_{K_2}]$, which are based on the local features of face images in previous section, we need to derive a high resolution residue training set $R^h = [R_1^h, R_2^h, \cdots, R_{K_2}^h]$ from F. This high resolution residue training set should represent the high frequency local features of the testing face image. For such purpose, we first down sample F to low resolution, then enhance them to the high resolution image set $\tilde{F} = [\tilde{F}_1, \tilde{F}_2, \cdots, \tilde{F}_{K_2}]$. As a result, R^h can be derived as follows:

$$R_{(i)}^h = F_{(i)} - \tilde{F_{(i)}} \tag{3.6}$$

where $i = (1, 2, \cdots, K_2)$.

In order to reduce the computational costs, we do not derive $\tilde{F}$ through sparse representation enhancement for each training image. Instead, we adopt the Bicubic interpolation (Hou and Andrews, 1978) to obtain a smooth high resolution training set $\tilde{F}$. Note that R^h represents the high frequency components of the selected training images, which are then down sampled to the low-resolution version $R^l = [R^l_1, R^l_2, \cdots, R^l_{K_2}]$. Now we have a testing low resolution image s and a set of training samples R^h and R^l. We first decompose both s, R^l and R^h into Curvelet subspace and derive the corresponding Curvelet coefficients. Let $C^l_s\{j\}\{t\}$ denote the coefficients of s, $C^l_{r_{(i)}}\{j\}\{t\})$ denote the coefficients of $R^l_{(i)}$ and $C^h_{r_{(i)}}\{j\}\{t\}$ denote the coefficients of $R^h_{(i)}$ $(i = 1, 2, \cdots, K_2)$, we formulate an optimization problem to obtain $C^h_s\{j\}\{t\}$, which is the Curvelet coefficients of s^h. For each element in $C^l_s\{j\}\{t\}$, which is a matrix, there is a corresponding matrix in $C^l_{r_{(i)}}\{j\}\{t\})$ and $C^h_{r_{(i)}}\{j\}\{t\}$. Here, we first formulate the following least square problem:

$$\begin{aligned} \underset{W_i}{\operatorname{argmin}} \parallel C^l_s\{j\}\{t\} - \sum_{i=1}^{K_2} W_i C^l_{r_{(i)}}\{j\}\{t\} \parallel_2^2 \\ s.t. \sum_{i=1}^{K2} W_i = 1 \end{aligned} \tag{3.7}$$

where $(i = 1, 2, \cdots, K_2; j = 1, 2; , t = 1, 2, \cdots, 8)$.

This is a standard optimization problem. The solution can be obtained in a close form as below: For each value of j and t, $C^l_s\{j\}\{t\}$ is a matrix. This matrix can be resized into a vector and represent one variable. There are total $N = j \times t$ $C^l_s\{j\}\{t\}$. These vectors are combined as a matrix. Similarly, each $C^l_{r_{(i)}}\{j\}\{t\})$ is resized to be a vector. Now we follow the following algorithm:

Step 1: For each j and t, there exists a unique vector that represents the Curvelet feature in this position. This vector is denoted as vector C_s.

Step 2: For each vector C_s, the K_2 neighbors have been found at the same position of the Curvelet features in training data. We denote this training matrix as C_r. C_r

is determined in our training sample pre-selection with K_2 columns.

Step 3: Vector C_s is subtracted from each column of matrix C_r. We denote this matrix as $C_r s$. The covariance matrix is computed as $C = C_r s' * C_r s$.

Step 4: Solve the linear system $C * W_i = 1$.

Step 5: The solved weight W_i is normalized to be 1.

Step 6: Repeat Step 1 to Step 5.

Step 7: The high-resolution Curvelet coefficients are reconstructed by Equation 3.8.

Then for each $C_s^l\{j\}\{t\}$, we can derive a set of W_i. The corresponding $C_s^h\{j\}\{t\}$ can be derived by:

$$C_s^h\{j\}\{t\} = \sum_{i=1}^{K2} W_i C_{r_{(i)}}^h\{j\}\{t\} \tag{3.8}$$

We now have the complete coefficients $C_s^h\{j\}\{t\}$ of s^h, By using the Inverse Discrete Curvelet Transformation (IDCT) in $C_s^h\{j\}\{t\}$, we can obtain the high frequency feature s^h. Finally we can derive:

$$y_f = y + s^h \tag{3.9}$$

It should be remembered that in the residue step, we only used the K_2 selected images for high frequency learning.

3.3 Experimental Results

In this chapter, we use FERET face database (Phillips *et al.*, 2000) and CASPEAL database (Gao *et al.*, 2008) to test our approach. The FERET database includes 839 individuals and each individual has 2 to 10 images. We choose 239 people as testing samples and the other 600 as training samples. CASPEAL database has

1040 individuals and we choose 440 for testing and the remain 600 for training. In the experiment only one frontal image is collected for each person, including various illuminations, races and genders. Before the experiment, we first align all the images manually and crop the faces by fixing the centers of the eyes and the mouths. The size of the cropped high-resolution image is set as 128×96 and that of the low-resolution image is set as 32×24.

3.3.1 Hallucinating Faces in Curvelet

There are two steps in our experiment. Before each step, there is a pre-selection process, which is designed to locate a set of proper training samples for a testing image to achieve a better performance compared with randomly selected training samples in the hallucinating process. Then in the first step, we try to construct a smooth, high-resolution image with global features. In other words, we want to construct a high-resolution face image with low frequency information based on the coarse coefficients in the Curvelet domain. Since we construct a high-resolution global face with low frequency information, we need to find the local features of each individual. These local features are also called residues, located in the high frequency domain. For this reason, we aim to find the high frequency information through the fine coefficients in Curvelet domain in the second step.

In summary we perform four types of experiments in this chapter.
Experiment 1: By randomly selecting K_1 training samples, we perform sparse representation super-resolution algorithm (Equation. 3.2).
Experiment 2: By randomly selecting K_1 training samples, we first perform sparse representation super-resolution algorithm (Equation. 3.2), then perform Curvelet residual compensation (Equation. 3.7) based on those K_1 training samples.
Experiment 3: We first adopt the proposed Curvelet based pre-selection algorithm to

select K_1 training samples, and then perform sparse representation super-resolution algorithm (Equation. 3.2).
Experiment 4: We adopt our Curvelet based pre-selection algorithm to select K_1 and K_2 training samples, then perform sparse representation super-resolution algorithm (Equation. 3.2). And finally obtain the final results by combining the proposed Curvelet residual compensation (Equation. 3.7).

Figure 3.5 illustrates the hallucination results of our method. Specifically, columns a and b show the original low-resolution and high-resolution images, respectively. Columns c indicates the global face without the pre-classification step, while column d shows the global face with the pre-classification step. Column e illustrates the final result of our three-step approach. It can be seen from Fig.3.5 that the images in column d have better quality than those in column c.

Comparisons in terms of the Peak Signal Noise Ratio (PSNR) and the Root Mean Square Error (RMSE) for our four experiments with other approaches can be found in Table 3.2 and Fig. 3.6, where $K_1 = 30$, $K_2 = 20$, and (e),(f),(g) and (h) are proposed approaches. In Table 3.2, we show the PSNR values of randomly picked 6 training samples and the average results of the whole 679 testing samples in each column. It can be seen from Table 3.2 that when combined with sparse representation technique separately, both our Curvelet Residual compensation approach (Experiment 2) and pre-selection approach(Experiment 3) can improve the results of generic sparse representation method (Experiment 1). A more significant improvement can be seen in our final result (Experiment 4), where we adopt both proposed pre-selection approach and Curvelet residual compensation approach. Similarly, Fig. 3.6 demonstrates the average RMSE values of the 679 testing samples, where both our Curvelet Residual compensation approach and pre-selection approach can reduce the errors, no matter whether they are used separately (Experiment 2 and 3) or used together (Experiment 4).

(a) (b) (c) (d) (e)

Figure 3.5: Our approach. (a) Original low-resolution images. (b) Original high-resolution images. (c) Hallucinated global faces without pre-classification. (d) Hallucinated global faces with pre-classification. (e) Our final output.

Table 3.2: PSNR Comparison of six randomly selected images and the average values of 679 Testing Samples. (a) Sparse Representation approach Yang *et al.* (2010). (b) LPH super-resolution and neighbor reconstruction Zhuang *et al.* (2007). (c) Eigen-Transformation hallucination Wang and Tang (2005). (d) A two-step face hallucination Liu *et al.* (2007). (e) Experiment 1. (f) Sparse Representation combined with our Curvelet residual compensation approach (Experiment 2). (g) Sparse Representation combined with our pre-selection approach (Experiment 3). (h) Our final approach (Experiment 4). Average demonstrates the average PSNR results of all the 679 testing images for each approach.

Images	1	2	3	4	5	6	Average
(a)	28.69	29.02	30.68	27.75	29.26	31.59	29.23
(b)	24.72	22.47	26.47	21.23	24.32	25.54	24.16
(c)	21.44	18.35	23.23	15.81	24.79	20.24	21.92
(d)	25.79	27.06	28.36	24.45	27.85	29.54	26.96
(e)	28.01	28.06	29.36	27.45	28.85	31.54	28.96
(f)	30.89	30.47	31.55	29.74	30.01	32.27	30.03
(g)	30.85	30.64	31.49	29.90	29.48	32.02	29.97
(h)	31.75	31.40	32.22	30.49	31.24	32.41	30.63

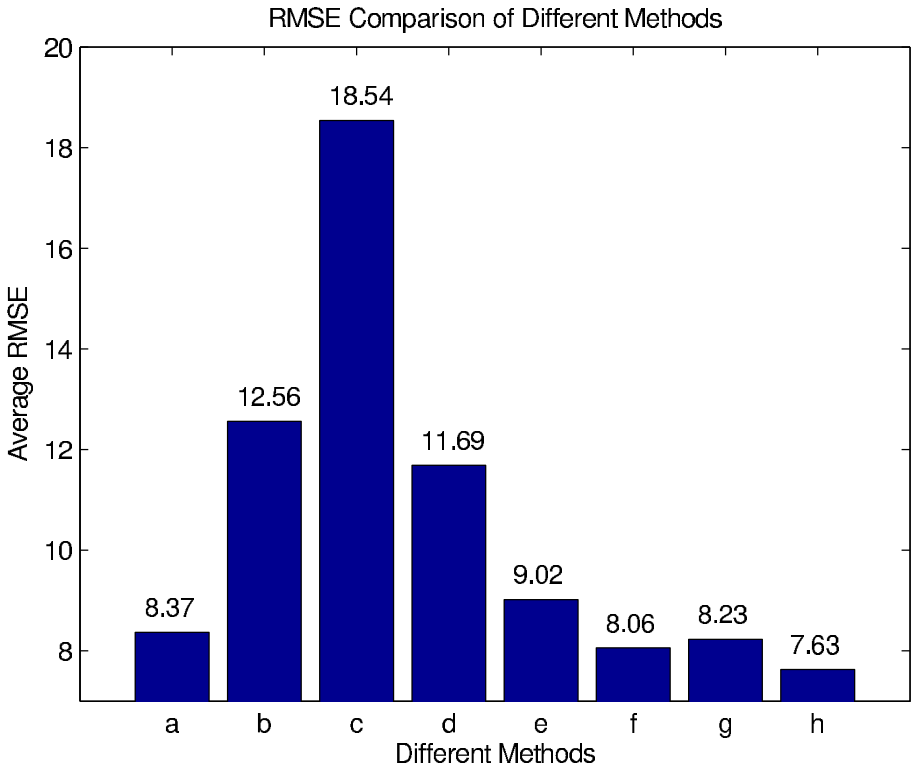


Figure 3.6: Average RMSE Comparison of 679 Testing Samples. (a) Sparse Representation approach Yang *et al.* (2010). (b) LPH super-resolution and neighbor reconstruction Zhuang *et al.* (2007). (c) Eigen-Transformation hallucination Wang and Tang (2005). (d) A two-step face hallucination Liu *et al.* (2007). (e) Experiment 1. (f) Sparse Representation combined with our Curvelet residual compensation approach (Experiment 2). (g) Sparse Representation combined with our pre-selection approach (Experiment 3). (h) Our final approach (Experiment 4).

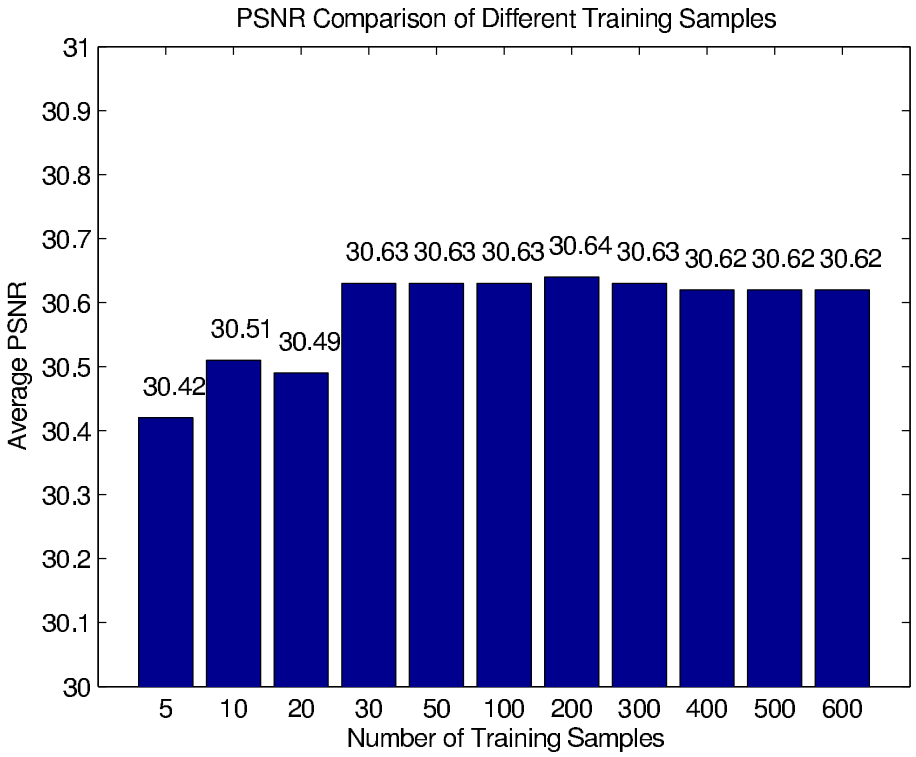


Figure 3.7: Our PSNR results in terms of different K_1.

