# Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

**By reading and using the thesis, the reader understands and agrees to the following terms:**

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.

2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.

3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMAGE-PROCESSING TECHNIQUES FOR

ROBUST FACE RECOGNITION

MUWEI JIAN

Ph.D

The Hong Kong Polytechnic University

2014

THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF ELECTRONIC AND INFORMATION ENGINEERING

# Image-Processing Techniques for Robust Face Recognition

Muwei Jian

A thesis submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

March, 2014

# CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

(Signed)

  Muwei Jian    (Name of student)

# Abstract

Current face-recognition algorithms can achieve a highly accurate performance under controlled conditions, such as unchanged light sources, frontal-view pose, no occlusion, neutral facial expression, etc. Face recognition has a wide range of applications, however it still has many technical and challenging issues to be solved, in particular when the faces under consideration have a very low resolution, different illumination conditions, arbitrary poses and are under occlusions. In order to achieve robust face recognition, we have investigated efficient techniques to solve some typical challenging problems for robust face recognition.

First, among the facial features, the eyes play the most important role in face recognition and face hallucination. Most of the existing facial-feature detection and localization algorithms cannot work accurately when the faces are rotated or under poor lighting conditions. Therefore, in this research, an efficient algorithm for eye detection in face images is proposed. As the eye region always has the most variations in a face image, our algorithm uses a wavelet-based salient map, which can detect and reflect the most visually meaningful regions for coarse eye detection. With the aid of a pose-adapted eye template - which can handle eye regions with large rotation and pose variations, accurate eye positions can be localized. Furthermore, the position of the nose and mouth can be determined by considering both the saliency values in the salient map and the detected eye positions as geometric references.

Second, face images under different illuminations represent a challenge for face recognition. In this research, we have discussed an efficient scheme for illumination compensation and the enhancement of face images. Our illumination model is universal; it does not require the assumption of a single-point light source. Thus, it circumvents and overcomes the limitations of the Lambertian model. The proposed approach can learn the average representations of face images under changing illuminations so as to compensate or enhance the face images, and also to eliminate the effect of different and uneven illuminations, while retaining the

intrinsic properties of the face surface. Our experiments have provided promising results, demonstrating that our proposed methods are effective.

Third, in order to achieve robust face recognition and to make face-recognition systems capable of identifying people at very low resolution, super-resolution (SR) technology is investigated. In this thesis, we first introduce facial-image super-resolution, which is also called face hallucination. In this research, an efficient mapping model is first proposed for face hallucination. Since we can observe and prove that the singular values of an image at one resolution, represented by singular value decomposition (SVD), have approximately linear relationships with their counterparts at other resolutions, the estimation of the corresponding singular values of the high-resolution (HR) face images becomes more reliable. From the signal-processing viewpoint, this can effectively preserve and reconstruct the dominant information in the reconstructed HR face images. The mapping scheme can be viewed as a "coarse-to-fine" estimation of HR face images. Compared to other, state-of-the-art algorithms, experiments have shown that our proposed face-hallucination scheme is practicable and effective.

Fourth, a framework based on singular value decomposition (SVD) for performing both face hallucination and recognition simultaneously is also proposed. Conventionally, low-resolution (LR) face recognition is carried out by super-resolving the LR input face first, and then performing face recognition to identify the input face. By considering face hallucination and recognition simultaneously, the accuracy of both hallucination and recognition can be improved. In our algorithm, each face image is represented by using SVD. For each LR input face, the corresponding LR and high-resolution (HR) face-image pairs can then be selected from the face gallery. With the aid of face recognition, using the selected LR-HR pairs, the estimation of the mapping functions for interpolating the two matrices in the SVD representation of the corresponding HR face image can be more accurate.

All these techniques can be integrated with both existing and new face recognition algorithms so as to achieve a robust and good performance level.

# List of Publications

[1] **Muwei Jian**, Kin-Man Lam, Junyu Dong, "A Novel Face-Hallucination Scheme Based on Singular Value Decomposition," *Pattern Recognition*, vol. 46, issue 11, pp. 3091-3102, November 2013. (Impact factor: 2.6)

[2] **Muwei Jian**, Kin-Man Lam, Junyu Dong, "Facial-Feature Detection and Localization Based on a Hierarchical Scheme, *Information Sciences*. vol. 262, pp. 1-14, 2014. (Impact factor: 3.6)

[3] **Muwei Jian**, Kin-Man Lam, Junyu Dong, "Illumination-insensitive Texture Discrimination Based on Illumination Compensation and Enhancement", *Information Sciences*, vol. 269, pp. 60-72, 2014. (Impact factor: 3.6)

[4] **Muwei Jian**, Kin-Man Lam, "Face-Image Retrieval Based on Singular Values and Potential-Field Representation." *Signal Processing*, vol. 100, pp. 9-15, 2014. (Impact factor: 1.85)

[5] **Muwei Jian**, Junyu Dong, Kin-Man Lam, "FSAM: A fast self-adaptive method for correcting non-uniform illumination for 3D reconstruction", *Computers in Industry*, vol. 64, Issue 9, pp. 1229-1236, December 2013. (Impact factor: 1.7)

[6] **Muwei Jian**, Kin-Man Lam, Junyu Dong, "Visual-patch-attention-aware Saliency Detection", accepted to appear in *IEEE Transactions on Cybernetics*.

[7] **Muwei Jian**, Kin-Man Lam, "Simultaneous Hallucination and Recognition of Low-resolution Faces Based on Singular Value Decomposition", revised version submitted to *IEEE Trans. on Circuits and Systems for Video Technology*.

[8] **Muwei Jian** and Kin-Man Lam, "Face Super-Resolution Based on Singular Value Decomposition," Proceedings, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA'2012), December 2012, Hollywood, U.S.A.

[9] **Muwei Jian**, Kin-Man Lam and Junyu Dong, "Dynamic Textures Indexing and Retrieval Based on Intrinsic Properties," Proceedings, IEEE International Symposium on Circuits and Systems (ISCAS'2012), pp. 866-9, May 2012, Seoul, Korea.

[10] **Muwei Jian** and K.M. Lam, "Fast Eye Detection and Localization Using a Salient Map," Proceedings, Pacific-Rim Conference on Multimedia (PCM'2011), pp. 74-83, December 2011, Sydney, Australia.

[11] **Muwei Jian**, K.M. Lam and Junyu Dong, "Illumination Compensation and Enhancement for Face Recognition," Proceedings, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC'2011), paper Wed-AM.RS6, October 2011, Xi'an, China.

[12] Yujuan Sun, **Muwei Jian**, Junyu Dong, Haozhi Yang and K.M. Lam, "Facial shape and albedo reconstruction based on a trained prototype," Proceedings, IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC'2012), pp. 425-9, August 2012, Hong Kong.

# Acknowledgment

I feel much indebted to many people who have instructed and favored me in the course of writing this thesis.

First of all, I would like to express my sincere heartfelt thanks to my nice supervisor Prof. Kin-Man Lam, for his helpful and constructive comments and suggestions, constant encouragement and precise modification. I admire his knowledge and personality. He not only grants me the opportunity to enter the world of academic research, but also give me the consistent and illuminating instruction and support in this colorful three years of study. Without the help and support of him, I could not have completed this thesis.

My thanks also go to Prof. Guoping Qiu, and Prof. Junyu Dong. The insightful and fairly extensive discussions with them have broadened my view and enlightened my mind, and I have learned a lot from their knowledge and experience.

I am fortunate to have the chance to work with Kwok-Wai Wong, Kuong-Hon Pong, Zhuo Hui, Yu Hu, Chensheng Sun, Sanyuan Zhao, Chao Li, Wentao Liu, Xiaolong Ma, Deliang Yang, Dong Li, Hailiang Li, Jiwei Hu, Huiling Zhou and Cigdem Tuguan during my Ph.D. study. I am thankful to their share of knowledge and thoughts at work and meeting, and the life we have enjoyed together.

Moreover, my special thanks should go to my mother, who has not a change to go to school when she is young. Although she is an illiterate person, she always gives consistent encouragement and endless support to me during my whole study. I would like again to thank my mother for her endless love. At last, my deepest gratitude also goes to my wife and my five-year-old son for their love and great

confidence in me through these years. They are always my motivation and strong backing.

# Contents

x

# List of Figures

# List of Tables

# Chapter 1 Introduction

## 1.1 Face recognition

Face recognition is a biometric technique which can be formulated as the problem of identifying or verifying one or more persons in a scene by comparing input faces (probes) with the face images stored in a database (galleries) [1]. In contrast to other types of biometrics (such as fingerprint, hand geometry, iris, and retina scans), face recognition offers a non-intrusive and the most natural way of person identification/authentication [1]. As is well known, the advantage of face recognition is intuitive and convenient, and often effective without the participants' cooperation or knowledge.