In order to clarify the influence of training samples, we perform our experiment when K_1 is set as 5, 10, 20, 30, 50, 100, 200, 300, 400, 500 and 600 respectively in Experiment 4 ($K_2 = 20$). Figure 3.7 describes the average PSNR values in term of different training samples(K_1) in Experiment 4. It can be seen that our approach does not depend too much on the number of selected training samples. It performs quite well even the number of training samples is small.

By combining the pre-selection approach and Curvelet residual compensation, our approach utilizes the advantages in both spatial domain and frequency domain. Figure 3.8 indicates the comparison between our approach ($K_1 = 30, K_2 = 20$) and other four typical methods (Wang and Tang, 2005; Liu *et al.*, 2007; Yang *et al.*, 2010; Zhuang *et al.*, 2007) when the number of training data is 200. It can be identified that our results have much smoother and clearer views even with a smaller training data. Especially when compared with those approaches with holistic face as training

samples (Wang and Tang, 2005; Liu *et al.*, 2007; Zhuang *et al.*, 2007), the proposed method has less noises around the chin area.

The comparison of different methods in terms of the Peak Signal Noise Ratio (PSNR) is also shown in Table 3.2, which includes six randomly selected people in both two databases. And the last column shows the average results of the whole experiment (239 people in FERET and 440 in CASPEAL). Face hallucination results of Yang *et al.* (2010), Zhuang *et al.* (2007), Wang and Tang (2005) and Liu *et al.* (2007) are shown in $(a), (b), (c)$ and (d) respectively. Figure 3.6 describes the average Root Mean Square Errors of the above methods as well. From these comparisons, one can see that our approach outperforms other existing approaches.

3.3.2 Holistic Model vs Patch based Model

In previous chapter, we propose a holistic model to enhance the resolution of faces, while in this chapter we propose a patch-based model instead. When compare both these two types of models, we can conclude from Figure 1.2 that the holistic based hallucination models have better performance in facial details. When comparing both these models in PSNR and RMSE values, we can see from Table 3.2 that, the holistic models (Liu *et al.*, 2007),(Wang and Tang, 2005),(Zhuang *et al.*, 2007) have smaller PSNR values (less than 27), while the patch based model Yang *et al.* (2010) has higher PSNR performance (more than 30). This means that the patch-based models performs better in terms of PSNR evaluation method. In Figure 3.3, we specialize the hallucination algorithm to be the same (Sparse Representation Super-resolution (Yang *et al.*, 2010)) in both holistic and patch based models, which reduce the impact caused by different hallucination methods. All the experiment settings are exactly the same, such as training data, sparse parameters, testing data, and etc. As displayed in Figure 3.3, the hallucinated faces in terms of holistic and

(a) (b) (c) (d) (e) (f) (g)

Figure 3.8: Comparison with other methods. (a) Original low-resolution face. (b) Original high-resolution face. (c) Hallucinated faces by Yang *et al.* (2010). (d) Hallucinated faces by Zhuang *et al.* (2007). (e) Hallucinated faces by Wang and Tang (2005). (f) Hallucinated faces by Liu *et al.* (2007). (g) Hallucinated faces by our approach.

patch-based models with sparse sensing theory have the similar results compared with Figure 1.2. In this thesis, we propose a holistic based model in Chapter 2 and a patch based model in Chapter 3. The comparison of visual quality between the proposed approaches can be seen from Figure 2.11 and Figure 3.5. The common features in terms of holistic-based and patch-based models could be concluded as follows:

1. The detailed facial features, such as eyes, eyebrows, mouths, and etc., of hallucinated faces based on holistic models look better than patch-based models. However, the visual comparison shows the edges of hallucinated faces, especially around chin area, of holistic models are quite noisy. The patch-based models could display smoother faces, though some of the facial features are lost during the hallucinating process.

2. In terms of PSNR and RMSE, the patch-based models performs better than holistic models, as they produce much smoother results.

3.4 Summary

In this chapter, a Curvelet feature based face hallucination approach is proposed via sparse sensing technique. We first select the training samples according to the Curvelet coefficients of the low-resolution testing image. Secondly, we use the general sparse representation idea to reconstruct the global face based on the selected training samples. Compared with the general sparse representation method, this pre-selected training samples can help improve the hallucination results. Residue compensation is then carried out. Since the residues can be thought of as the high frequency information of the face images, we select the residue training samples by

locating the nearest neighbors of the high frequency coefficients in Curvelet domain. Through the learning process, the Curvelet coefficients of the high-resolution residue images are estimated. The high-resolution residue image can be derived by employing the Inverse Discrete Curvelet Transformation. Finally, by combining the global face with the residue, the final high-resolution face can be derived.

Most of the face hallucination methods only adopt Peak Signal Noise Ratio (PSNR) or Root Mean Square Error (RMSE) as evaluation method. However, in some occasions, PSRN and RMSE can not represent hallucination quality properly. Since our aim of face enhancement is to improve face recognition performance, recognition performance should be adopted to evaluate hallucination quality. How much face hallucination can improve face recognition in different situation? This will be discussed in the next chapter.

Chapter 4

Face Hallucination for Recognition Performance Improvement

4.1 Introduction

4.1.1 Face Hallucination Overview and Related Work

As described in Chapter 1, most of the face recognition approaches are performed in the same resolution. When testing images and gallery images are in different resolutions, the simplest way is to resize the testing and/or gallery images to have the same resolution. As the gallery faces are generally in high resolutions while the testing faces vary from low to high resolutions, the easiest way is to down-sample the gallery faces to match the testing size or choose a low resolution and down-sample both gallery and testing images to this chosen size. However, this will have a compromise on face recognition performance.

Another way of matching resolution is to change the resolution of testing images and make them to have the same resolution as gallery images. Thus super-resolution techniques are required for face recognition, which is called face hallucination. Face hallucination adds additional pixels to increase the resolution of face images in order to improve the performance of face recognition systems for very low-resolution

testing face images.

As to face hallucination, Baker and Kanade (2000, 2002) was the first to introduce the face hallucination theory. Based on such techniques, they proposed a learning based algorithm, which learns the prior on the spatial distribution of the face image gradient and yields high-resolution face images. Numerous face hallucination approaches have been proposed ever since, such as Wang and Tang (2005); Liu *et al.* (2007); Zhuang *et al.* (2007); Yang *et al.* (2008a); Zhang and Cham (2011); Chang *et al.* (2004); Yang *et al.* (2010); Ma *et al.* (2010); Chakrabarti *et al.* (2007). As discussed previously, Theses hallucinating approaches can be divided into two types according to the way they dealing with testing faces. One is holistic based face hallucination and the other is patch based face hallucination. For example, Wang and Tang (2005) proposed an efficient hallucinating algorithm. They took the low-resolution and high-resolution faces into two groups, and tried to find the linear relation between those two groups. They derived this through Principal Component Analysis (PCA), where both the low and high resolution images are projected into their eigen-subspaces, and the linear transformation could be interpreted in these two subspaces. A statistical modeling approach was proposed by Liu *et al.* (2007), who solved the enhancement problem through two steps. The global features and local features were separately derived from a global parametric model and a local non-parametric model. A hallucinated human face can be derived by combining the global and local features. Based on their work, Zhuang *et al.* (2007) adopted Locality Preserve Projection (LPP) and Radial Basis Function (RBF) to produce the global faces and simplify the non-parametric model to generate the local features. Yang *et al.* (2008a) adopted the Non-negative Matrix Factorization (NMF) algorithm to generate global faces and found the local residues through sparse representation method. All these above approaches are holistic models. As for patch based face hallucination approaches. Zhang and Cham (2011) proposed an approach in frequency domain. They transformed faces through the Discrete Cosine

Transformation (DCT), and estimated the high-resolution DC components and AC components separately. Through the inverse DCT, hallucinated faces can be constructed. Other super-resolution approaches for generic images also can be used in face hallucination. Chang *et al.* (2004) proposed a super-resolution approach based on the Locality Linear Embedding. Yang *et al.* (2010) proposed a Sparse Representation super-resolution approach. All of these patch based models can generate smooth hallucinated faces efficiently and with good performances in terms of Peak Signal Noise Ratio (PSNR) and Root Mean Square Error (RMSE) performance.

4.1.2 Research Gap

One original motivation of face hallucination is to improve face recognition performance. However, many of the previous hallucinating approaches only used the PSNR or RMSE values to evaluate the hallucinating quality. The proposed approaches claim to have good hallucination results in terms of high PSNR or low RMSE values. However, the did not validate their performance in terms of face recognition. Also, they did not prove the performances of their algorithms on images with extremely low resolutions, where face hallucination is actually most needed for face recognition. In this thesis we define the extremely low-resolution faces as facial image whose resolutions are equal or lower than 16×16.

4.1.3 Contributions

In this chapter, we aim to fill the above gap by studying the relationships among image resolution, recognition performance and hallucination performance. We have three contributions here.

Firstly, through extensive experiments we prove that in case of extremely very low-resolution, the effectiveness of hallucinating on improving face recognition is actually debatable.
Secondly, we reveal that the recognition performance can be improved if the image resolution is sufficiently large. When the resolution is very low, hallucinating faces can not actually enhance the recognition performance.
Finally, by studying the relationship between the recognition rate with PSNR and RMSE values, we found that these two commonly used evaluation metrics for hallucination algorithms are not able to accurately reflect how much hallucination could assist in face recognition.

The remaining part of this chapter is organized as follows. In Section 4.2, we will investigate the relationship of image resolution and recognition performance. In Section 4.3, the face hallucination and recognition are investigated. The Hallucination metrics are discussed in terms of recognition performance in Section 4.4. The summary of this chapter is presented in Section 4.5.

4.2 Relationship between Resolution and Recognition

Since one of the the most important purposes of face hallucination is to assist performance in face recognition, it is an important issue to determine what kind of face images cannot be recognized by human perceptions and machine perceptions. For human perceptions, faces are able to be recognized in reasonably high-resolution images as there are more details available in those images. When the image resolution decreases to a certain threshold, faces are difficult to be recognized by human. Figure 4.1 demonstrates a set of faces in YaleB database (Georghiades *et al.*, 2001;

(a) 128×128 (b) 64×64 (c) 32×32 (d) 16×16 (e) 8×8

Figure 4.1: Face Image Display in terms of Different Resolutions.

Lee *et al.*, 2005) from high-resolution 128×128 to very low-resolution 8×8. We can see that the faces become hardly recognizable when the resolution is below 32×32.

With machine recognition, things are surprisingly different. Experiments are conducted on the extended YaleB database (Georghiades *et al.*, 2001; Lee *et al.*, 2005), where face images are down-sampled from high-resolution (128×128) to a set of low-resolution images: 64×64, 32×32, 16×16 and 8×8. Face recognitions are then performed on images in different resolutions. In all the recognition experiments, half of the images in each class are randomly selected as training data and the remaining half as testing data. Recognition experiments are repeated twenty times for each class and the recognition rates are taken from the average values.

If the resolutions are set as the variable, it is found that when the resolutions vary from 8×8 to 128×128, the recognition rates have different trends with different recognition approaches. Figure 4.2 shows the results of our experiments in which we recognize faces using different recognition approaches in the Extended YaleB database. We implemented Principal Component Analysis (PCA) Turk and Pentland (1991), Linear Discriminant Analysis (LDA) Belhumeur *et al.* (1997), Locality

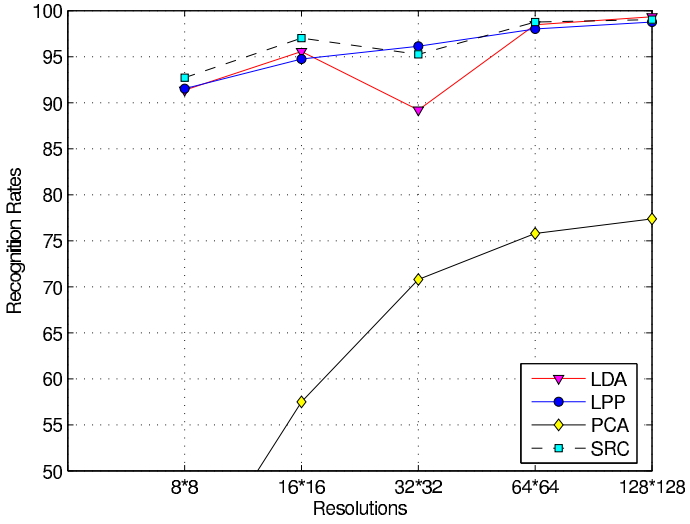


Figure 4.2: Recognition Rates in terms of Different Recognition Algorithms and Resolutions in Extended YaleB Face Database

Preserving Projections (LPP) He and Niyogi (2004) and Sparse Representation (SRC) Wright *et al.* (2009) which are the most popular face recognition algorithms for images with resolutions of 8×8, 16×16, 32×32, 64×64 and 128×128. Recognition by the PCA algorithm has an obvious trend, where the recognition rate increases when face resolution increases. However, the overall recognition rates by PCA is not good enough, ranging from 37.54 to 77.37. LDA, LPP and SRC produce satisfactory recognition rates. However, the recognition performances do not strictly follow the resolution sizes. It can be seen that in the low resolution, SRC has a very good performance. Specifically in the resolution of 16×16, SRC produces a similar recognition rate as in the highest resolution of 128×128. In the resolution of 32×32, recognition rates drop for both the SRC and LDA methods.

The same experiment is also performed in the AR database (Martinez and Benavente, 1998) in Figure 4.3, where the trend of recognition rate varies but still does not increase along with resolution.

From both these experiments, it has been shown that though it is not always the

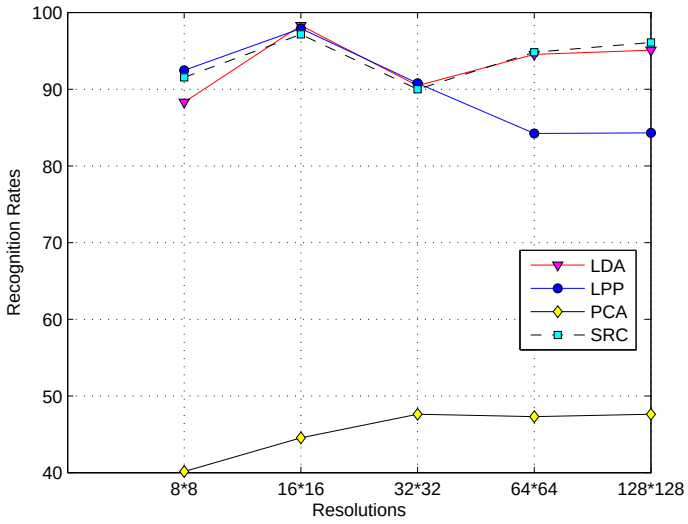


Figure 4.3: Recognition Rates in terms of Different Recognition Algorithms and Resolutions in AR Face Database.

case that higher resolutions leads to the higher recognition rates, the recognition performance is in general becoming better with increasing high resolution. In other words, when the resolution of facial images decreases, face recognition performances do not drop obviously. Instead, they remain in a similar performance range (see Figure 4.2 and Figure 4.3). Resolution is not the only factor that influences the face recognition performances. There might be some other factors that affect the recognition rate, such as recognition algorithms, types of cameras and different face databases which include illumination, poses, expressions, gender and human races.

It is worth to point out that the high performances with different classifiers are possibly due to the face database patterns. Also the low resolution images are obtained only from down sampling technique and some innate features of its high resolution images are inherited. In practice, it is more interesting to see the recognition performance for hallucinated images by using different approaches. We will investigate this issue in next section.

4.3 Face Hallucination and Recognition

The previous section has shown that high resolutions do not guarantee the improvement of face recognition in YaleB database (Georghiades *et al.*, 2001; Lee *et al.*, 2005) and AR database(Martinez and Benavente, 1998). In order to clearly evaluate the performance of face hallucination in recognition context, another experiment is conducted in which we enhance the low-resolution face images through four typical face hallucination approaches: Eigen Transformation (Eigen) (Wang and Tang, 2005), Two Step Face Hallucination Theory (TwoStep) (Liu *et al.*, 2007, 2001), Sparse Representation Super Resolution (ScSR) (Yang *et al.*, 2010) and Cubic Interpolation (Cubic) (Hou and Andrews, 1978). Wang and Tang (2005) and Liu *et al.* (2007, 2001) are examples of holistic models. Yang *et al.* (2010) represents patch based models. Hou and Andrews (1978) is a non-learning based interpolation approach. The image resolutions are enhanced in three types of resolutions: from 8×8 to 32×32, from 16×16 to 64×64 and from 32×32 to 128×128. We take each low-resolution image in the Extended YaleB and AR databases as a test image, and use the FRGC Phillips *et al.* (2006) face database as the training data of the hallucination experiment. We use a third face database (FRGC) as the training data in order to fairly verify the robustness of hallucination approaches.

As a result a set of hallucinated high-resolution faces are then derived. In the recognition experiment, the testing faces are randomly selected from the hallucinated faces and the training data are randomly selected from the original high-resolution images. LDA Belhumeur *et al.* (1997), LPP He and Niyogi (2004), PCA Turk and Pentland (1991) and SRC Wright *et al.* (2009) face recognition algorithms are adopted. Similarly, the recognition experiment is repeated twenty times and the recognition rates are averaged from them.

The recognition rates of the low-resolution faces, hallucinated faces and high-resolution

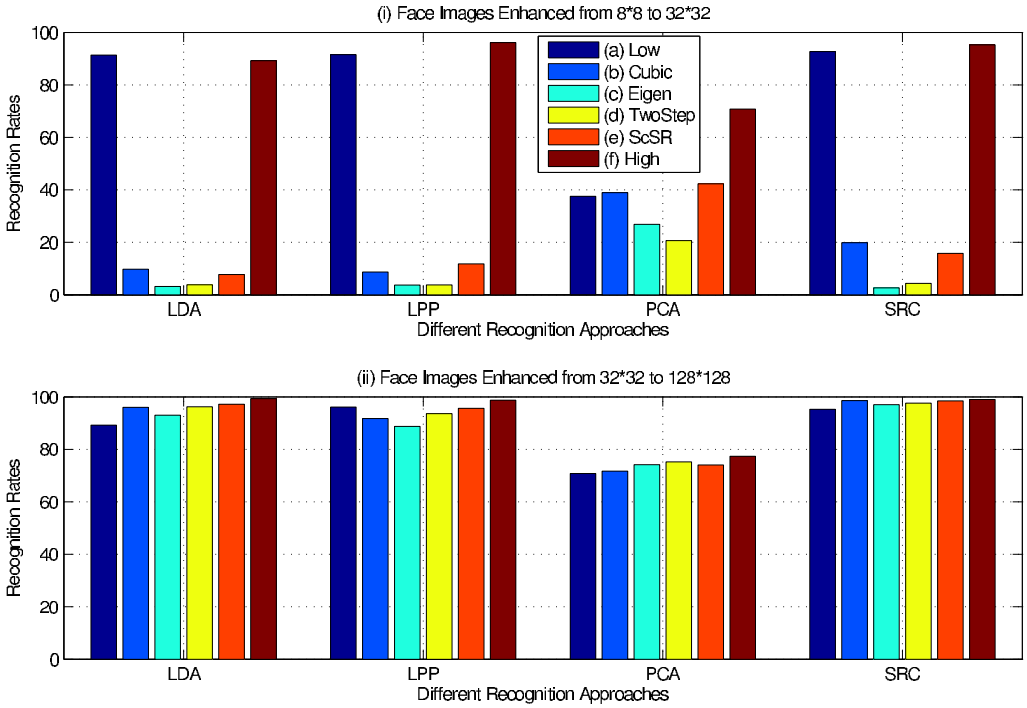


Figure 4.4: Recognition Rates of Hallucinated Faces in YaleB database. (a) Recognition Rate of Low Resolution Faces. From (b) to (e) are Recognition Rates of High Resolution Faces Hallucinated through Cubic Interpolation Hou and Andrews (1978), Eigen-Transformation Wang and Tang (2005), Two Step Hallucinating Theory Liu *et al.* (2007) and Sparse Representation Super-Resolution Yang *et al.* (2010). (f) Recognition Rate of Original High Resolution Faces.

faces are compared by using different face recognition approaches. Figure 4.4 (i) demonstrates the experimental results when the low-resolution is 8×8 and the hallucinated high-resolution is 32. It can be seen that when the image resolutions is extremely very low, hallucinated faces actually do not provide much help to recognition rate. In fact, most hallucinated high-resolution faces have lower recognition rates compared with the low-resolution faces. This gives us a conclusion that when the image resolution is sufficiently low, image hallucination will not help for face recognition. This is very surprising.

However, the situation is totally different when the images resolution is enhanced

from 32×32 to 128×128. Figure 4.4 (ii) shows the experimental result for such cases. For most hallucinated 128×128 images, face recognition rates increase when compared with the original 32×32 faces using all four typical recognition approaches. This means that if the original face is in the resolution of 32×32, the face recognition rates can be improved, sometimes significantly, by enhancing the image resolutions through hallucinating faces, while in the resolution of 8×8, the recognition performance can hardly be enhanced by these super-resolution approaches. Figure 4.5 displays hallucinated facial images in terms of different hallucinating approaches.

Similar results have been obtained by performing the same experiment in the AR database (Martinez and Benavente, 1998). Figure 4.6 shows the results of our experiment in AR database. Figure 4.7 illustrates the visual display of hallucinated faces in AR database. These faces are hallucinated from the resolution of 32×32 to the resolution of 128×128.

4.4 How to Evaluate the Hallucination Results

Most of the existing face hallucination algorithms use the Root Mean Square Error (4.1) and Peak Signal Noise Ratio (4.2) to evaluate the enhancement results. They are defined respectively as:

$$RMSE = \sqrt{\frac{\sum_{1}^{m}\sum_{1}^{n}(I - \hat{I})^2}{m \times n}} \tag{4.1}$$

$$PSNR = 20 \cdot \log_{10} \frac{255}{RMSE} \tag{4.2}$$

where m and n are the numbers of rows and columns of the high-resolution images. I and $\hat{I}$ represent the original high-resolution testing images and hallucinated high-

(a) (b) (c) (d) (e) (f)

Figure 4.5: Examples Display of Faces through Different Hallucinating Methods In YaleB (Georghiades *et al.*, 2001) Database. (a) Original Low-resolution Facial Image (with the resolution of 32×32). (b) High-resolution Facial Image Hallucinated by approach of Wang and Tang (2005) (with the resolution of 128×128). (c) High-resolution Facial Image Hallucinated by approach of Liu *et al.* (2007) (with the resolution of 128×128). (d) High-resolution Facial Image Hallucinated by approach of Cubic Interpolation (with the resolution of 128×128). (e) High-resolution Facial Image Hallucinated by approach of Yang *et al.* (2010) (with the resolution of 128×128). (f) Original High-resolution Facial Image. (with the resolution of 128×128).

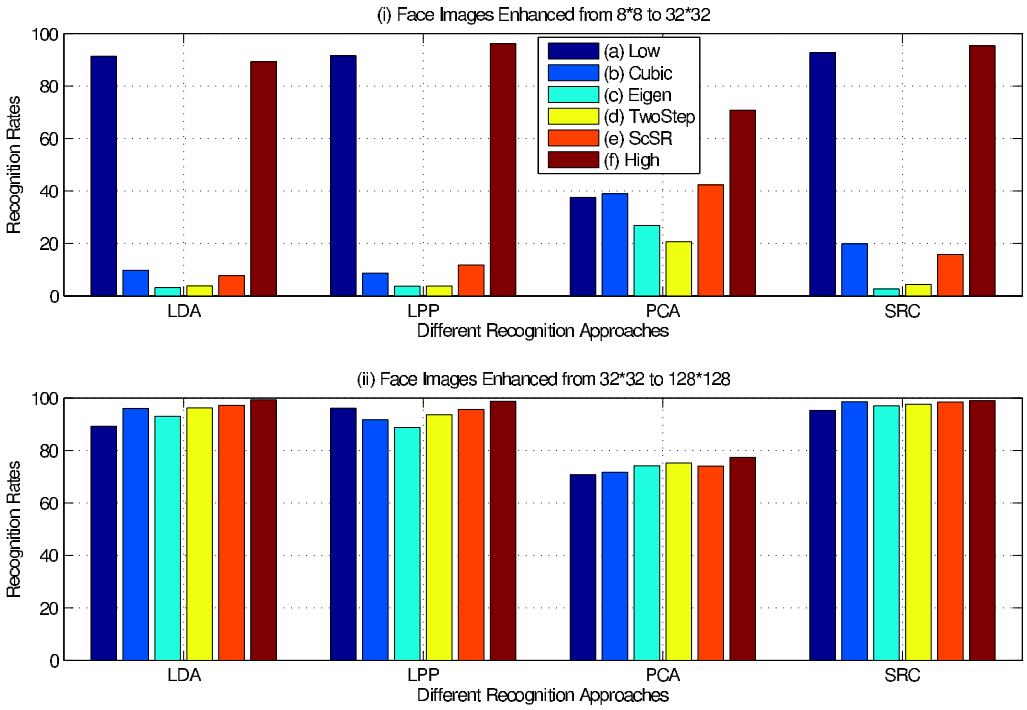


Figure 4.6: Recognition Rates of Hallucinated Faces in AR database. (a) Recognition Rate of Low Resolution Faces. From (b) to (e) are Recognition Rates of High Resolution Faces Hallucinated through Cubic Interpolation Hou and Andrews (1978), Eigen-Transformation Wang and Tang (2005), Two Step Hallucinating Theory Liu *et al.* (2007) and Sparse Representation Super-Resolution Yang *et al.* (2010). (f) Recognition Rate of Original High Resolution Faces.

(a) (b) (c) (d) (e) (f)

Figure 4.7: Examples Display of Faces through Different Hallucinating Methods In AR (Martinez and Benavente, 1998) Database. (a) Original Low-resolution Facial Image (with the resolution of 32×32). (b) High-resolution Facial Image Hallucinated by approach of Wang and Tang (2005) (with the resolution of 128×128). (c) High-resolution Facial Image Hallucinated by approach of Liu *et al.* (2007) (with the resolution of 128×128). (d) High-resolution Facial Image Hallucinated by approach of Cubic Interpolation (with the resolution of 128×128). (e) High-resolution Facial Image Hallucinated by approach of Yang *et al.* (2010) (with the resolution of 128×128). (f) Original High-resolution Facial Image. (with the resolution of 128×128).

resolution images respectively.

However, both of these two parameters only measure the average differences between the original low-resolution images and hallucinated high-resolution images, which cannot count the correctness pixel by pixel. In other words, PSNR and RMSE measurements can not exactly measure the face features.

In next experiment, we aim to verify high PSNR or RMSE values do not necessarily lead to good hallucination metrics in terms of recognition performance. An experiment is conducted to enhance the face resolution from 32×32 to 128×128, and compares the PSNR and RMSE values with recognition rates using Principal Component Analysis (PCA) (Turk and Pentland, 1991), Linear Discriminant Analysis (LDA) (Belhumeur *et al.*, 1997), Locality Preserving Projections (LPP) (He and Niyogi, 2004) and Face Recognition via Sparse Representation (SRC) (Wright *et al.*, 2009).