Face recognition has a wide range of applications (including access control, security monitoring, facial-paralysis diseases diagnosis, and video surveillance), and has drawn significant attention from various fields, e.g. cognitive psychology, neural networks, image processing, pattern recognition, computer vision, computer graphics. Over the past several decades, many face recognition approaches have been proposed: Principal Component Analysis (PCA) [2][3] and Linear Discriminant Analysis (LDA) [4][5][6] are two of the most common approaches, and have long been used as benchmarks for face recognition. However, these classical and holistic-based face recognition algorithms require the accurate positions of the two eyes for normalization and alignment. Furthermore, all the holistic-based algorithms require a huge training set with multiple images in different poses and expressions for each person. On the other hand, feature-based methods extract local features, like eyes, nose, and mouth, to form feature vectors, which are fed into a structural classifier for face recognition. Typical algorithm is based on Elastic Bunch Graph Matching (EBGM) [7]. Compared to those holistic approaches, a major advantage of EBGM is that it can recognize human faces by comparing their corresponding parts instead of requiring a large training set for efficient recognition. Under controlled environments, both holistic-based and feature-based face recognition algorithms can achieve a high performance.

# 1.2 Review of typical face recognition approaches

Since face recognition has the additional advantage of being a passive and non-intrusive method, in the past three decades, automatic recognition of faces has become an active research topic in the fields of image processing and pattern recognition. In the following sections, a brief overview of some typical face recognition approaches is given.

### 1.2.1. Principal Component Analysis (PCA)

Among the different face recognition approaches, Principal Component Analysis (PCA), also known as Eigenface, is the most classical method. The objective function of PCA is to find a transformation that can represent high-dimensional data in lower dimensions such that the maximum information about the data is present and retained in the transformed space [2][3]. The basic idea of the Eigenface method is introduced in the following [2][3].

A face image can be considered as a column vector by concatenating its rows or columns one by one, denoted as $\Gamma$. Let $\Gamma_1$, $\Gamma_2$,..., $\Gamma_M$ be a training set of face images. The average of the training samples is given as follows:

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n .$$ (1.1)

The difference of each training face image from the average is computed as $\Phi_n = \Gamma_n - \Psi$. The covariance matrix is then calculated as follows:

$$COV = \frac{1}{M}\sum_{n}^{M}\Phi_n\Phi_n^T = AA^T ,$$ (1.2)

where the matrix $A = \left[\Phi_1\Phi_2...\Phi_M\right]$. In practice, the matrix $COV$ is large for computing its eigenvectors. A trick [2, 3] was suggested to determine the eigenvectors $v_n$ of $A^TA$ first, i.e.

$$A^TAv_n = \lambda_n v_n .$$ (1.3)

Multiply both sides by $A$, we have

$$AA^T A v_n = \lambda_n A v_n . \tag{1.4}$$

Thus, the $A v_n$ on the right represents an eigenvector of $COV = AA^T$, and the eigenvector, denoted as $\mu_n$, is a linear combination of the $M$ training samples:

$$\mu_n = \sum_{k=1}^{M} v_{nk} \Phi_k = A v_n, \ n = 1,......,M . \tag{1.5}$$

The eigenvectors represents the principal components of the face images, and are called as the Eigenfaces.

For face recognition, a new or query face image, denoted as $\Gamma$, is projected onto the Eigen-space as follows:

$$\omega_k = \mu_k^T (\Gamma - \Psi) . \tag{1.6}$$

The weights form a vector $\Omega^T = [\omega_1, \omega_2,..., \omega_M]$ that describes the contribution of each eigenvector in representing the input face image. The simplest method for determining which face class provides the best description of an input face image is to find the face class $k$ that minimizes the Euclidean distance :

$$\varepsilon_k = \| \Omega - \Omega_k \| , \tag{1.7}$$

where $\Omega_k$ is a vector describing the *kth* face class. A face is classified as belonging to class $k$ when the minimum $\varepsilon_k$ is below some chosen threshold $\theta_k$. Otherwise the face is classified as "unknown" [2].

Eigenface is a well-known practical approach that is computationally efficient for face recognition. However, all face images have a similar structure, so they are highly correlated with each other. In other words, PCA is more effective for facial-image representation than for face recognition. In addition, the images must be aligned accurately, otherwise the performance of the Eigenface method degrades significantly.

## 1.2.2. Linear Discriminant Analysis (LDA)

Besides PCA, another well known linear transformation methods for face recognition is Linear Discriminant Analysis (LDA) [7]. Contrast to PCA, which aims to find a linear transformation that maximum information of the data is

preserved in the transformed space, the objective of LDA is to preserve as much of the class discriminant information as possible in the transformation.

The main purpose of Linear Discriminant Analysis (LDA) method is that it attempts to maximize the ratio of between-class scattering to within-class scattering by using the Fisher's Linear Discriminant (FLD). The Fisherface method is briefly described below [4, 5, 6].

Consider a set of $N$ sample images $\{x_1, x_2, ..., x_N\}$, and assume that each image belongs to one of $C$ classes $\{X_1, X_2, ..., X_C\}$. Let the between-class scatter matrix be defined as follows:

$$S_B = \sum_{i=1}^{C} N_i (\mu_i - \mu)(\mu_i - \mu)^T , \tag{1.8}$$

and the within-class scatter matrix be given as:

$$S_W = \sum_{i=1}^{C} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T , \tag{1.9}$$

where $\mu_i$ is the mean image of class $X_i$ and $N_i$ is the total number of face images in class $X_i$. If the within-class scatter matrix $S_W$ is non-singular, the optimal projection $W_o$ is chosen as the matrix with those columns, which maximizes the ratio of the between-class scatter matrix of the projected samples to the within-class scatter matrix of the projected samples [4]:

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

$$= [\mathbf{W}_1 \quad \mathbf{W}_2 \quad ... \quad \mathbf{W}_m], \tag{1.10}$$

where $\{\mathbf{W}_i \mid i = 1, 2, ..., m\}$ is the $m$ leading eigenvectors of the following eignvalue problem:

$$S_B \mathbf{W}_i = \lambda_i S_W \mathbf{W}_i, \ i = 1, 2, ..., m. \tag{1.11}$$

It should be note that there are at most $c-1$ nonzero generalized eigenvectors, so the upper bound on $m$ is $c-1$, where $c$ is the number of classes. For face recognition, the difficulty is that the within-class scatter matrix $S_W$ is always singular. In order to slove this problem, PCA can be firstly used to reduce the dimension of the

feature space into $N-c$, and then the standard FLD of equation 1.10 is applied to reduce the dimension to $c-1$. $W_o$ is more formally defined as follows:

$$W_o^T = W_{fld}^T W_{pca}^T, \tag{1.12}$$

where

$$W_{pca} = \arg\max_W \left| W^T S_T W \right|;$$

$$W_{fld} = \arg\max_W \frac{\left| W^T W_{pca}^T S_B W_{pca} W \right|}{\left| W^T W_{pca}^T S_W W_{pca} W \right|}.$$

Note that the Eigenface method uses PCA for dimensionality reduction, which yields projection directions that maximize the total scatter across all classes of images. This projection is best for the reconstruction of images from a low-dimensional basis. However, this method does not make use of any between-class scatter. The projections are not optimal for discrimination between different classes. Although Fisherface is similar to Eigenface, it considers between-class scatter and can improve the discriminant capability of classifying faces from different classes. With the aid of Fisher's Linear Discriminant (FLD), the Fisherface method can classify different people with varying facial expressions [4]. Nevertheless, Fisherface is more complex than Eigenface in terms of finding the projection vectors for face recognition, resulting in a larger storage requirement for the representation and more computation time for face recognition.

### 1.2.3. Independent Component Analysis (ICA)

In addition to PCA and LDA, Independent Component Analysis (ICA) is another widely used subspace-projection method, projecting data from a high-dimensional space to a lower-dimensional space. Motivated by the fact that much of the important information may be represented in the high-order relationships, face recognition based on ICA is also proposed. Compared to PCA, which uses those eigenvectors capturing the maximum image variance to determine the basis vectors, ICA provides a set of basis vectors that possesses maximum statistical independence. The basic idea of ICA algorithm is briefly introduced as follows [7, 129].

Let $\mathbf{S}$ be the vectors of unknown sources signals and $\mathbf{X}$ be vectors of observed mixtures. If $\mathbf{A}$ is an unknown mixing matrix, the mixing model can be expressed as $\mathbf{X} = \mathbf{AS}$. The main task is to estimate the independent source signals $\mathbf{U}$ by computing the separating matrix $\mathbf{W}$ that corresponds to the mixing matrix $\mathbf{A}$ using the following relation [129]:

$$\mathbf{U} = \mathbf{WX} = \mathbf{WAS}. \tag{1.13}$$

Let the observed samples are whitened and denoted by $\mathbf{Z}$. Then the kurtosis of $\mathbf{U}_i = \mathbf{W}_i^T \mathbf{Z}$ can be obtained and the separating vector $\mathbf{W}_i$ is computed via maximizing the kurtosis:

$$kur(\mathbf{U}_i) = |E\{(\mathbf{U}_i)^4\} - 3(E\{(\mathbf{U}_i)^2\})^2|. \tag{1.14}$$

Traditionally, ICA can be applied to face recognition in two different architectures [7]. The goal in Architecture I was to use ICA to find a set of spatially independent basis images, whereas the ICA architecture II finds statistically independent coefficients that represent input images. Face recognition performance was evaluated in [7] shows that there was no significant difference in the performances of the two ICA representations .

As discussed in [7], both PCA and ICA can be derived as generative models of the data, where PCA uses Gaussian sources, and ICA typically uses sparse sources for face recognition. It has been shown that for many natural signals, ICA is a better model in that it assigns higher likelihood to the data than PCA [3]. The ICA basis dimensions presented here may have captured more likelihood of the face images than PCA, which provides a possible explanation for the superior performance of ICA for face recognition [129].

## 1.2.4. Elastic Bunch Graph Matching (EBGM)

Face recognition techniques based on Elastic Bunch Graph Matching (EBGM) [12] shows successful experimental results. The EBGM representation of facial feature is based on Gabor wavelet transform. Gabor wavelet, which is also called Gabor filtering, can also produce a multi-resolution representation of the sample texture image. Gabor filtering provides a flexible scheme for designing efficient algorithms

to capture more orientation and scale information in signals. A two dimensional Gabor function $g(x, y)$ is defined as

$$g(x, y) = \frac{1}{2\pi\sigma_X\sigma_y}\exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_X{}^2} + \frac{y^2}{\sigma_y{}^2}\right) + 2\pi jWx\right), \qquad (1.15)$$

where $\sigma_X$ and $\sigma_y$ are the standard deviations of the Guassian envelopes along the $x$ and $y$ direction [1]. Then a set of Gabor filters can be obtained by appropriate dilations and rotations of $g(x, y)$:

$$g_{mn}(x, y) = a^{-m}g(x', y'),$$

$$x' = a^{-m}(x\cos\theta + y\sin\theta);$$
$$y' = a^{-m}(-x\sin\theta + y\cos\theta). \qquad (1.16)$$

where $a > 1$, $\theta = n\pi/K$, $n = 0, 1, \ldots, K\text{-}1$, and $m = 0, 1, \ldots, S\text{-}1$. $K$ and $S$ are the number of orientations and scales. The scale factor $a^{-m}$ is to ensure that energy is independent of m. Given an image $I(\vec{x})$ around a given pixel $\vec{x} = (x, y)$, its Gabor transform is defined as a convolution:

$$J_j(\vec{x}) = \int I(\vec{x})g^*_{mn}(\vec{x} - \vec{x}_1)d^2\vec{x}_1 = \int I(x, y)g^*_{mn}(x - x_1, y - y_1)dx_1dy_1, \quad (1.17)$$

where * indicates the complex conjugate. A jet $J$ is defined as the set $\{ J_j \}$ of 40 complex coefficients obtained for one image pixel. The set of 40 coefficients obtained for one image point is referred as a jet. However, jets taken from image points only a few pixels apart from each other have very different coefficients due to phase rotation. This may decrease the accuracy of matching. Therefore, a method is proposed to compare the jets by taking into account the phase difference in comparison, and the phase similarity function is defined as follows:

$$S_a(J, J') = \frac{\sum\limits_j a_j a'_j \cos(\phi_j - \phi'_j - \bar{d}\bar{k}_j)}{\sqrt{\sum\limits_j a_j^2 \sum\limits_j a'^2_j}}, \qquad (1.18)$$

By using this phase function, the phase difference $(\phi_j - \phi'_j)$ is compensated by the displacement $\bar{d}$, which is estimated using Taylor expansion. The displacement estimation could be done using the disparity estimation.

In order to represent a face, EBGM method builds an image graph from a set of fiducial points like the pupils, the corner of the mouth, the tip of the nose, the top and bottom of ears, etc. A labeled graph $G$ representing a face consists of $N$ nodes on the fiducial points at position $\bar{x}_n$, $n = 1, ..., N$ and $E$ edges between them.

For an automatic face recognition system, it has to locate the fiducial point and build the image graph from an input image automatically. This can be done by matching the input image with a stack like general representation of faces, Face Bunch Graph (FBG). For the matching between an input graph and the Face Bunch Graph (FBG), a function called graph similarity is employed. The graph similarity function depends on the jet similarity mentioned before and the distortion of the image grid relative to the FBG grid. For an image graph $G^I$ with nodes $n = 1, ..., N$ and $e = 1, ... ,E$ and an FBG $B$ with model graphs $m = 1, ... ,M$. The similarity is defined as

$$S_B(G^I, B) = \frac{1}{N} \sum_n \max_m (S_\phi(J_n^I, J_n^{B_m})) - \frac{\lambda}{E} \sum_e \frac{(\Delta \bar{x}_e^I - \Delta \bar{x}_e^B)^2}{(\Delta \bar{x}_e^B)^2}, \qquad (1.19)$$

where $\lambda$ determines the relative importance of jets and metric structure. $J_n$ are the jets at nodes $n$, and $\Delta \bar{x}_e$ are the distance vectors used as labels at edge e. Because the FBG provides several jets for each fiducial point, the best one is selected and used for comparison. These best fitting jets serve as local experts for the image face [ 7].

For face recognition, after having extracted model graphs from the gallery images and image graphs from the probe images, recognition can be conducted by comparing an image graph to all model graphs and selecting the face with the highest similarity value. The similarity function is an average over the similarities between pairs of corresponding jets. If $G^I$ is the image graph, $G^M$ is the modal graph, and node $n_{n'}$ is the modal graph corresponds to node $n'$ in the image graph, the graph similarity is defined as

$$S_G(G^I, G^M) = \frac{1}{N'} \sum_{n'} S_a(J_{n'}^I, J_{n_{n'}}^M), \qquad (1.20)$$

where the sum runs only over the $N'$ nodes in the image graph with a corresponding node in the modal graph.

Experiment result show that Elastic Bunch Graph Matching (EBGM) works well with different facial expression and with different scales. However, this algorithm has certain drawbacks. It is quite complicated to build the FBG at the initial stage. A large amount of grid placements has to be done manually at the beginning. What's more, it is difficult to implement because of the complexity of the algorithm in automatically finding the position of the fiducial points. And it requires huge storage of convolution images for better performance.

In short, despite the success of face recognition techniques in many practical applications, there are still many technical and challenging problems under uncontrolled environments to be solved [1].

## 1.3 Challenging issues of face recognition

Although current face-recognition algorithms can achieve a highly accurate performance under controlled conditions (such as unchanged light sources, frontal-view images, no occlusion, neutral facial expression, etc.), it still has many technical and challenging issues to be solved. Especially, face recognition with very low resolution and under different illuminations, are typical difficult problems, in particular in outdoor circumstances. Fig. 1.1 shows two typical challenging problems for current face recognition.



Very low resolution　　　　　Illumination

Fig. 1.1 Two typical challenging problems for face recognition.

In reality, face recognition is always under outdoor or uncontrolled conditions

in real-world applications. For example, face images are captured at a long distance from the camera in a surveillance system. For face images in surveillance videos, mostly are of low resolution and are compressed with a high compression ratio, etc. All these problems constitute a difficult and challenging issue in face recognition. Fig. 1.2 shows an example of faces captured in a surveillance system.

As shown in Fig. 1.2, low resolution problem is one of the hard issues in current face recognition. Constructing a high-resolution (HR) image from its low-resolution (LR) inputs is called super-resolution in the image-processing research field. For face images, this technology is also called face hallucination [8], which has become one of the most important fields for face recognition. Face hallucination is helpful for face recognition. Especially, this technique is useful when face-recognition systems confront the low-resolution of face images.

Apart from low-resolution problem, illumination variation is another major problem for current face recognition. Human faces share a similar shape and structure, but illumination variations and different lighting directions always make images of the same person look dissimilar. As shown in Fig. 1.3, the different face images of the same person with variations in illuminations are not discernible as the same man.



(a)                                   (b)                                   (c)

Fig. 1.2. Low-resolution of faces captured in a surveillance system. (a) and (c) Low-resolution of faces in video-surveillance videos; (b) video-surveillance videos compressed with a high compression ratio.

In real-world applications, face recognition with different illuminations is a difficult problem, in particular in outdoor circumstances. Illumination variations remain an unsolved problem in face recognition, despite a lot of research having been devoted to solving it [1]. In the past decade, the illumination problem has received considerable attention in both face-recognition-related industries and academic circles. However, it is still one of the most prominent issues for appearance- or image-based face recognition approaches. The development of illumination-compensation techniques for face recognition is important, and modeling face variations in realistic settings is still a heuristic issue. Without solving this problem, accurate and robust face recognition cannot be achieved [1] [9].
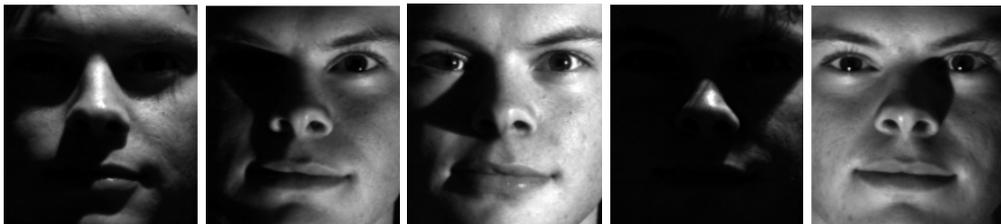


Fig. 1.3. Five face images of the same person from the YaleB face database under different illumination conditions.