This experiment is conducted in Extended YaleB Face Database (Georghiades *et al.*, 2001; Lee *et al.*, 2005). And two processes are performed as follows:

Stage 1: Hallucinating Faces. In this stage, an individual face database, Face Recognition Grand Challenge face database (Phillips *et al.*, 2006), is adopted as training data for hallucinating faces. Frontal faces with different illumination conditions in each human subject are selected as testing data. This means each subject contains 64 face images. These $64 * 28 = 1792$ testing images are firstly four times down-sampled to low resolutions (32×32). And then hallucinated to the high resolution (128×128) by four different hallucination algorithms, which are Eigen Transformation (Eigen) (Wang and Tang, 2005), Two Step Face Hallucination Theory (TwoStep) (Liu *et al.*, 2007, 2001), Sparse Representation Super Resolution (ScSR) (Yang *et al.*, 2010) and Cubic Interpolation (Cubic) (Hou and Andrews,

1978).

Stage 2: The hallucinated faces have the same order as the original high-resolution faces. We mix them together and adopt face recognition algorithms (PCA (Turk and Pentland, 1991), LDA (Belhumeur *et al.*, 1997), LPP (He and Niyogi, 2004) and SRC (Wright *et al.*, 2009)). For one human subject (64 images), we first mark the numbers for each face image from 1 to 64. Then we randomly select half of the images (32 images) from hallucinated face images as testing data. And choose the left half from the original high-resolution face images as the training data. The marking numbers of testing data and training data are not overlapped. PSNR and RMSE evaluation is also performed for the hallucinated face images. These values are then compared with face recognition performances.

The comparison is shown in table 4.1, where the best super-resolution result appears when using the ScSR approach in terms of the PSNR and RMSE evaluation. However, in PCA recognition method the best recognition rate appears at the TwoStep approach and in SRC recognition method the best recognition rate appears at the Cubic approach. In PCA recognition method, Cubic interpolation approach performs the worst in recognition rate while it performs better than Eigen-transformation and Two-step approaches in terms of PSNR and RMSE. The experiment clearly shows that neither PSNR nor RMSE could provide a good evaluation for the performance of hallucination algorithms in terms of assisting face recognition.

Similar experiment is conducted to enhance the face resolution from 8×8 to 32×32. And we compare the PSNR and RMSE values with recognition rates in the LDA, LPP, PCA and SRC methods. The comparison is shown in table 4.2, where the best super-resolution result appears at the PCA approach in terms of the PSNR and RMSE evaluation.

If we look the Table 4.1 and Table 4.2 further, we can find the best performance classifiers in two cases are different, SRC in table 4.1 and PCA in table 4.2. With better classifier selected, its performance is increasing with better PSNR values for different hallucinated images. This is coincident with our logical reasoning and motivates us to develop hallucinated techniques with selected classifiers in future. If we look these two tables with Figure 4.4 and Figure 4.6 together, we will find that it is unnecessary to hallucinate face images if the resolution is extremely low (8x8 in our case) as the hallucinated images will not improve the recognition performance in this case as seen in Figure 4.4 and Figure 4.6.

In a more precise study of Table 4.1 and Table 4.2, we can find the best recognition performance appears when facial images are enhanced by Cubic interpolation (Hou and Andrews, 1978) and recognized by SRC method (Wright *et al.*, 2009) in Table 4.1. In this case the recognition performance is almost as good as original high resolution images. However, the best performance in terms of PSNR/RMSE appears when faces are enhanced by ScSR (Yang *et al.*, 2010). Though ScSR (Yang *et al.*, 2010) hallucination approach often has the best recognition performance regardless of the recognition approaches in most of the time, Cubic interpolation (Hou and Andrews, 1978) performs better occasionally in both recognition and PSNR/RMSE evaluation. Regardless of different recognition approaches, the performance of face hallucination in terms of PSNR/RMSE is not consistent with the performance in terms of recognition rates. However, we still could make some conclusions. As we have discussed in previous two chapters, patch-based models usually have better performance in terms of PSNR/RMSE. Patch-based models also have better performance in terms of recognition rates, though they are not always exactly consistent with each other.

Table 4.1: Comparison between Recognition Rates and PSNR/RMSE Values when Hallucinating Faces from 32×32 to 128×128.

	ScSR	TwoStep	Eigen	Cubic
LDA	97.21	96.23	93.03	96.01
LPP	95.63	93.61	88.77	91.75
PCA	74.07	75.22	74.17	71.71
SRC	98.48	97.63	97.02	98.60
$PSNR$	31.52	27.57	22.48	30.96
$RMSE$	7.04	11.15	20.65	7.51

Table 4.2: Comparison between Recognition Rates and PSNR/RMSE Values when Hallucinating Faces from 8×8 to 32×32.

	ScSR	TwoStep	Eigen	Cubic
LDA	7.74	3.82	3.20	9.91
LPP	11.81	3.77	3.73	9.93
PCA	42.30	20.66	26.84	37.98
SRC	12.87	4.34	3.60	19.18
$PSNR$	22.76	13.00	12.01	22.25
$RMSE$	19.40	58.99	64.00	20.61

4.5 Summary

One of the main motivations for face hallucination is to help enhance the face recognition performance by both machine and human perceptions. Through extensive experiments in several public face databases, we found that when the face image resolution is below 32×32, they can hardly be recognized by human perception, but can still be well recognized by machines if such resolution is obtained by down-sampling, through some popular recognition algorithms such as LDA, LPP and SRC. It is also found that resolutions are not the only factor that influences face recognition rate. Higher resolution does not necessarily mean higher recognition rate for all classifiers in general. Many other factors may affect the performance of face recognition.

Four typical face hallucination approaches (Wang and Tang, 2005; Liu *et al.*, 2007; Yang *et al.*, 2010; Hou and Andrews, 1978) are implemented to enhance the low-resolution face images to high-resolution: from 8×8 to 32×32, from 16×16 to 64×64 and from 32×32 to 128×128. Face recognitions are then performed on those hallucinated face images using four popular face recognition algorithms: LDA, LPP, PCA and SRC. From our experiments, when face images are enhanced from 32×32 to 128×128, the hallucinated high-resolution face images can be better recognized than the low-resolution images. However, in extremely low dimension (8×8), some of the face hallucination approaches do not work properly. Recognition rates on the hallucinated faces could be even lower than those on the original low-resolution face images.

PSNR and RMSE are the common evaluation metrics for face hallucination results. We compared PSNR and RMSE values of hallucinated faces with recognition rates. The comparison shows that in some circumstances, PSNR and RMSE values are not able to exactly reflect the hallucinating performance in terms of assisting recognition.

More factors are to be considered to develop a robust and effective face hallucination algorithm in future so that it could effectively help enhance face recognition with a selected classifier. More effective evaluation metrics that can directly connect the hallucination quality and the recognition correctness are needed.

Chapter 5

Face Recognition in Surveillance Scenarios

5.1 Introduction and Related Works

As seen from previous chapters, popular face recognition approaches can achieve very high recognition performance in publicly released databases, e.g. Zhao *et al.* (2003); Turk and Pentland (1991); Belhumeur *et al.* (1997); He and Niyogi (2004); Wright *et al.* (2009), where the resolutions of the captured facial images are usually higher than 100×100. Some can even achieve similar high performances in a very low resolution (Wright *et al.*, 2009), where the resolution of face images can be even less than 10×10. However, most of these works have been conducted on databases where face images are captured in controlled environments with high definition cameras. The so-called "low-resolution" face images are derived from high-resolution faces by down-sampling and/or smoothing methods. In Chapter 2, we have discussed and compared different down-sampling methods. Their influences in face recognition and face hallucination are also investigated. However, these down-sampling methods in Chapter 2 are performed by mathematical techniques. The low-resolution images produced are totaly different from the low-resolution images that are directly captured . When face images are captured directly in a "real" low resolution, the high performances of current face recognition approaches are yet to be proven.

5.1.1 Related Works

Recently, face recognition research in real-life surveillance environments has become popular. Surveillance cameras generally produce images in low resolutions, and face images captured directly by surveillance cameras are usually very small. Besides, images taken by surveillance cameras are generally with noises and corruptions, due to the uncontrolled circumstances and distances. Zou and Yuen (2012) proposed a super-resolution approach to increase the recognition performance for very low-resolution face images. They employ a minimum mean square error estimator to learn the relationship between low and high resolution training pairs. A further discriminative constraint is put on the learning approach using the class label information. Biswas *et al.* (2012) proposed a matching algorithm through using Multidimensional Scaling (MDS). In their approach both the low and high resolution training pairs are projected into a kernel space. Transformation relationship is then learned in the kernel space by using iterative majorization algorithm, which is used to match the low-resolution test faces to the high-resolution gallery faces. Similarly, Ren *et al.* (2012) proposed the Coupled Kernel Embedding approach, where they map the low and high resolution face images into different kernel spaces and then transform them to a learned subspace for recognition.

5.1.2 Research Gap

Only a small portion of existing researches are specifically for real surveillance scenarios, where the captured face images are quite different compared with images captured under controlled circumstances with high-definition cameras. Most of the existing works are based on the down-sampled low-resolution face images captured by high definition cameras under controlled environments. Even in those works under surveillance cameras (Zou and Yuen, 2012; Biswas *et al.*, 2012; Ren *et al.*, 2012),

the claimed low-resolution (e.g. 16×16) surveillance face images are in fact down-sampled from the original images captured in the resolution of 64×64 (Grgic *et al.*, 2011). Face recognition performance based on low-resolution (lower than 32×32) face images in uncontrolled surveillance scenarios remains an issue to be explored.

5.1.3 Our Contributions

In this chapter, we systematically analyze the key issues for face recognition in surveillance scenario, where the captured face images are usually with uncontrolled illumination, motion, poses and are generally taken in a far distance. Moreover, the off-the-shelf commercial surveillance cameras come with low-quality sensors and can only capture images in low resolutions.

Through our analysis, we found out that three factors impact significantly on face recognition performances, including the distance between the camera and the human subject, types of cameras including sensor sizes and quality, and the resolutions of captured face images. Three experiments are designed to show the impact of these factors. We first demonstrate that the recognition performances on the low-resolution face images directly captured in real surveillance circumstances are much lower than those on the down-sampled low-resolution images from high-resolution images. This clearly indicates that the down-sampled face images are not able to represent the true low-resolution images. By changing the types of cameras and the values of distances and resolutions, we demonstrate that face image resolution plays a key role in face recognition although the types of cameras and capturing distances are also important factors.

Based on these observations, we propose an approach for face recognition in real surveillance environment. In this chapter we focus on the indoor surveillance envi-

ronment, e.g., in a corridor where people's motions are generally walking in a single direction in a relatively slow and steady pace. Our focus is hence on face recognition based on directly captured face images from surveillance cameras with low resolutions, varied illumination conditions, small pose variation, and slow motions. Due to the very low resolution of the captured face images, many face features might be lost. Image pre-processing ideas are employed to remove illumination variations as much as possible. In order to accumulate more features, we fuse a video sequence into one frame in the frequency domain. Curvelet features are adopted in the fusion process. The fused image is further improved through image super-resolution methods in order to increase the image resolution. Experimental results demonstrate that the proposed system is able to improve the face recognition performance significantly.

5.1.4 Chapter Structure

The remaining parts of this chapter are organized as follows: the proposed approach is illustrated in Section 5.2 and Section 5.3 followed by description of our experiments in Section 5.4. The conclusion is shown in Section 5.5.

5.2 Face Image Pre-precessing

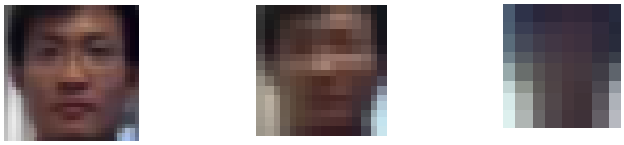
5.2.1 Histogram Equalization for Illumination

In real surveillance scenarios, directly captured low resolution images are different from those which are captured in controlled circumstances. Various factors influence the performance of face recognition, such as motion blur, illumination and noises

in images. In this paper we will focus on the surveillance of an ordinary indoor environment, where a normal range of illumination condition and distortion are considered without motion blurring.

With surveillance cameras, video pictures are usually captured in low resolutions. The generic commercial surveillance cameras record pictures with resolutions varying from 400 to 800 pixels. For example the "$SWANNDVR4-1300$" commercial surveillance system used in CurtinFaces database (Li *et al.*, 2013) captures video sequences with the resolution of 576×704. While working in the indoor circumstance, the camera system captures very small faces in a distance. In the "$SWANNDVR4-1300$" commercial surveillance system, face resolutions are around 32×32 in the distance of approximately 2.5 metres, 16×16 in the distance of 5 metres and 8×8 in the distance of 10 metres respectively. Aliasing problem is apparent in this surveillance circumstance, especially in the distance beyond 5 metres. Fig 5.1 shows the captured face images in the three distances with resolution of 32×32, 16×16 and 8×8 respectively. It is apparent that in the surveillance system, aliasing exists because of the under-sampling problem according to the Nyquist-Shannon sampling theorem. Anti-aliasing techniques are not adopted currently in order to preserve the original features of the captured faces. Most of the time, the captured motions are slow and regular. There are few motion blur effects on the cropped faces. The rarely appeared blurring images will not be used for face recognition in this chapter.