At present, most of these face hallucination and face recognition algorithms assume that the face images under consideration have been aligned accurately with respect to the important facial features, such as the eyes and mouth. This assumption makes these algorithms semi-automatic. Most of the existing facial feature detection and localization algorithms [8] cannot work accurately and reliably when the eyes are closed, the faces are rotated, or the face are under a poor lighting condition. This will, in turn, have an adverse effect on the performance of both face hallucination and face recognition algorithms. In addition, a lot of research has found that the performances of these algorithms will degrade significantly if the face images under consideration are not aligned properly, regardless of whether these systems are based on local or global methods

[10][11][12][13]. In [10], Shan et al. called this phenomenon "Curse of mis-alignment in face recognition". Fig. 1.4 shows that the rank-1 recognition rate of the Fisherface method [6], which degrades seriously with the increase of the mis-alignment [10]. Wang et al. [11] also found that, even if the eye-location error is only about 5%, the face recognition accuracy reduces dramatically – by about 50%. Therefore, the accurate detection and localization of facial features play a very important role in face recognition systems.
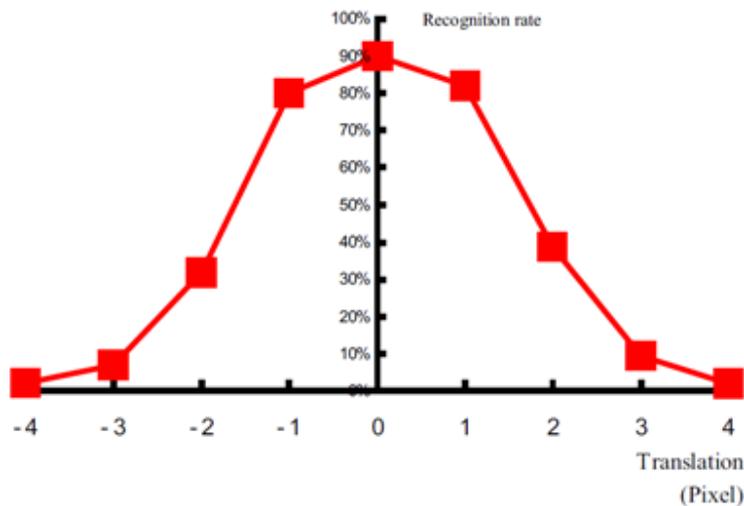


Fig. 1.4. Relationship between the rank-1 recognition rate of the Fisherface [6] and the mis-alignment of translation [10].

To sum up, although current face-recognition algorithms, under controlled conditions, achieve a highly accurate performance, there are still many hard and challenging issues to be solved for robust face recognition.

## 1.4 Summary

This chapter is started with an introduction to face recognition. Then, we have reviewed some different challenging issues of face recognition. Three main challenging issues, which are emphasized in this thesis, have been briefly introduced, namely (1) the accurate eye-detection problem, (2) the illumination-variation issue, and (3) the low-resolution problem, respectively.

# 1.5 Scope of this thesis

In this thesis, we aim to investigate the challenging issues of current face recognition. Face recognition has a wide range of applications, nevertheless still has many technical and challenging issues to be solved. For robust face recognition, the efforts toward achieving this objective consist of three technical and inseparable parts:

    (1) the accurate eye-detection problem,

    (2) the illumination-variation issue, and

    (3) the low-resolution problem.

In this thesis, we will put emphasis on the above three technical and challenging issues one by one. The brief introduction of each chapter and content of this thesis are outlined as follows:

- A brief introduction to Chapter 2 - We put our effort on the accurate eye-detection problem for robust face recognition. In this chapter, a novel scheme is proposed, which uses a two-level hierarchical structure based on a wavelet-based salient map and PCA-based verification to locate eye regions coarsely, and uses a pose-adapted eye template, based on the prior knowledge of the sclera's photometric and geometric properties, for precise human eye localization.

  First of all, face regions are located using a cascade of boosted classifiers [110]. In the first level of our algorithm, a wavelet-based saliency map is proposed and applied to the detected face region for a coarse detection of its facial-feature regions. It is observed that the eye regions are always the most varied blocks in a face image. Therefore, those rectangular regions in the saliency map having the largest corresponding sums of saliency values should be the potential candidates for the two eye regions. These eye-region candidates are verified using PCA, so the false-alarm rate can be significantly reduced.

  Within the eye regions, the accurate locations of the eyes are then further determined. The first level of detection can greatly narrow down the search space for finding the actual eye positions, and thus can reduce the

computational cost needed. In the second level of our hierarchical scheme, a pose-adapted eye template is constructed using the intrinsic photometric and geometric properties of eyes for accurate eye detection and localization. Based on the detected eye positions, the other facial features (nose and mouth) can be determined by considering their respective sums of the saliency values and by using the two eye positions detected as a reliable geometric reference.

In brief, an efficient hierarchical scheme, which is robust to illumination and pose variations in face images, is proposed for accurate facial-feature detection and localization. Our proposed algorithm is non-iterative and computationally simple. Experimental results show that our algorithm can achieve a superior performance compared to other state-of-the-art methods.

- A brief introduction to Chapter 3 - The illumination-variation issue: We will present a facial-image lighting-compensation and -enhancement algorithm for face recognition in this chapter.

According to the Retinex theory, the intensity of an image can be represented as the product of illumination and surface reflectance. Face images of the same person have identical facial structures and patterns, sharing a similar shape surface reflectance structure, so it is reasonable to assume that the surface-reflectance-representation matrix is a slowly-changing matrix, which reflects the intrinsic property of a face surface. Consequently, the dissimilarity between images of the same person under different illumination conditions is mainly caused by the differences in the illumination-effect matrix. Due to the fact that images under uneven illuminations produce shadows, and look different in those regions with insufficient illumination. If we can learn a mean illumination-effect matrix so as to compensate the component of the illumination-effect matrix of the images with uneven lighting and shadows, it will make images lighter and shadowless. We can learn the mean illumination-effect matrix to compensate the images with uneven illuminations and shadows. Instead of deriving universal representations, illumination compensation and enhancement utilizing specific individual information can possibly provide an effective and useful way to achieve a better appearance and a higher

recognition rate. In this chapter, we present a computationally efficient method for generating illumination-invariant texture and face images using the scheme of illumination compensation and enhancement. We use an illumination model which is universal and does not require the assumption of a single-point light source, thereby overcoming the limitation of the Lambertian model. Our proposed approach captures the mean illumination-effect matrix representations of images under a variety of different illumination conditions for each class, so as to compensate or enhance the images, and consequently.

In particular, we will aim at devising a simple and effective scheme to compensate/enhance illumination, and to remove the shadow caused by uneven illumination, to achieve better face recognition performance. The advantage of the proposed method is that the assumption of a single-point light source is not required, so it circumvents and overcomes the limitations of the Lambertian model and is also suitable for outdoor circumstances. Experimental results have produced promising results, which demonstrate the effectiveness of our proposed method.

- A brief introduction to Chapter 4 - The low-resolution problem: In this chapter, we will introduce a novel face-hallucination scheme based on singular value decomposition (SVD).

We firstly verify that a simple mapping model in the image space $R^{m \times n}$ is inappropriate and unfeasible, as the mapping scheme is too coarse to generate satisfactory results. Then, we propose a more sophisticated mapping model in the eigenspace based on SVD. This proposed face-hallucination framework consists of three steps. In the first step, $M$ example faces that share the most similarity to the input LR face image are searched from a database, and the optical-flow method is then used to align these $M$ example image pairs. In the second step, we compute the leading eigenvectors, which account for most of the information contained in the image. We also observe and prove that, based on the Frobenius norm, the corresponding singular values of an image across different resolutions have approximately linear relationships. Furthermore, we can interpolate the other two data matrices generated by SVD to a higher

resolution, as the interpolation of these matrices will not change the holistic structure or the pattern of the face image. The mapping scheme, which utilizes the interpolated SVD matrices multiplied by the learned corresponding mapping matrices to generate more details of face images, can be viewed as a "coarse-to-fine" manner. In the third step, a residual-error matrix, which represents the high-frequency information or the detailed local features missed in the previously predicted HR face image, is generated and added to the one produced in the second stage. Experimental results show that our proposed face-hallucination scheme is effective in terms of producing plausible HR images with both a holistic structure and high-frequency details.

- A brief introduction to Chapter 5 - A framework based on SVD for performing both face hallucination and recognition simultaneously is proposed in this chapter.

In our scheme, face images are represented using SVD, and the hallucination and recognition of LR faces are taken into account simultaneously. We have proved [124] that, based on the Frobenius norm, the corresponding singular values of an image across different resolutions have approximately a linear relationship. This makes the estimation of the singular values of HR face images more reliable. Furthermore, the left and right matrices in the SVD representation can be interpolated to a higher resolution using bicubic interpolation; this interpolation method applied to these two matrices will not change the holistic structure or the pattern of the face image. Our proposed approach can be applied to both face verification and identification.