In an indoor corridor with no obvious side lighting, the face images captured demonstrate quite obvious illumination effects from the natural overhead lightings during a walking motion. A histogram equalization approach is adopted here for reducing illumination variations. There are generally two types of histogram equalization for image pre-processing (Štruc *et al.*, 2009). One is the rank normalization where each pixel of the image is ranked and mapped to a new image between the values of 0



(a) (b) (c)

Figure 5.1: Captured Low Resolution Faces in Surveillance Camera. (a) Captured in the Distance of 2.5 Meters with the Resolution 32×32. (b) Captured in the Distance of 5 Meters with the Resolution 16×16. (c) Captured in the Distance of 10 Meters with the Resolution 8×8.

and 255. Another one is to pre-define a distribution of an image's pixels and re-map the image into the pre-defined model. Due to the similar feature on most part of face images, we adopt the second method in our approach.

In detail, for a 32×32 grey scale face image x, the rank for each pixel is normalized to be $r_{i,j}$ ($r_{i,j} \in [1, 1024]$) and the number of pixels is 1024 and the grey scale image level is 256. A general mapping function for pixel $x_{i,j}$ is defined as:

$$p_{i,j} = \frac{1024 - r_{i,j} + 0.5}{1024} = \int_{x=-\infty}^{t_{i,j}} f(x)dx = F(x) \tag{5.1}$$

where $t_{i,j}$ is the rank of pixel $x_{i,j}$ in the re-mapped space with distribution function $f(x)$ and $F(x)$ is the cumulative distribution function (CDF) for a given distribution $f(x)$.

In order to remove the illumination variation, we assume that the intensity distribution of face images matches the standard normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{5.2}$$

The re-mapped face images can be derived from the inverse cumulative distribution function. For the pixel $x_{i,j}$, the re-mapped rank $t_{i,j}$ is derived from:

$$t_{i,j} = F^{-1}(p_{i,j}) \tag{5.3}$$

where F^{-1} is the inverse cumulative distribution function and

$$F(x) = \int_{-\infty}^{t} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x=-\infty}^{t} e^{-(x-\mu)^2/2\sigma^2} dx \tag{5.4}$$

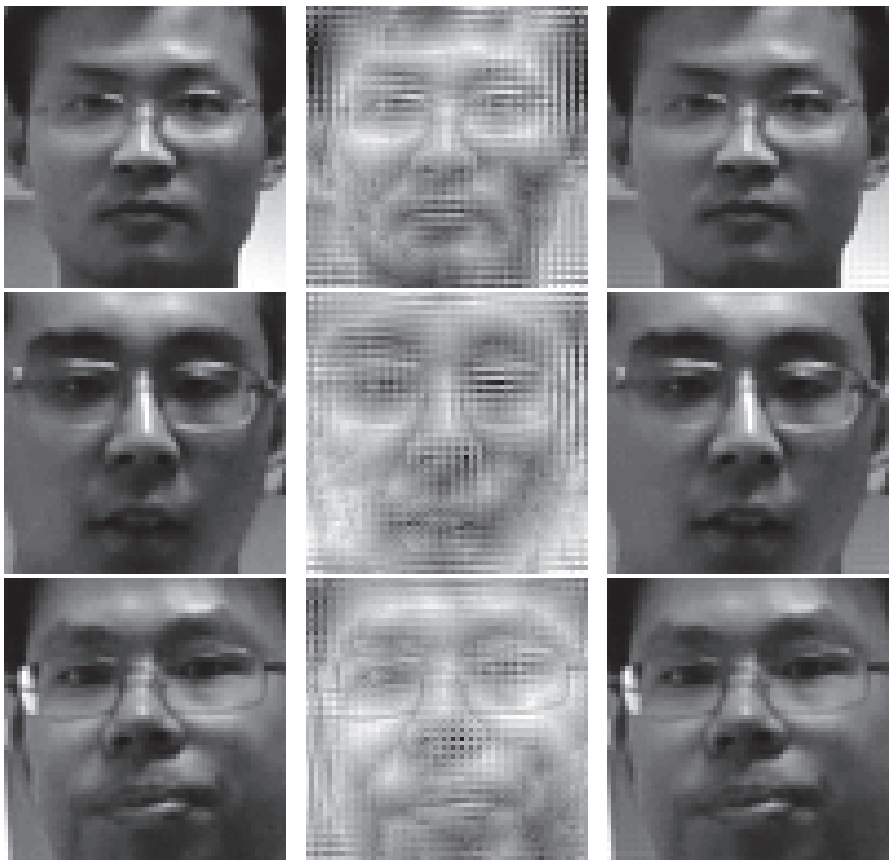
The grey scale face image after histogram equalization is derived through adjusting the pixel rank $t_{i,j}$ to the interval $[0, 255]$.

Figure 5.2 displays the histogram equalization quality for the faces captured by High-Definition cameras in distance. The originally captured face images have the resolution of 64×64. It can be seen that the histogram equalization can remove the illumination effectively.

5.2.2 Fusion of Video Sequence

Surveillance cameras usually capture whatever happens in a given environment into a video sequence. A set of images belong to one person with minor differences in poses and expressions can be extracted from the video. Illumination differences could be minimized after histogram equalization as described in last section. In order to enhance the spectral features for face recognition, image fusion method (Mitchell, 2010) is adopted here. Generally there are two ways for image fusion. One is fusion in the spatial domain and the other is fusion in the frequency domain. In this paper, we utilize the Curvelet coefficients to represent the face features (Candès and Donoho, 1999; Candes *et al.*, 2006). The introduction of Curvelet based algorithms in face feature representation can be found in Section 3.2.1.1 of Chapter 3.

As from Candes *et al.* (2006), Curvelet coefficients can be derived as follows for gray scale 2D face images:



(a) (b) (c)

Figure 5.2: The Results of Histogram Equalization. (a) Original Face Images Captured in Surveillance System. (b) Removed Illuminations. (c) Face Images after Histogram Equalization.

1. Apply the 2D FFT and obtain Fourier samples $\hat{f}[n_1, n_2]$ of $f[t_1, t_2]$, where $0 \leq t_1, t_2 \leq n, -n/2 \leq n_1, n_2 < n/2$.

2. For each scale j and angle l , compute the product $\tilde{U}_{j,l}[n_1, n_2]\hat{f}[n_1, n_2]$, where $\tilde{U}_{j,l}[n_1, n_2]$ is the discrete localizing window.

3. Wrap this product around the origin and obtain $\tilde{f}_{j,l}[n_1, n_2] = W(\tilde{U}_{j,l}\hat{f}[n_1, n_2])$, where W is the wrapping function.

4. Apply the inverse 2D FFT to each $\tilde{f}_{j,l}$ and collect the discrete Curvelet coefficients $C\{j\}\{l\}(k_1, k_2)$.

where j and l represent the scales and angles and k_1, k_2 denote the position of Curvelet coefficient matrix.

As discussed in previous chapter, Curvelet features are good at representing the objects with edges. For example, human faces can be well represented by Curvelet features. For this reason, Curvelet features are adopted for image fusion in this chapter.

The proposed Curvelet based image fusion is represented in Figure 5.3 which indicates that several video frames can be fused into one image in order to derive rich features. It is expected to generate a face image which provides more features for face recognition. From Mandal *et al.* (2009) we can see that fine coefficients represent the character of a human better. For a sequence of facial images, we first transfer them into Curvelet Coefficients. The smallest low-frequency components which are represented by the coarse Curvelet coefficients and the biggest high-frequency components which are represented by the fine Curvelet coefficients are therefore used in the proposed approach.

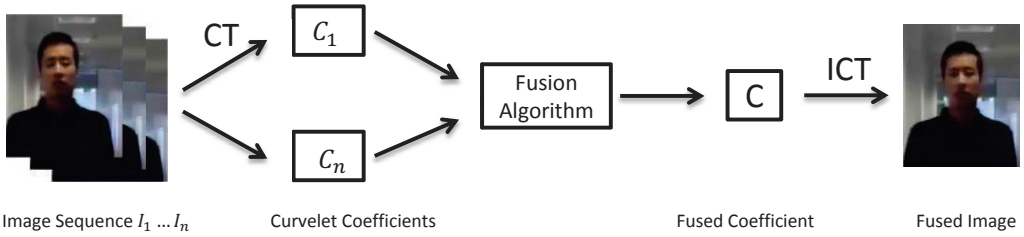


Figure 5.3: Image Fusion Process Diagram.

For the image sequence $I_1, I_2, \cdots, I_n$, their coefficients are represented as $C_i\{j\}\{l\}(k_1, k_2)$ $(i = 1, 2, \cdots, n)$. The components of the first scale where $j = 1$ represent the low-frequency parts of the face image and the components of other scales represent the high frequency parts. The minimum components between each $C_i\{1\}\{l\}(k_1, k_2)(i = 1, 2, \cdots, n)$ and the maximum components among $C_i\{j\}\{l\}(k_1, k_2)(i = 1, 2, \cdots, n; j \neq 1)$ are kept for the fused Curvet coefficients $C\{j\}\{l\}(k_1, k_2)$. After inverse Curvelet transformation, the fused face image can be derived as shown in Figure 5.3.

The process steps of fusing face images are shown as follows:

. Step 1: We first transfer each face image in video sequence $I_1, I_2, \cdots, I_n$ into Curvelet domain. And Curvelet coefficients are derived as $C_i\{j\}\{l\}(k_1, k_2)(i = 1, 2, \cdots, n)$.

Step 2: For each face image I_i, the Curvelet coefficients are a set of matrices. With each j and l, there is a matrix. So for every face image, its Curvelet coefficients contain $j * l$ matrices. If the video sequence has n frames, the total coefficient matrices would be $n * j * l$.

Step 3: When $j = 1$, the coefficients $C_i\{j\}\{l\}(k_1, k_2)(i = 1, 2, \cdots, n)$ show the low frequency features of face images. When $j > 1$, $C_i\{j\}\{l\}(k_1, k_2)$ demonstrate the high frequency features of face images. When faces are captured in surveillance system, the captured face images usually loose high frequency components. This is due to the far distance between objects and cameras and also the poor sensor quality of surveillance cameras. Consequently, the high frequency components face images are desired in surveillance systems. In the fusion process, we keep the high frequency

components of video sequence and suppress the low frequency components. When n face images are fused to one, we adopt the minimum values of low frequency components ($C_i\{j\}\{l\}(k_1, k_2)j = 1$) and the maximum values of high frequency components ($C_i\{j\}\{l\}(k_1, k_2)j > 1$) in each position of Curvelet coefficients. Thus the Curvelet coefficients ($C\{j\}\{l\}(k_1, k_2)$) of fused face image can be obtained.
Step 4: After Inverse Curvelet Decomposition, the fused face image can be derived.

This fusing step is optional because in many cases there is probably only one facial image in the captured image sequences which is suitable for face recognition. In surveillance environment is an uncontrolled system. Lighting, motion, pose, capturing angle, capturing distance and so on affect the quality of captured faces. Some faces with variances in post, illumination, blurring and etc. are not suitable for face recognition. In many situations, only one frontal facial image is available for further processing, thus the fusion step is optional in the proposed aprroach.

5.3 Super-resolution based Face Recognition

As demonstrated in previous work (Wright *et al.*, 2009), even with a very small resolution, some face recognition approaches can still achieve high recognition rates. However, these image are generally obtained in controlled circumstances. Furthermore, most of the very low-resolution face images are derived from down-sampling and smoothing from high-resolution face image database, which are captured by high definition cameras in labs. They are not directly captured low-resolution images. In this chapter, we try to explore face recognition problem with directly captured low-resolution images from surveillance cameras. As shown in Fig 5.1, in real surveillance video sequences, face images taken beyond certain distance always come with noticeable noises and corruptions. When the captured face images are

below 32×32, corruptions are obvious. Directly applying existing face recognition approaches on them generally will not achieve acceptable recognition performances. In order to enhance the face features, we propose a super-resolution based face recognition algorithm.

As discussed previously, face hallucination techniques can be divided into two types. One is the patch based super-resolution, where face images are divided into overlapped patches and each patch is enhanced separately. The final hallucinated image is the combination of the enhanced patches. Such approaches can achieve a smooth high-resolution face image. However, they take face images as generic images and sometimes the enhanced faces are too smooth to preserve the specific human face features. Moreover, in surveillance scenario, the captured low-resolution faces are usually with noises and corruption. The patch based face hallucination would bring these noises and corruptions into the hallucinated high-resolution patches. The other type of approaches is to take the face image as one entity and enhance it directly. Most of such approaches can keep the holistic human face features after enhancing low-resolution faces to high-resolution ones. However, when mapping low-resolution faces into high-resolution, it will produce more noises in high-resolution face image reconstruction if we take the face image as one unit.

In order to make use of the advantages from both the patch based and holistic based hallucination techniques, we propose a method to enhance low-resolution face images by utilizing both of them. Inspired by Yang *et al.* (2010), we make use of the sparsity of signal representation to train low-resolution image patches p_l through a dictionary D_l and transfer the trained relationship α onto the corresponding high-resolution dictionary D_h to reconstruct the high-resolution patch p_h. This dictionary is trained in the FRGC (Phillips *et al.*, 2006) face database independently with both the high-resolution and low-resolution pairs. The high-resolution patch p_h is reconstructed through adopting the same coefficients in the low-resolution training relationship,

where a low-resolution patch p_l is represented by a low-resolution dictionary D_l with the relationship of α. A high-resolution face image can be derived by combining all the high-resolution patches together. The low-resolution sparse representation is formulated as:

$$\begin{aligned} \hat{\alpha} &= argmin \parallel \alpha \parallel_{l_1} \\ s.t. \quad & D_l \alpha = p_l \end{aligned} \tag{5.5}$$

where α is the sparse representation coefficients in l_1 norm.

This sparse representation relationship is mapped to the high dimension space. The high-resolution image patch is derived from:

$$p_h = D_h \hat{\alpha} \tag{5.6}$$

After combing the two high-resolution patches, the hallucinated face image y can be derived.