Our simultaneous face-hallucination and -verification algorithm is denoted as SHV, As the claimed identity of the query is known, the claim will simply be rejected if the difference between the singular values of the query and those of the claimed faces in the database is larger than a certain threshold. If the difference is smaller than the threshold, super-resolution will be performed based on the mapping models learned from the claimed LR-HR face pairs. Then, hallucination is performed based on the LR-HR face pairs of the claimed identity, and the LR-to-HR mapping matrices of the respective

claimed face pairs are learned for estimating the high-frequency information or any detailed local features missed in the estimated HR faces generated by interpolating the two SVD matrices. The hallucinated faces are then used for verification again On the other hand, the algorithm for simultaneous face hallucination and identification is denoted as SHI. In this algorithm, $Q$ faces that are the most similar to the input LR face image are first searched from a gallery database of LR-HR pairs based on its singular values. Suppose that these $Q$ faces belong to $M$ distinct subjects, where $M < Q$. For each of these $M$ identities, the corresponding mapping models are learned and used to super-resolve the query input. Therefore, $M$ HR face images for the LR query are generated. Then, the differences between each of the $M$ HR face images and the corresponding HR face images in the database are computed based on PCA. The input LR face is assigned to the class of the face with the smallest difference.

Conventionally, by considering face hallucination and recognition simultaneously, the accuracy of both the hallucination and the recognition can be improved.

● A brief introduction to Chapter 6 - The discussions and conclusions of this thesis will be given in this chapter. We will also pinpoint the contributions of our work in robust face recognition. The chapter ends with some pointers to possible future works related to robust face recognition. Some possible future of research on robust face recognition will also be discussed.

# Chapter 2 Facial-Feature Detection and Localization Based on a Hierarchical Scheme

Accurate facial-feature (such as eyes, nose, mouth, etc.) detection and localization are essential for a wide range of computer-vision applications (such as driver-fatigue monitoring systems [81], face recognition [2][3][4][5][6][82] face tracking, iris recognition [83], human-computer interaction [84], etc.) and biometric/cognitive psychological tasks (such as facial-paralysis diseases diagnosis/evaluation [85], control devices for disabled people, facial-expression analysis for human affective states [86], age estimation, user attention and gaze for marketing analysis, etc.). Since the detection and localization of facial features play a vital role in many tasks, they have attracted many researchers' attention from different industrial sectors and different academic disciplines.

In the past few decades, a lot of face recognition approaches have been proposed: Principal Component Analysis (PCA) [2][3] and Linear Discriminant Analysis (LDA) [4][5][6] are two of the most classical and representational approaches, and have long been used as benchmarks for face recognition. However, these holistic-based face recognition algorithms require the accurate positions of the two eyes and the mouth for normalization and alignment. In addition, most of the current face-hallucination (or face super-resolution) algorithms [8] also require an precise alignment of the key facial features. Most face-hallucination and current face recognition algorithms assume that face images have been aligned accurately with respect to the important facial features, such as the eyes, or both eyes and mouth. This assumption makes these algorithms semi-automatic if high accuracy, in terms of the facial-feature positions, is needed. What's more, many research has found that the performances of these algorithms will degrade dramatically if the face images under consideration are not aligned properly, regardless of whether these

systems are based on local or global/holistic methods [10][11][12][13]. For instance, Wang et al [11] found that, even if the eye-location error is only about 5%, the face recognition accuracy reduces dramatically − by about 50%. Therefore, the accurate detection and localization of facial features play a very important role in face recognition systems.

The remainder of this chapter is organized as follows. First, we provide a brief overview of the state-of-the-art work on facial-feature detection and localization. Then, we introduce our proposed algorithm for accurate facial-feature detection and localization based on a two-level hierarchical scheme. Finally, experiment results are presented and a conclusion is given at the end.

# 2.1 Previous work on facial-feature detection

Although research in different subject areas has paid attention to the issue of accurate detection of facial features, there are still many challenging technical problems to be solved. Most of the existing facial-feature detection and localization algorithms [1] cannot work accurately and reliably when a face is rotated, under a poor lighting condition, of low resolution, or when the eyes are closed. This will, in turn, have an adverse effect on the performance of applications such as face hallucination and face recognition.

There is a general agreement that, among the facial features, the eyes are the most important. Furthermore, the position of the eyes can be used as a reliable geometric reference for localizing the other facial features. Therefore, a large number of methods have been proposed for detecting the eye positions in face images and video clips. The literature on eye-detection and -localization methods can be categorized into two classes: active methods and passive methods. Active methods refer to those that employ a sensing device, while passive methods do not use any sensing device. For active methods, infrared illumination sources are usually used to produce a red-eye effect to detect the eyes [87, 88]. However, the obvious drawbacks of this approach are that intrusive sensors are used, and the subject concerned is lit by infrared light. Therefore, these active methods are utilized in some specific circumstances and conditions only.

Most of the existing approaches to eye detection are passive. In [89], a template-based approach was proposed to detect and locate the eyes in a frontal-view face image. This approach has difficulty when the appearance of the features changes significantly: for example, closed eyes, open mouth, etc. The deformable-template method [90] proposed by Yuille et al. defined parameterized templates for eyes and mouths. By minimizing energy functions defined in terms of the edges, peaks, valleys, image intensities, and prior knowledge, the template parameters are changed iteratively so as to fit the templates to the corresponding facial features in a face image. The Active Shape Model [91] was proposed to handle shape variations in human faces using statistical models. However, the statistical-shape model has to be learned via the manual labeling of landmarks, which are used to represent the facial features in the training face images; the fitting of the model to the face is also done iteratively. Later, an extended Active Shape Model (ASM) [92] was proposed to locate the facial features in frontal views of upright faces. In [93], a corner-detection scheme is first employed to locate possible facial-feature candidates in a head-and-shoulders image, and the approximate positions of the features are estimated by means of average anthropometric measures. Based on the rough positions of the facial features, deformable eye templates can be initialized and used to localize the eyes more accurately. Many other methods have also been devised for eye detection. A multi-layer perception (MLP) eye finder was proposed by Jesorsky et al. in [94]. In [95], a generalized projection function (GPF) was defined, and the hybrid projection function (HPF) was used to determine the optimal parameters of the GPF for eye detection. In [96], an approach for human-face detection and facial-feature extraction was proposed, using a genetic algorithm and the eigenface technique. As the genetic algorithm is, in general, computationally intensive, the searching space is limited to around the eye regions, so that the required runtime can be greatly reduced. Recently, a method using pixel-to-edge information was proposed in [97], which employs the length and the slope information about the closest edge pixels to detect and locate eyes. In [98], an eye-location algorithm based on the radial-symmetry transform was proposed. A symmetry magnitude map is computed using the transform to identify possible eye

candidates. Campadelli et al. [99] proposed a method for eye detection using the support vector machine (SVM) and the Haar wavelet coefficients. In [100], a two-stage hybrid face-detection scheme was proposed using Probability-based Face Mask Pre-Filtering (PFMPF) and the Pixel-based Hierarchical-Feature AdaBoosting (PBHFA) method to effectively solve the complexity in AdaBoosting. Hamouz et al. [101] proposed a method using the Gabor features. A feature-based, affine-invariant method using isophote curvature was proposed by Valenti et al. in [102] for eye localization. A ternary eye-verifier was proposed in [29] for eye verification, which uses skin information and compensation in the color spaces for eye-feature extraction.

As eye localization can be used to approximately identify the other facial features, compared to eye localization, less of the literature has addressed the detection of other facial features such as the nose and mouth. Perlibakas et al. [104] utilized a sequence of bottom-hat morphological operations to locate dark regions which correspond to the eyes and lips; geometrical constraints are then employed to locate the facial-feature candidates. In [105], two thresholds are applied on the image edge map for the extraction of head/face boundaries. Then, projections along the $x$ and $y$ axes are performed on the binary edge image for the detection of the eyes, nose and mouth regions. Cristinacce et. al [106] proposed a multi-stage approach to locate interesting points around the eyes, the mouth, the nose, the eyebrows and the chin. A shape constraint is also used to improve the detection accuracy of the facial features. In [107], a simple transformation is outlined to convert color images to gray-scale images in order to enhance the mouth region during the extraction; the shape of the mouth is then extracted using a binarization method. Asteriadis et al. [108] used the located eye centers to define a mouth region, and then used the hue component in the mouth region to locate the mouth corners. In addition, the mouth candidates in the area were compared to the mean distance vector field. Although the eye positions can be used to coarsely localize the other facial features, different people have faces with different geometric structures. Therefore, it is hard to precisely detect the nose and mouth positions in practice.

An efficient approach to locate the facial features is to identify the salient regions; localization is then focused within those salient regions. A preliminary work [109] uses a salient map to locate pairs of eye regions, and then to identify the best pair of eyes using a fitness function. However, the detection of the nose and mouth has not been investigated. Furthermore, using the fitness function only to identify the best pair of eyes is difficult in terms of locating the eye positions accurately. In this chapter, we aim for a precise localization of the eyes, as well as the nose and mouth. Our proposed method is based on a two-level hierarchical scheme, which has the following advantages:

- Compared to those statistical-learning-based methods [83, 101] (which usually require a lot of training samples and are time-consuming), our proposed method does not need extensive training or learning in advance to locate the eye positions in a face region.

- In contrast to the projection-based methods [95] and the template-based approaches [89, 90], our proposed method can locate accurately the eye regions in face images under illumination variations. This is due to the fact that our proposed scheme is based on multi-scale analysis using wavelet transform, which is more robust and less sensitive to changes caused by illumination variations.

- Since our method computes saliency from three different directional wavelet subbands, the detection is reliable even if the face is tilted, or under different poses, facial expressions, and/or resolutions.

- The proposed two-level hierarchical scheme can efficiently narrow down the search space when finding the accurate eye positions. Therefore, the computational complexity of our proposed method is low: it mainly requires the computation of a salient map only. The computation required is therefore linearly proportional to the image size, and the method is suitable for real-time applications.

## 2.2 Hierarchical structure for eye detection and localization

In this section, a novel scheme is proposed, which uses a two-level hierarchical structure based on a wavelet-based salient map and PCA-based verification to locate eye regions coarsely, and uses a pose-adapted eye template for precise human eye localization. First of all, face regions are located using a cascade of boosted classifiers [110]. In the first level of our algorithm, a wavelet-based saliency map is proposed and applied to the detected face region for a coarse detection of its facial-feature regions. It is observed that the eye regions are always the most varied blocks in a face image. Therefore, those rectangular regions in the saliency map having the largest corresponding sums of saliency values should be the potential candidates for the two eye regions. These eye-region candidates are verified using PCA, so the false-alarm rate can be reduced. Within the eye regions, the accurate locations of the eyes are then determined. The first level of detection can greatly narrow down the search space for finding the actual eye positions, and thus can reduce the computational cost needed. In the second level of our hierarchical scheme, a pose-adapted eye template is constructed using the intrinsic photometric and geometric properties of eyes for accurate eye detection and localization. Based on the detected eye positions, the other facial features (nose and mouth) can be determined by considering their respective sums of the saliency values and by using the two eye positions detected as a reliable geometric reference.

The framework of the proposed novel scheme, based on a two-level hierarchical structure for facial-feature detection and localization, is illustrated in Fig. 2.1. A cascade of boosted classifiers [110] is first used to detect the face regions in an image. The idea of the cascade classifiers is that simple features are used to remove negative candidates in the early stages, while more complicated features are used in the later stages [110]. This method is robust and can be used for real-time face detection. Then, an improved wavelet-based salient map is computed so that salient rectangular regions in the face regions can be determined. Those rectangular

regions with the largest saliency values in the salient map are possible eye-candidate regions. Two candidate regions form an eye-window candidate, which is then verified to be an eye window (or not) by using PCA. Finally, a pose-adapted eye template is used to accurately determine the eye locations. After the eye positions are localized, they are used as a reliable and geometric reference to further locate the approximate positions of the nose and mouth, which can then be detected according to their respective saliency values in the salient map. The following sections of this chapter will present the details of the proposed method.

**2.2.1 Detection of eye-candidate regions using a wavelet-based salient map**

Due to the resemblance between multi-resolution filtering and human visual processing, wavelet-transform techniques have been successfully used to analyze spatial-frequency content [111]. Psychophysical investigation has also shown that the human visual system (HVS) performs a frequency analysis when we see things [112]. Following [111], we denote the detail images (i.e. subbands) as LH (contains the horizontal high-frequency information), HL (contains the vertical high-frequency information), and HH (contains the diagonal high-frequency information). The decomposition/transform also produces one approximation image, denoted as LL, containing the low-frequency information. The wavelet transform can recursively decompose the LL band. Since a two-level wavelet decomposition yields 6 detail images, we use LH1, HL1, HH1, LH2, HL2, HH2, and an additional approximation image, LL2, to denote all the subband images. Fig. 2.2 shows an example of two-level wavelet decomposition.

Wavelet-based saliency detection is an effective approach for describing different parts of a face image, as it can express image variations at different resolutions. An extensive comparison of saliency-detection techniques can be found in [113][114][115][116][117][118]. An orthogonal wavelet transform with a compact support, i.e. its value is zero outside a bounded interval, can lead to a non-redundant and a complete representation of signals. Using a set of orthogonal wavelets, the wavelet coefficient at each signal point is computed at each scale of $2^j$; the wavelet transform can provide information about the local variations in a

signal at different scales. A local absolute maximum value of the wavelet coefficients at a coarse resolution corresponds to a region with high global variation in the signal. In other words, salient points can be detected by finding relevant points at finer resolutions to represent the global variation.

For wavelets with a compact support, a wavelet coefficient can be computed as $W_{2^j} f(n)$ with $2^{-j} p$ signal points, where $W_{2^j}$ and $f(n)$ are the wavelet function and the signal at the scale of $2^j$, respectively, and $p$ is the wavelet regularity [113][114][115][116][117][118]. We can further investigate the wavelet coefficients at the finer scale of $2^{j+1}$. At this scale, a set of coefficients is computed for the same signal points as a coefficient at the scale of $2^j$. We call these coefficients, denoted as $C(W_{2^j} f(n))$, the children of the coefficient $W_{2^j} f(n)$, and they are related as follows:

$$C(W_{2^j} f(n)) = \{W_{2^{j+1}} f(k), 2n \leq k \leq 2n + 2p - 1\}, \tag{2.1}$$

where $0 \leq n \leq 2^j N$, and $N$ is the length of the signal. The children coefficients $C(W_{2^j} f(n))$ reflect the variations of the $2^{-j} p$ signal points, and the most salient point should have the wavelet coefficient with the largest absolute value. The salient points can therefore be detected by considering this maximum, and the corresponding children are then searched. By applying this process recursively, the salient points can be identified. The following formulation has been used to compute saliency values for detecting salient points in [113][114][115][116][117][118]:

$$Sal(n) = \sum_{k=1}^{-j} | C^{(k)}(W_{2^j} f(n)) |, \quad 0 \leq n \leq 2^j N, \quad -\log_2 N \leq j \leq -1. \tag{2.2}$$

However, the largest absolute values of the wavelet coefficients at different scales have different mean values and ranges. The set of maximum wavelet coefficients at the first level have larger magnitudes than those at the second level, and so on. This result actually follows the wavelet transform theory [111]. In order to extract the salient points more accurately, the maximum values at different scales should be normalized. In our algorithm, we utilize an improved saliency detector as follows [119, 120]:
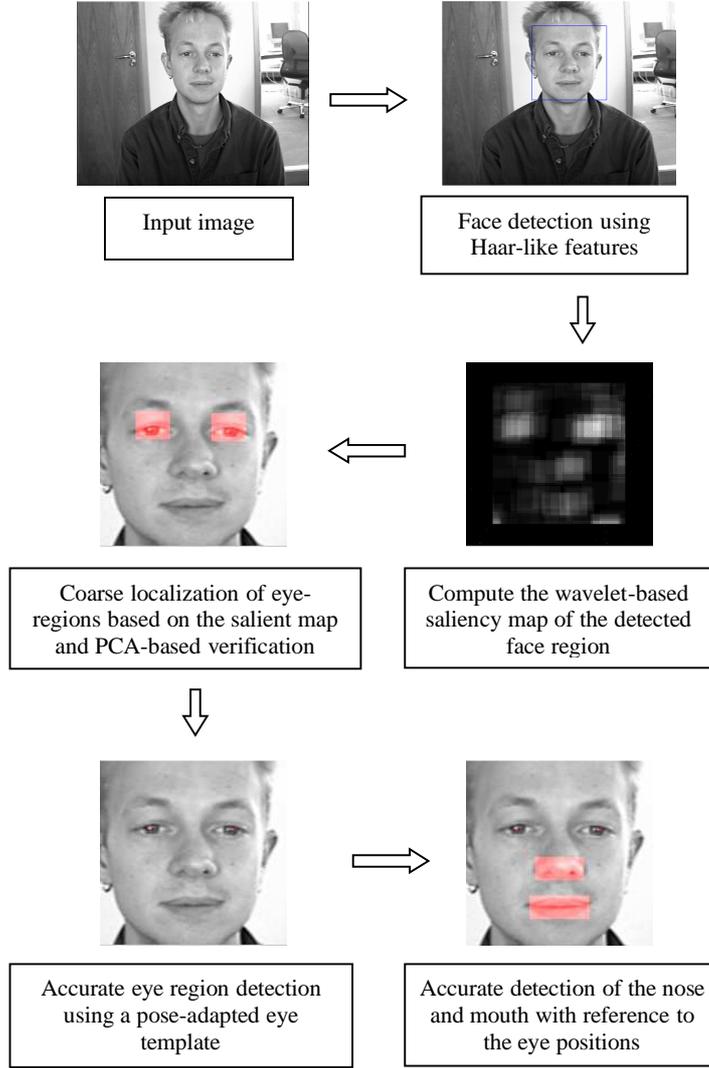
Fig. 2.1. The framework of the proposed facial-feature detection and localization algorithm using a two-level scheme.