Meanwhile, we adopt the idea of Wang and Tang (2005) to enhance the same low-resolution face image into a high-resolution one. This process utilizes the Eigen-subspace features of human faces, which has been proved to have a good and stable performance in face feature representation (Turk and Pentland, 1991). For a set of training data (FRGC (Phillips *et al.*, 2006) in this chapter), the covariance of zero mean face images L is: $C = L \times L^T$. A zero mean low-resolution face image x can be represented by the Eigenvectors E as:

$$x = E \times w + m \tag{5.7}$$

where w is the weight of Eigenfaces and m is the mean face.

Equation (5.7) can be rewritten as:

$$x = (L \times V \frac{1}{\sqrt{\Lambda}})w + m = L \times \alpha + m \tag{5.8}$$

where V is the Eigenvectors of covariance matrix $C = L^T \times L$ and $E = L \times V\frac{1}{\sqrt{\Lambda}}$. The high-resolution face y can be derived from:

$$y = H \times \alpha + m_h \tag{5.9}$$

where H is the corresponding high-resolution training data of L and m_h is the high-resolution mean face.

After obtaining two high-resolution faces from the same low-resolution one, a decision is made for each pixel based on the low-resolution face image. For example, for a 16×16 face image, we first enhance it into two high-resolution images using the methods described above. Both these high-resolution face images are then combined into one image with a pixel by pixel decision making. For each pixel $x_{i,j}$ in the low-resolution image, the corresponding pixels in high-resolution is a 4×4 block. For a 16×16 low-resolution face, there are 256 blocks in the high-resolution image. Assume the blocks from the two different enhanced face images are b_1 and b_2 respectively. In order to decide which block is to be kept, we down-sample both the 4×4 blocks into one pixel and keep the one which produces the pixel value closer to the value of the original low-resolution pixel $x_{i,j}$. The final enhanced block image is:

$$\begin{aligned} &\arg\min_{\lambda} \quad Down(b) - x_{i,j} \\ s.t. \quad &b = \lambda \times b_1 + (1 - \lambda) \times b_2 \end{aligned} \tag{5.10}$$

where λ equals to 0 or 1.

After combing the 256 blocks together, the final enhanced face image is obtained which will be used for recognition.

The selection of holistic model and patch based model is optional. The purpose of this holistic and patch combined approach is to adopt the advantages of both holistic and patch based hallucinating model. Thus the hallucinated faces obtained

by proposed approach have smooth appearances and in the meanwhile can keep the facial features.

The proposed approach can be performed as the following steps:
Step 1: Perform histogram equalization on a captured video sequence to remove illumination condition and noises;
Step 2: The image sequence derived after Step 1 is then fused into one image by Curvelet based image fusion;
Step 3: The fused face image is then enhanced to high-resolution by a holistic hallucination model;
Step 4: The fused face image in Step 2 is enhanced to high-resolution by a patch-based hallucination model;
Step 5: The final hallucinated face image is decide by Equation 5.10.

5.4 Experiments and Results

In this chapter, the experiments are performed on four databases: FRGC (Phillips *et al.*, 2006), AR (Martinez and Benavente, 1998), ScFace (Grgic *et al.*, 2011) and CurtinFaces (Li *et al.*, 2013). FRGC and AR databases are captured with high definition cameras. Low-resolution images are down-sampled and smoothened from high-resolution ones. ScFace and CurtinFaces databases contain face images from both high definition cameras and surveillance cameras. All the face images are cropped and aligned before being used. The high definition cameras used in AR, ScFace and CurtinFaces databases are $SONY3CCDs$, $CanonEOS10D$ and $PanasonicLumix$ respectively. The surveillance cameras used in ScFace database are: $BoschLTC0495/51$, $ShannyWTC-8342$, $ShannyMTC-L1438$, $JSJCC-915D$ and $VFD400-12B$. The surveillance camera used in CurtinFaces database

is $SWANNDVR4-1300$.

In real world, there are generally two reasons for a captured face image to be very small. One is that the distance between the camera and the person is too large and the other is that the camera sensor's limitation. Although the focal length of a camera can always be changed, when the distance between a camera and an object is too far away, the captured images become very small. For simplicity, we assume that all cameras in our experiments have fixed focal length.

In our experiments, the resolutions of face images are the originally captured sizes unless specified otherwise. None of the images are down-sampled from high resolution images. For simplicity, we divide face image resolutions into five levels: 128×128, 64×64, 32×32, 16×16 and 8×8. The face images are directly cropped from the surveillance images, and if the cropped images are not exactly the desired sizes, they are slightly changed through Cubic interpolation to the nearest resolution level.

Four experiments are conducted here.
Experiment 1 compares recognition performances between two different types of low resolution image. One type is directly captured with large distance between the camera and the person. Another type is down-sampled from high-resolution images. Results from Experiment 1 demonstrate that the recognition performances for the directly captured images are much lower than the down-sampled low-resolution images.
In Experiment 2 the distance between the camera and the person is fixed. We compare the recognition performances between different types of cameras, resulting in different resolutions in the captured images.
In Experiment 3: the image resolution is fixed. Recognition performances are compared for face images from various sources, whereas the types of camera and cap-

turing distances vary.

The recognition performance of the proposed approach on surveillance face images is demonstrated in Experiment 4.

5.4.1 Down-sampling vs Directly Captured Images

Lots of works have been done on low-resolution face recognition. However, most of the existing works are on low-resolution face images down-sampled from high-resolution images. In real life, most low-resolutions are due to the large distances between the cameras and the face. It is hence worthwhile to evaluate whether the down-sampled images provide a good representation of the true low-resolution images. Here we compare the recognition difference between the down-sampled images and images captured by cameras in a far distance. Face recognition is first performed on images from the popular AR database. Fig 5.4 (a) shows the recognition rates in terms of different down-sampled resolutions on AR database. In this experiment, we randomly select 13 out the 26 images per person for training and the other 13 for testing. This procedure is repeated 10 times to obtain the average recognition rate. Similarly, face recognition results on the CurtinFaces High Definition database are shown in Fig 5.4 (b). Here, only 25 images are selected per person from the available 92 images among which images with large pose and illumination variations are excluded. 12 images out of the 25 are randomly selected for training and the other 13 are for testing.

It can be seen from Figure 5.4 (a) and (b) that when low-resolution face images are down-sampled from high-resolution ones, their recognition rates do not reduced much. Even very low-resolution (8×8) faces can still achieve a satisfactory recognition rate (around 90%).

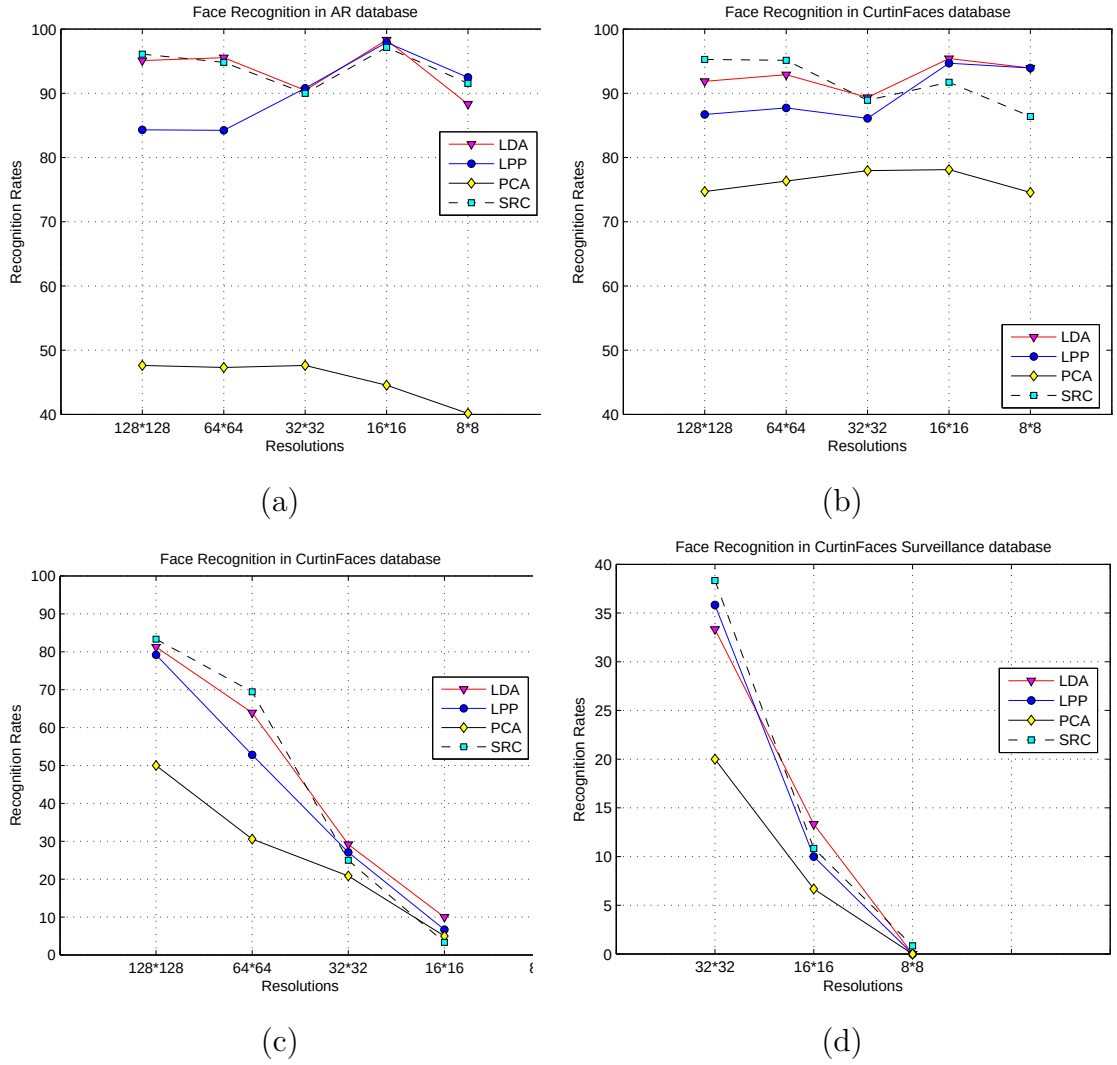


Figure 5.4: Down-sampling vs Distance Sampling. (a) Face Recognition Performance of AR Database when Resolutions are Produced by Down-sampling. (b) Face Recognition Performance of CurtinFaces HD Database when Resolutions are Produced by Down-sampling. (c) Face Recognition Performance of CurtinFaces HD Distance Database when Resolutions are produced by Distances. (d) Face Recognition Performance of CurtinFaces Surveillance Database when Resolutions are produced by Distances.

However, when image resolution drops due to the increased distances, recognition rates decrease very sharply, as shown in Figure 5.4 (c). In this figure, face images are captured using the same High Definition camera as in b. Instead of down-sampling images to low-resolution, images in this figure are captured from various distances in the same environment. The resolutions of the captured face images are approximately in the resolution levels of 128×128, 64×64, 32×32, and 16×16 in the distances of 2.5 meters, 5 meters, 10 meters and 20 meters respectively. On the contrary to the various resolutions from down-sampling, decreasing of resolutions due to the increased distance from camera caused the recognition rates drop very sharply, which is shown in Figure 5.4 (c).

It can be concluded that the down-sampled face images are not good representations of captured low-resolution images for face recognition. Face recognition performance with directly captured images in distances through High Definition cameras is very low when the capturing distances decrease.

To further demonstrate the difference between down-sampling and distance sampling, databases captured through surveillance cameras are adopted. Figure 5.4 (d) and Table 5.1 show the face recognition rates in CurtinFaces Surveillance Camera database and ScFace database respectively. The face images of them are captured with different surveillance cameras in far distances. Figure 5.4 (d) shows the face recognition performance in commercial available surveillance cameras. The capture environment is the same as CurtinFaces HD Distance Database. Due to the quality of surveillance camera, the resolution of captured face images drops to 32×32 with the distance of 2.5 meters, 16×16 with the distance of 5 meters, and 8×8 with the distance of 10 meters. It can be seen from Figure 5.1 that, face images captured in the distance of 10 meters are unable to be recognized with the resolution of 8×8. Further recognition experiment in Figure 5.4 (d) shows the recognition performance in this situation is near 0. Table 5.1 demonstrates the face recognition performance

Table 5.1: Face Recognition Performance in ScFace database.

	LDA	LPP	PCA	SRC
$Camera1$	3.08	4.62	13.85	13.85
$Camera2$	5.38	4.62	18.46	14.62
$Camera3$	3.85	4.62	16.42	10.00
$Camera4$	1.54	7.69	20.77	12.31
$Camera5$	3.08	6.15	12.31	3.08

in ScFace database Grgic *et al.* (2011). It can be seen that regardless of different recognition approaches, the recognition rates are very low when images are captured in far distances instead of down-sampled from high-resolution images.

This experiment shows the obvious difference between directly captured low-resolution facial images and those which are down-sampled from high-resolution ones in terms of recognition performance. As discussed in this subsection, the recognition performance of down-sampled images are much better than those directly captured ones. This is due to the environmental noises between the cameras and faces which are usually in far distances. The down-sampled faces do not have this problem because they are usually derived from mug shots. Mug shots are captured in very shot distances under controlled environment, where there is very few noises.

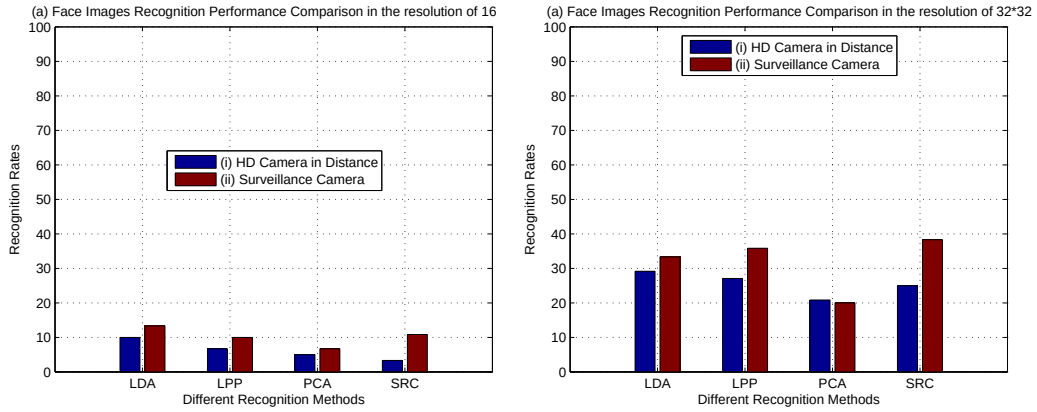
5.4.2 High Definition Camera vs Surveillance Camera

It has been shown in Fig 5.4 (c) that even images captured from a high definition camera are unable to warrant a good recognition performance. In this experiment we evaluate the performance of surveillance camera in an indoor surveillance scenarios. The CurtinFaces Surv database contains video sequences from a surveillance camera which captures human faces in the same environment as the high definition camera

used above. The surveillance camera is a commercial video surveillance camera with the image resolution of 704×576. The original resolutions of the cropped face images from the surveillance camera are approximately 32×32, 16×16 and 8×8 taken in the distances of 2.5 meters, 5 meters and 10 meters respectively. Face recognition performance by the popular LDA, LPP, PCA and SRC methods are shown in Fig 5.4 (d). The recognition rates can be observed to be similar to those of the high definition cameras with different distances (Fig 5.4 (c)). However, when the distances are fixed, e.g., in 5 metres, the differences of cameras and resolutions lead to huge differences in recognition rates. In this distance the SRC recognition rate for high definition camera is around 70% with the resolution 64×64, while the SRC recognition rate for surveillance camera is around only 11% with the resolution 16×16.

5.4.3 Camera, Distance and Resolution

This experiment aims to explore the influences of the types of cameras, distances and resolutions on recognition performances in surveillance system. From Experiment 1, we can see that when the same camera is used, images taken in different distances result in totally different recognition performance. As shown in Experiment 2, when the distance is fixed, images taken by different cameras have large differences in the recognition performance. What would a given resolution lead to? We select two different resolutions in this experiment. Fig 5.5 (a) shows the recognition performance for images with the resolution of 16×16. Images with this resolution is captured by the high definition camera at the distance of 20 meters and by the surveillance camera only from 5 meters away. Fig 5.5 (b) shows the performance in the resolution of 32×32, where HD camera is at a distance of 10 meters and surveillance camera is at a distance of 5 meters. We can see from both figures that despite the differences in camera types and shooting distances, face images with same resolutions result in



(a) (b)

Figure 5.5: Recognition Comparison with The Same Resolution. (a) Comparison of Face Recognition Performance between High Definition Camera and Surveillance Camera in the Same Resolution of 16×16. (b) Comparison of Face Recognition Performance between High Definition Camera and Surveillance Camera in the Same Resolution of 32×32.

similar recognition performances, which is surprising.

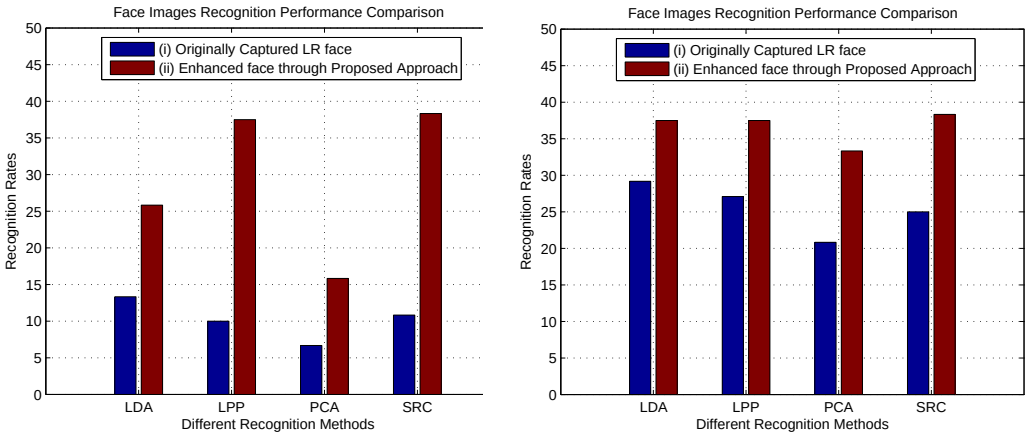
5.4.4 Face Recognition by Super Resolution

In this experiment, the proposed face recognition method is applied and tested. Here, we carry out the experiment on the surveillance camera. Figure 5.6 (a) demonstrates the recognition performance comparison between the captured faces in the distance of 5 meters by the surveillance camera and the enhanced images by the proposed approach. In this setting, the original face resolution is 16×16 and the enhanced face resolution is 64×64. The directly captured face sequence by HD camera have the resolution of 16×16. They are firstly processed through our histogram equalization method and then fused into one face image by the proposed fusion method. The pre-processed face image for each human subject is then hallucinated

by the proposed super-resolution method. Two hallucinated high-resolution face image are derived separately by one holistic hallucination model and one patch-based model. As we have discussed in the proposed algorithm, the two hallucinated high-resolution face images are merged into by proposed algorithm, which balance the advantages and disadvantages of both holist hallucination model and patch-based hallucination model.

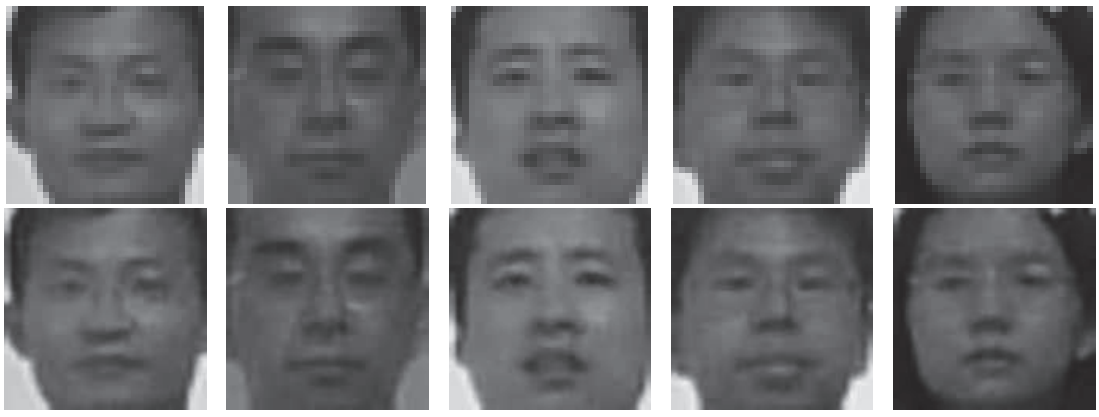
Figure 5.6 (b) shows the recognition performance comparison between the captured faces in the distance of 10 meters by the HD camera and the enhanced faces by the proposed approach. The directly captured human faces with HD camera in surveillance environment have higher resolutions compared with surveillance cameras. The resolution of captured face image in the distance of 10 meters is 32×32 and enhanced to the resolution of 128×128. Similar as Figure 5.6 (a), the captured face images enhanced by the proposed approach can achieve higher performance compared with the directly captured low-resolution face images before hallucination. As shown in Figure 5.6, the face recognition rates are greatly improved after the images are processed using the proposed method, no matter which recognition method is used.

The visual display of Figure 5.6 (b) is shown in Figure 5.7. Here, low-resolution face images (32×32) captured in the surveillance environment are hallucinated to high resolution (128×128) by proposed approach. The first row shows the directly captured low-resolution face images and the second row displays the hallucinated high-resolution face image by proposed approach. We can see that compared with directly captured low-resolution face images, the hallucinated face image by proposed approach have more detailed facial textures and both global and local face features. Thus the improvement of face recognition performance as shown in Figure 5.6 (b) is reasonable.



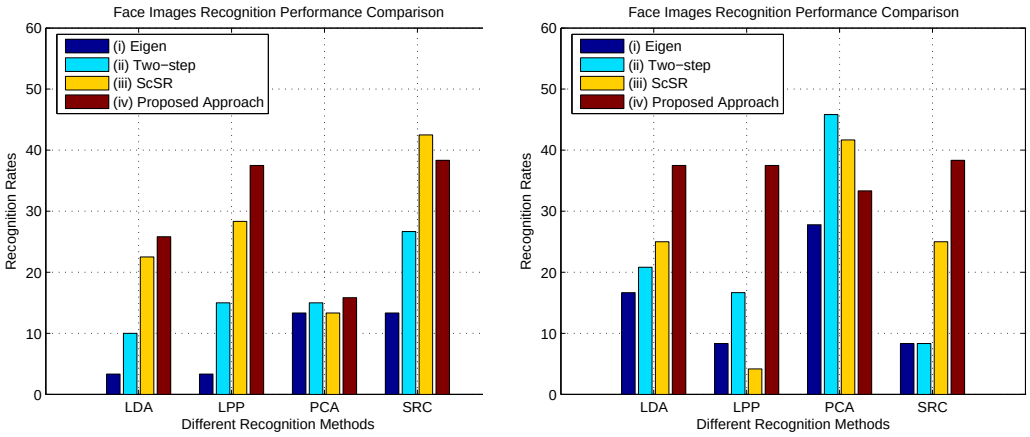
(a) (b)

Figure 5.6: Recognition Performance Comparison between originally captured face images and proposed approach. (a) Originally Captured Low-Resolution (16×16) face images in Surveillance Camera *vs* Enhanced face images (64×64) through Proposed Approach. (b) Originally Captured Low-Resolution (32×32) face images in HD Distance Camera *vs* Enhanced face images (128×128) through Proposed Approach.



(a) (b) (c) (d) (e)

Figure 5.7: Visual Display of Hallucinated Faces of Proposed Approach.



(a) (b)

Figure 5.8: Recognition Performance Comparison in terms of different face hallucination approaches. (a) Face Images (64×64) Enhanced from Originally Captured Low-Resolution (16×16) in Surveillance Camera through Wang and Tang (2005); Liu *et al.* (2007); Yang *et al.* (2010) and Proposed Approach. (b) Face Images (128×128) Enhanced from Originally Captured Low-Resolution (32×32) in HD Distance Camera through Wang and Tang (2005); Liu *et al.* (2007); Yang *et al.* (2010) and Proposed Approach.

We also compare the proposed approaches with other face hallucination methods, including Wang and Tang (2005); Liu *et al.* (2007); Yang *et al.* (2010). Figure 5.8 (a) shows the comparison in Curtin Surveillance Database. Faces are captured with surveillance cameras and the directly captured image resolution is 16×16. After the same image pre-processing proposed in this chapter, those 16×16 facial images are enhanced through different face hallucination approaches to the resolution of 64×64. A similar comparison in Curtin High-Definition camera database is shown in Figure 5.8 (b). It can be seen that the proposed approach performs better in both situations.

5.5 Summary

The experiment in previous sections show that traditional face recognition approaches can hardly achieve satisfactory performance with low-resolution images, especially on those directly captured by surveillance cameras. Till now little work has been done specifically on face recognition based on surveillance cameras. In this chapter, we analyze the factors which impact on face recognition performances in surveillance scenarios. Experiments indicate that other than camera types and capturing distances, image resolution is the major factor affecting the performance of face recognition in surveillance circumstance.

According to the special conditions of a surveillance system, we proposed a super-resolution based face recognition approach. Experiments demonstrate that our approach outperforms traditional face recognition approaches significantly.

Although the proposed approach performs well for very low resolution face recognition in surveillance system, more practical surveillance conditions need to be considered, such as motion blur, extremely low resolution (less than 10×10) and face recognition in outdoor conditions and from very far distances.

Chapter 6

Conclusion and Future Works

6.1 Summary

This thesis has investigated facial image super-resolution techniques. Holistic and patch-based hallucinating models are analyzed respectively. A new holistic face hallucination model is proposed, which adopts the face features in Eigen-subspace to keep and enhance the facial global features. A patch-based model is further proposed, which generates smooth facial images and compensates global facial features meanwhile. After studying face hallucination, we further explore the intrinsic reasons behind it, i.e., face recognition with low resolution facial images. Recognizing faces in low resolution and in hallucinated high resolution are studied. Experiments show that the recognition improvements of hallucinated faces are not so evident in many of the current face databases. In fact, there is a threshold in resolutions. Facial images with the size of 32×32 can have higher recognition performance after hallucination while faces with the size of 8×8 can hardly be improved. Traditional evaluations for hallucinating results are also analyzed and compared with recognition performance. According to the recognition experiments, tradition PSNR and RMSE measurements can not exactly represent the hallucination quality. As a result, we propose to adopt face recognition performance to evaluate the quality of face hallucination instead of PSNR and RMSE. Practical face recognition scenarios often happen in surveillance environments. We further analyze low resolution face recognition in surveillance cameras and experiments demonstrate that resolutions

play a key role in face recognition. Thus for those low resolution images captured in far distance with surveillance cameras, face hallucination can be very useful. In order to make use of the advantages of both holistic and patch based hallucination models, we proposed a new approach which is able to deal with the special situation of low-resolution images captured in surveillance environment. Specifically, our research in this thesis can be divided into the following four aspects:

Firstly, we propose a holistic based face hallucination method. It learns face features from Eigen-subspace and reconstruct the high-resolution faces keeping the global facial features. This reconstructed face image has noises and lacks some of the local face features. A residual compensation is then proposed. It renders local facial features through holistic based learning in Eigen-subspace. This residual compensation is designed to be implemented iteratively, which can render the lost local features as many as possible while keep the global features. Moreover, a two-stage method is proposed. It divides targeted resolution into two or more stages and enhance low-resolution faces stage by stage. This stage based method helps learn the face features in close resolutions, which help reduce noises. Compared with patch-based models, this holistic based face hallucination approach greatly keeps the global features of human faces and reduces noise at the same time. As for its Eigen-subspace learning, the computational cost is also very low.