$$Sal(n) = \sum_{k=1}^{-j} | \varsigma_k C^{(k)}(W_{2^j} f(n)) |, \quad 0 \le n \le 2^j N, \quad -\log_2 N \le j \le -1, \tag{2.3}$$

where $\varsigma_k$ is the weight to be assigned to the maximum wavelet coefficients at different scales. The weight $\varsigma_k$ is the reciprocal of the standard deviation of the coefficients, which is defined as follows:

$$\mu_k = \frac{1}{S} \sum_{z=1}^{S} | W_{2^k} f(z) |, \tag{2.4}$$
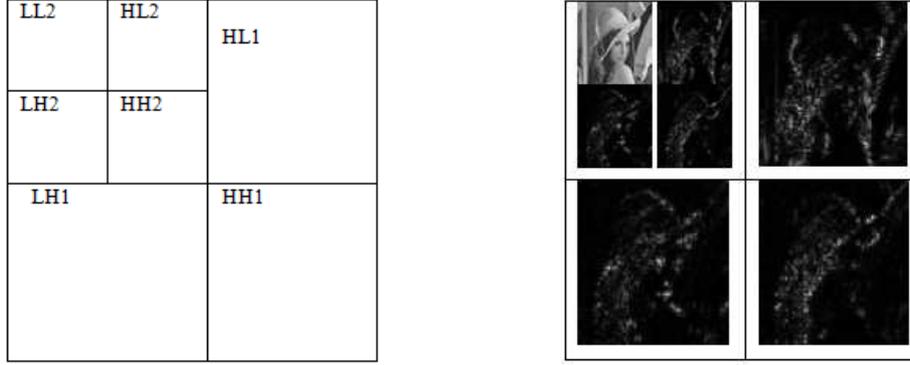
(a) The framework of two-level wavelet decomposition. (b) Two-level wavelet decomposition of the image "Lena"

Fig. 2.2. Two-level wavelet decomposition: (a) the different subbands generated in the decomposition, and (b) the decomposition of the image "Lena", where the approximation image, denoted as *LL*2, contains the low-frequency information; the *LH*1 and *LH*2 subbands contain the horizontal high-frequency information; the *HL*1 *and HL*2 subbands contain the vertical high-frequency information; and the *HH*1 and *HH*2 subbands contain the diagonal high-frequency information.

$$\sigma_k = \sqrt{\frac{1}{S}\sum_{z=1}^{S}(W_{2^k}f(z)-\mu_k)^2} \text{ , and} \tag{2.5}$$

$$\varsigma_k = 1/\sigma_k \text{ ,} \tag{2.6}$$

where $W_{2^k}f(z)$ is an element in the set of maximum coefficients with $0 \le z \le S$, and $S$ is the number of maximum wavelet coefficients at level $k$. In practice, if $K$ salient points are detected in an image, we can set $S=1.5K$. $S$ is the number of the set of maximum wavelet coefficients at each level [119], e.g. $S=60$.

As the two eye regions contain the largest intensity variation in a face image, many salient points can therefore be detected in these regions. Thus, the salient values of the regions are first computed. The two blocks in a face image which have the largest total saliency values are identified as the positions of the eye regions. Define a rectangular region Rect($x$, $y$), whose center coordinates are ($x$, $y$) and whose length and width are denoted as *len* and *wid*, respectively. Consider the wavelet coefficients at the scale of $2^j$ in Rect($x$, $y$), we can further investigate the

region at the finer scale of $2^{j+1}$, where a set of coefficients is computed using the same points at the scale of $2^j$, i.e. the children in Rect($x$, $y$). These children coefficients reflect the variations inside Rect($x$, $y$), and the most salient Rect($x$, $y$) is the one that has the largest sum of absolute wavelet coefficients at the scale of $2^{j+1}$.

The algorithm is applied to the three different subbands, i.e. the horizontal, vertical, and diagonal subbands. If the most salient rectangles of the different subbands lead to the same position, the corresponding sums of saliency values of the three subbands are added to form the saliency values. The saliency at ($x$, $y$) of the resulting salient map of an image is computed by summing the saliency values inside Rect($x$, $y$). If the size of the face images obtained from a face detector is $L{\times}W$, the width and the height of Rect($x$, $y$) are approximately set at $len = 0.15{\times}L$ and $wid = 0.12{\times}W$, respectively; these are set empirically. It is obvious that the eye regions are always the most varied blocks in a face image, and the $T$ rectangular regions with the maximum saliency values are selected as the coarse eye-region candidates for further verification.

## 2.2.2 Verification of eye-region candidates using PCA

The $T$ selected rectangular regions, which possibly contain an eye, can be verified using PCA. Each of these possible eye-region candidates is projected onto an eigen-space, constructed using PCA based on eye-region training samples, and is then reconstructed. The two candidates with the minimum reconstruction errors are selected as the best or the correct eye regions for the left and right eyes. This approach can achieve a better performance for face images in which people wear glasses. The eigen-space verification procedure is described as follows [2, 3]:

An eye window can be considered as a column vector, denoted as $\Gamma$, of dimension $len \times wid$ by concatenating its rows one by one. Let $\Gamma_1$, $\Gamma_2$,…, $\Gamma_M$ be a training set of eye regions. The average of the eye-region training samples is given as follows:

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n .\tag{2.7}$$

The difference of each eye region from the average is computed as $\Phi_n = \Gamma_n - \Psi$. The covariance matrix is calculated as follows:

$$MAT = \frac{1}{M}\sum_{n}^{M}\Phi_n\Phi_n^T = A^T A, \tag{2.8}$$

where the matrix $A = [\Phi_1\Phi_2...\Phi_M]$. In practice, the matrix $MAT$ is large for computing its eigenvectors. A trick [2, 3] can be used which determines the eigenvectors $v_n$ of $A^T A$ first, i.e.

$$A^T A v_n = \lambda_n v_n. \tag{2.9}$$

Multiply both sides by $A$, we have

$$AA^T A v_n = \lambda_n A v_n, \tag{2.10}$$

so $Av_n$ are the eigenvectors of $MAT = AA^T$. Thus, the eigenvector $\mu_n$ is a linear combination of the $M$ training samples:

$$\mu_n = \sum_{k=1}^{M} v_{nk}\Phi_k = A v_n, \; n = 1,......,M. \tag{2.11}$$

Usually, a smaller number of eigenvectors $M'$ is selected, which are those with the largest eigenvalues, so as to reduce the dimensionality of the eigen-space.



(a)

| Eye-candidate regions for the left eye | | | | |
|---|---|---|---|---|
| Corresponding reconstructed images | | | | |
| SSD | 4.07e+005 | 4.16e+005 | 4.09e+005 | 2.16e+005 |
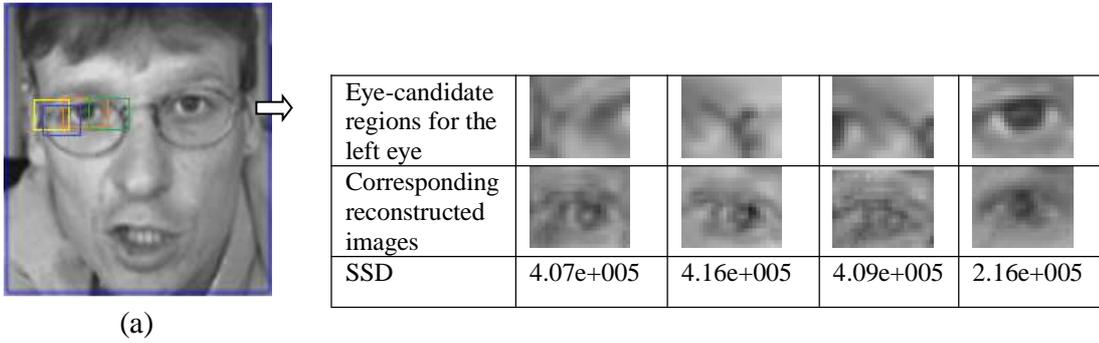
Fig. 2.3. An example of coarse eye-region verification based on PCA: (a) The colored rectangular regions represent the eye-candidate regions for the left eye, (b) the first row shows the original eye-candidate regions for the left eye, the second row shows the corresponding reconstructed images, and the third row shows the sum of squared differences (SSD) between each original eye region and the corresponding reconstructed image.

<div align="center">(a)          (b)          (c)</div>

Fig. 2.4. Examples of coarse eye-region detection results using the improved saliency-detection method with the Db4 wavelet and verification based on PCA: (a) face detection results, (b) the saliency maps, and (c) the detected coarse eye regions using the proposed approach.

A possible eye region, denoted as $\Gamma$, is projected onto the eigen-space as follows:

$$\omega_n = \mu_n^{\ T}(\Gamma - \Psi), \ n = 1,......,M'. \tag{2.12}$$

The weights form a vector $\Omega^T = [\omega_1, \omega_2, ..., \omega_{M'}]$ that describes the contribution of each eigenvector in representing the input eye region. The sum of squared difference (SSD) between the eye region and the corresponding reconstructed image is computed as follows:

$$\varepsilon^2 = \left\| \Gamma - \Gamma_f \right\|^2, \tag{2.13}$$

where $\Gamma_f = \Psi + \sum_{n=1}^{M'} \omega_n \mu_n$. The two eye windows with the minimum reconstruction errors among the $T$ eye-candidate regions will be selected.