Secondly, a patch-based face hallucination model is proposed to solve the shortage of holistic hallucination models. Face hallucinated by holistic models usually have good quality in global features while are not good at local features. The hallucinated faces often have noises especially around the chin area. However, patch-based models also have their shortages. They can hallucinate smooth face images by enhancing the resolution patch by patch and these patches are often overlapped. But when learning from patches, the global character of facial images may be lost. We proposed a sample pre-selection approach based on patch based models. The

patch based model is based on sparse representation algorithm. Overlapped patches are enhanced separately through this algorithm. In order to render the lost global face features, two methods are proposed. The first is sample pre-selection algorithm. Due to the loosing of global features in patch-based learning, training samples of this learning is first chosen instead of grabbing all the faces in the database or randomly selecting some faces for training. These training samples are selected through global features of low-resolution testing faces. We adopt Curvelet features to perform this selecting. Only faces who have similar global features with testing face can be selected as patch-based training samples. This pre-selection method helps testing face learn from the training images who have the similar global features. Experiments show the improvement compared with randomly selected training samples. Moreover, this pre-selection method helps reduce the computational cost. Only 30 to 50 training images are required to keep the PSRN/RMSE performances high. As the hallucinated facial images are required to be smooth, they should be similar to the low frequency parts of original faces. We further propose a frequency based residual compensation method. The high frequency parts are learned from Curvelet coefficients and rendered to the hallucinated faces. Through inverse Discrete Curvelet Decomposition, the high-resolution faces can be hallucinated.

Thirdly, we propose to use face recognition performance as a measurement of face hallucination. Experiments demonstrate that in terms of face recognition performance, PSNR/RMSE can not represent the hallucination quality accurately. We further explore the influence face hallucination on face recognition performance. One can find that there is threshold with the resolutions in general face databases. Both holistic models and patch-based models are tested in various popular face recognition approaches in this thesis. The hallucinating scale is four, which means the testing faces with resolution of 8×8 will be enhanced to 32×32 and 32×32 will be enhanced to 128×128. Experiments on Extended YaleB and AR databases show that when the resolution of testing faces is around 8×8, hallucination algorithms

can barely enhance the face recognition performances. However, when the resolution of low-resolution testing images is above 32×32, all the hallucinated approaches can increase the recognition performances more or less.

Fourthly, Most of the previous face hallucination approaches are tested with high-resolution face databases which usually captured under controlled environments. However, few works have been done on surveillance databases. In this part of thesis, we specifically focus on the surveillance environment where face hallucination methods are mostly required. The special scenarios of surveillance cameras are first analyzed. Many factors need to be investigated, e.g., poses, lighting, capturing angle, noise, blurring and etc. Three important factors are proposed for a corridor surveillance scenario: capturing distance, camera sensor size and image resolutions. The latter one is decided by the former two. Experiments are set to analyze these three factors. In our experiment, we first show the difference between controlled environment in laboratory and uncontrolled environment in surveillance system. Even capturing with the same high-definition camera, the recognition performances are quite different. The faces derived from lab are in a very high resolution and low resolution faces are obtained from down-sampling methods while the faces captured from surveillance system are in low resolutions due to the capture distances or sensor sizes. The recognition results show the down-sampled low-resolution faces can have high performances which are similar as the original high-resolution face images. The low-resolution face images captured in far distances have very low recognition performances with the decrease of resolution. This shows the significant difference between down-sampled faces and distance sampled face images. Furthermore, we set one of the three factors to be fixed and the other two could change. Experiments show that different cameras and capturing distances result in differences of resolutions, which lead to obviously different recognition performances. However, regardless of cameras types and capturing distances, facial images with same resolutions have similar recognition performances. This implies that resolutions of face

images are very important in the face recognition of surveillance systems. Hence face enhancement can play a key role in these systems. As such, a face hallucination approach is proposed further. After histogram factorization and image fusion, low-resolution faces captured in surveillance systems are then enhanced to high resolutions. This method chooses the result of hallucinated faces between a holistic model and a patch-based model in pixel level. The method can balance the hallucinating result between these two models. Experiments show the improvement in face recognition performance.

6.2 Future Works

Based what we have achieved in this thesis, we believe the following problems deserve further investigation in the future.

6.2.1 Tradeoff between holistic model and patch based model

In this thesis, holistic and patch-based face hallucination models are analyzed. New approaches are proposed to solve the noise problem in holistic models and over-smooth problem in patch-based models. However, current hallucination approaches have a tradeoff between these two problems. Only one problem can be solved at a time. One hallucinating approach can focus on one side, either holistic or patch-based. Our patch-based model adopts holistic residual compensation in Chapter 3 to render holistic face features after hallucinating with patch-based model. However, the over-smooth problem can be partly solved through this method. The proposed decision maker method in Chapter 5 chooses either holistic or patch-based in pixel

level. But this method can only select pixels between either holistic hallucinated pixels or patch-based hallucinated pixels. There is still not a systematic method which can perfectly solve both noise and over-smooth problem at the same time. We believe this deserves further investigation. For example, faces can be hallucinated by divided them into special patches. These patches are not normally divided patches. Each patch can be part of the face features, e.g., eyes, nose, mouth, cheeks, etc. In future, we will investigate face hallucination based on feature patches.

6.2.2 Surveillance based Face Recognition Database

As we discussed in Chapter 5, practically low-resolution face recognition usually happens in surveillance cameras. However, most of current face databases are produced in controlled laboratory environments and face recognition algorithms are designed for these databases. Grgic *et al.* (2011) provide a good database with faces captured by both a high-definition camera and a set of commercial surveillance cameras. They provide difference poses in high-definition faces and different capturing distances in surveillance cameras. However, lighting conditions are not considered in high-definition faces. And only one face is provided for each surveillance camera in a certain distance. This would be difficult for face recognition algorithms dealing with image set or video sequences. Li *et al.* (2013) also has its limitations, for example, the capturing angles are not the same as surveillance cameras, which should be located in up front and only one commercial surveillance camera is provided. A more comprehensive face database is required, which has plenty of high-definition face image with poses and illuminations, various capturing distances in surveillance systems.

6.2.3 Hybrid Resolution Face Recognition

In order to recognize low-resolution face images, we generally down-sample high-resolution gallery face into low-resolution faces. Then low-resolution testing faces can be matched to these gallery faces in the same resolution. Hallucinating low-resolution testing faces is another way for face recognition. The low-resolution testing faces are firstly hallucinated to high-resolution images. They are then matched to the high-resolution gallery faces. Based on results here, we can usually enhance the resolution of low-resolution testing faces and then match them to gallery faces through face recognition algorithm. However, there is another solution for low-resolution face recognition. A third resolution can be figured out where both low-resolution testing faces and high-resolution gallery faces can be transferred to. This third resolution can be a medium size resolution which both low and high resolutions can reach easily. It can also be special subspace where the relationship between low and high resolution faces can be learned more accurately. Kernel space is another choice where pioneer works have been done by Ren *et al.* (2012) and Biswas *et al.* (2012). However, there is still a lot of research can be done and this would be our future work.

Bibliography

Baker, S. and Kanade, T. (2000). Hallucinating faces. In *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 83–88. IEEE.

Baker, S. and Kanade, T. (2002). Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(9), 1167–1183.

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

Biswas, S., Bowyer, K. W., and Flynn, P. J. (2012). Multidimensional scaling for matching low-resolution face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**(10), 2019–2030.

Bose, N., Kim, H., and Valenzuela, H. (1993). Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 269–272. IEEE.

Candes, E., Demanet, L., Donoho, D., and Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling Simulation*, **5**(3).

Candès, E. J. and Donoho, D. J. (1999). Curvelet: A surprising effective nonadaptive representation for objects with edges. department of statistics. Technical report, Stanford University: Technical Report 1999-28.

Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, **25**(2), 21–30.

Chakrabarti, A., Rajagopalan, A., and Chellappa, R. (2007). Super-resolution of face images using kernel pca-based prior. *IEEE Transactions on Multimedia*, **9**(4), 888–892.

Chang, H., Yeung, D. Y., and Xiong, Y. (2004). Super-resolution through neighbor embedding. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–275. IEEE.

Chantas, G., Galatsanos, N., Likas, A., and Saunders, M. (2008). Variational bayesian image restoration based on a product of-distributions image prior. *IEEE Transactions on Image Processing*, **17**(10), 1795–1805.

Chantas, G. K., Galatsanos, N. P., and Woods, N. A. (2007). Super-resolution based on fast registration and maximum a posteriori reconstruction. *IEEE Transactions on Image Processing*, **16**(7), 1821–1830.

Chen, T. and Defigueiredo, R. (1985). Two-dimensional interpolation by generalized spline filters based on partial differential equation image models. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **33**(3), 631–642.

Donoho, D. (2006a). Compressed sensing. *IEEE Transactions on Information Theory*, **52**(4), 1289–1306.

Donoho, D. L. (2006b). For most large underdetermined systems of linear equations the minimal ??1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, **59**(6), 797–829.

Freeman, W. T., Pasztor, E. C., and Carmichael, O. T. (2000). Learning low-level vision. *International journal of computer vision*, **40**(1), 25–47.

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

Georghiades, A. S., Belhumeur, P. N., and Kriegman, D. J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(6), 643–660.

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

He, X. and Niyogi, P. (2004). Locality preserving projections. *Advances in Neural Information Processing Systems*, **16**, 153–160.

Hertzmann, A., Jacobs, C. E., Oliver, N., Curless, B., and Salesin, D. H. (2001). Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 327–340. ACM.

Hou, H. and Andrews, H. (1978). Cubic splines for image interpolation and digital filtering. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **26**(6), 508–517.

Huang, H., Wu, N., Fan, X., and Qi, C. (2010a). Face image super resolution by linear transformation. In *17th IEEE International Conference on Image Processing*, pages 913–916. IEEE.

Huang, H., He, H., Fan, X., and Zhang, J. (2010b). Super-resolution of human face image using canonical correlation analysis. *Pattern Recognition*, **43**(7), 2532–2543.

Jia, K. and Gong, S. (2008). Generalized face super-resolution. *IEEE Transactions on Image Processing*, **17**(6), 873–886.

Karayiannis, N. B. and Venetsanopoulos, A. N. (1991). Image interpolation based on variational principles. *Signal Processing*, **25**(3), 259–288.

Kim, S., Bose, N., and Valenzuela, H. (1990). Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **38**(6), 1013–1027.

Lee, K.-C., Ho, J., and Kriegman, D. J. (2005). Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(5), 684–698.

Li, B. Y., Mian, A. S., Liu, W., and Krishna, A. (2013). Using kinect for face recognition under varying poses, expressions, illumination and disguise. In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 186–192. IEEE.

Liang, Y., Lai, J.-H., Xie, X., and Liu, W. (2010). Face hallucination under an image decomposition perspective. In *20th International Conference on Pattern Recognition*, pages 2158–2161. IEEE.

Liu, C., Shum, H.-Y., and Zhang, C.-S. (2001). A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–192. IEEE.

Liu, C., Shum, H. Y., and Freeman, W. T. (2007). Face hallucination: Theory and practice. *International Journal of Computer Vision*, **75**(1), 115–134.

Ma, X., Zhang, J., and Qi, C. (2010). Hallucinating face by position-patch. *Pattern Recognition*, **43**(6), 2224–2236.

Mandal, T., Jonathan Wu, Q. M., and Yuan, Y. (2009). Curvelet based face recognition via dimension reduction. *Signal Processing*, **89**(12), 2345–2353.

Martinez, A. M. and Benavente, R. (1998). The AR face database. *CVC Technical Report*, **24**.

Milanfar, P. (2010). *Super-resolution imaging*, volume 1. CRC Press.

Mitchell, H. B. (2010). *Image fusion: theories, techniques and applications*. Springer.

Patti, A. J. and Altunbasak, Y. (2001). Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants. *IEEE Transactions on Image Processing*, **10**(1), 179–186.

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

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., and Worek, W. (2006). Preliminary face recognition grand challenge results. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 15–24. IEEE.

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

Schultz, R. R. and Stevenson, R. L. (1994). A bayesian approach to image expansion for improved definition. *IEEE Transactions on Image Processing*, **3**(3), 233–242.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, **37**(1), 10–21.

Štruc, V., Žibert, J., and Pavešić, N. (2009). Histogram remapping as a preprocessing step for robust face recognition. *image*, **7**(8), 9.

Tom, B. C. and Katsaggelos, A. K. (1995). Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. In *International Conference on Image Processing*, volume 2, pages 539–542. IEEE.

Tsai, R. and Huang, T. S. (1984). Multiframe image restoration and registration. *Advances in computer vision and Image Processing*, **1**(2), 317–339.

Turk, M. A. and Pentland, A. (1991). Face recognition using eigenfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591. IEEE.

Wang, X. and Tang, X. (2005). Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man, and Cybernetics*, **35**(3), 425–434.

Wolberg, G. (1990). *Digital image warping*, volume 10662. IEEE computer society press Los Alamitos.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S., and Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**(2), 210–227.

Xue, K., Winans, A., and Walowit, E. (1992). An edge-restricted spatial interpolation algorithm. *Journal of Electronic Imaging*, **1**(2), 152–161.

Yang, J., Tang, H., Ma, Y., and Huang, T. (2008a). Face hallucination via sparse coding. In *15th IEEE International Conference on Image Processing*, pages 1264–1267. IEEE.

Yang, J., Wright, J., Huang, T., and Ma, Y. (2008b). Image super-resolution as sparse representation of raw image patches. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

Yang, J., Wright, J., Huang, T. S., and Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, **19**(11), 2861–2873.

Zhang, W. and Cham, W. (2011). Hallucinating face in the dct domain. *IEEE Transactions on Image Processing*, **20**(10), 2769–2779.

Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, **35**(4), 399–458.

Zhuang, Y., Zhang, J., and Wu, F. (2007). Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation. *Pattern Recognition*, **40**(11), 3178–3194.

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

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.