Fig. 2.3 shows an example of verification of eye regions based on PCA. The right-most column in Fig. 2.3(b) is the candidate with the minimum reconstruction error (SSD), so it is selected as the optimal coarse eye region. Fig. 2.4 shows some detection results of coarse eye regions using the improved saliency-detection method with the Db4 wavelet and PCA-based verification. It can be seen that our proposed approach is effective with variations in illumination, pose, and facial expression, and with the presence of glasses.

### 2.2.3 Accurate eye localization using a pose-adapted eye template

The sclera, also known as the white of the eye, is the opaque outer layer of the eye. Human eyes are distinctive in the animal kingdom in that the sclera is visible whenever the eye is open. For other mammals, the visible part of the sclera matches the color of the iris, so the white part does not normally show [121]. This special intrinsic property of the human eye can be utilized for accurate eye detection and localization.

Based on the prior knowledge of the sclera's photometric and geometric properties, a pose-adapted eye template is proposed. In practice, human faces always have variations in facial expression and pose. Nevertheless, using the proposed template, eyes can always be localized accurately, except in some extreme conditions. The proposed template has the following characteristics:

● No training or learning is needed.

● It works on both frontal and rotated faces with on-plane rotations of up to 45° (a rotation of 45° means that the face almost touches the shoulder, and this is the rotation limit).

● It can locate the eyes in face images with variations in pose, facial expression, and illumination.

Fig. 2.5(a) illustrates three face images with variations in pose, and Fig. 2.5(b) shows the appearance of the corresponding left eyes. Fig. 2.5(c) shows the devised pose-adapted eye template based on the intrinsic-appearance property of the white of the eye under variations in pose.

The pose-adapted eye template has a black block, denoted as $B$, at its center, with a size of $r \times r$. This black block represents the black pupil in the human eye. There are two white blocks on the left and the right of the black block, which represent the white (sclera) of the eye, and are denoted as $W_1$ and $W_2$, respectively. Their sizes are equal, i.e. $r \times 2r$ (width×height). The height of the white blocks is set at $2r$ in order to deal with face images with the maximum rotation of 45°.



(a)                    (b)          (c)

Fig. 2.5. Face images with variations in pose and the devised pose-adapted eye template: (a) three face images with different poses, (b) the appearance pattern of each of the eyes, and (c) the devised pose-adapted eye template.

Let $I(x, y)$ be the gray-level intensity at position $(x, y)$ of a face region, and $I_{W1}$, $I_B$, and $I_{W2}$ be the summation of the pixel intensities in the regions $W_1$, $B$, and $W_2$, respectively. Specifically, $I_{W1}$, $I_B$, and $I_{W2}$ are defined as follows:

$$I_{W1} = \sum_{(x,y) \in W_1} I(x, y), \tag{2.14}$$

$$I_B = \sum_{(x,y) \in B} I(x, y), \text{ and} \tag{2.15}$$

$$I_{W2} = \sum_{(x,y) \in W_2} I(x, y). \tag{2.16}$$

We define a similarity function, denoted as $SIM$, to measure how well the template represents an eye as follows:

$$SIM = \frac{I_{W1} + I_{W2}}{I_B + \beta}, \tag{2.17}$$

where $\beta$ is a small positive number to avoid the denominator being zero.

Having coarsely detected the eye regions as described in Section 2.2.2, the pose-adapted (or pose-insensitive) eye template is applied to the two eye windows. The

template slides over the detected eye region from left to right and then from top to bottom to determine the best eye position. At each position, the corresponding *SIM* value is computed. Fig. 2.6 shows how the pose-adapted eye template searches over an eye region. To determine the best eye location, the $Q$ template positions with maximum *SIM* values are selected. The best eye position $(x_e, y_e)$ is determined as follows:

$$x_e = median(x_1, x_2, ......x_Q) \text{ and} \tag{2.18}$$

$$y_e = median(y_1, y_2, ......y_Q), \tag{2.19}$$

where $(x_i, y_i)$ are the coordinates of the $i$th candidate of the $Q$ eye candidates.
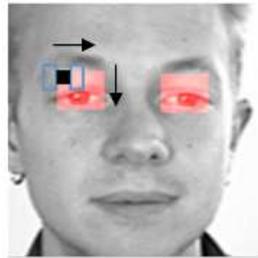


Fig. 2.6. The pose-adapted eye template searches over an eye region from left to right and top to bottom.
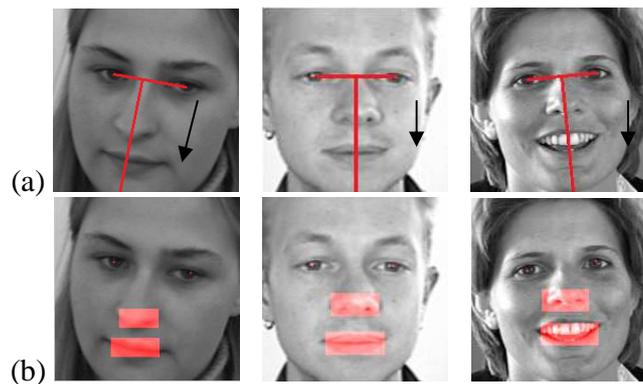


Fig. 2.7. Some examples, with variations in pose and facial expression, of the nose and mouth detection results. The black arrows show the search directions: (a) the nose and mouth are searched along the line perpendicular to the line joining the two eye centers and through the middle of the two eye centers, and (b) the nose and mouth detection results.

### 2.2.4. Nose and mouth detection and localization based on saliency values

Having determined the precise position of the two eyes, the position of the nose and mouth can be located using the fact that their centers are located roughly on the line which is perpendicular to the line joining the two eye centers and through the midpoint of the eye centers, as illustrated in Fig. 2.7(a). When a search is performed along the vertical line, the nose region and the mouth region have the largest variation in intensity, as well as the largest saliency values. Therefore, the two positions where the sum of the saliency values in the rectangles is the largest can be determined. The upper rectangular region is considered to be the nose position, while the lower one is the mouth position.

Fig. 2.7(b) shows some detection results for the nose and mouth in face images with variations in pose and facial expression. We can see that our method can accurately detect the nose and mouth in face images with variations in pose and facial expression.

# 2.3 Experimental results

### 2.3.1 Experiments on eye detection and localization based on the BioID face database

Extensive experiments have been conducted to evaluate the effectiveness of our proposed method. Experiments were conducted on a "difficult" dataset, the BioID face database, to demonstrate the robustness of our method. The BioID face database contains face images under "real-world" conditions, which were captured under uncontrolled illumination, different backgrounds, face sizes, with glasses, etc. The dataset consists of 1,521 gray-scale images with a resolution of 384×286 pixels.

In each of the following experiments, the eye-detection and -localization algorithm is tested using a wide range of face images, and the upper half of a detected face region is considered for eye detection. All of the face images are obtained from a face detector, and the saliency of the detected face regions is computed using the Db4 wavelet with 3 levels of decomposition [111]. Two rectangular regions from five eye-region candidates ($T = 5$) on the left and the right,

respectively, of each face region are selected to form the coarse eye-region candidates. The number of eye-region training samples is set at $M = 15$, and the number of leading eigenvectors used, i.e. those with the highest associated eigenvalues, is set at $M' = 10$. The pose-adapted eye template has the parameter $r = 7$, and the number of candidates for an eye region is experimentally set at $Q = 7$, which can produce a good and robust result by experiments. The detection accuracy is measured as follows [94]:

$$e = \max(d_l, d_r)/d_{lr} ,\tag{2.20}$$

where $d_l$ and $d_r$ are the Euclidean distances between the located positions and the corresponding ground-true positions of the left and the right eyes, respectively, and $d_{lr}$ is the Euclidean distance between the two eyes of the ground-truth. In the experiments, $e \leq 0.25$ and $e \leq 0.1$ are used as the criteria for measuring the accuracy. If a detected eye position has its $e$ value smaller than a certain threshold, the detection will be considered correct.

Fig. 2.8 shows some facial-feature detection results using our proposed method for faces under pose variations, different facial expressions, and wearing glasses. The blue rectangles represent the detected face regions, and the detected eyes are marked with a red dot. From the results, we see that the proposed method can detect eyes accurately even if the faces are under different illuminations and poses, or wearing glasses. As the saliency is computed based on the summation from three subbands, the detection is insensitive to face orientations. Furthermore, the proposed pose-adapted eye template based on the intrinsic-appearance property of eyes produces precise localization.

We also compare the detection accuracy of our proposed algorithm with six state-of-the-art eye-detection approaches using the BioID database. The six methods are:

(1) multi-layer perceptron (MLP) eye finder proposed by Jesorsky et al. [94],

(2) method using pixel-to-edge information proposed by Asteriadis et al. [97],

(3) enhanced method based on the radial-symmetry transform proposed by Bai et al. [98],

(4) method using the SVM and Haar wavelet coefficients proposed by Campadelli et al. [99],

(5) method using Gabor features proposed by Hamouz et al. [101], and

(6) method using isophote curvature proposed by Valenti et al. [102].



|     (a)     |     (b)     |     (c)     |

Fig. 2.8. Examples of coarse eye-region detection results using the improved saliency-detection method with the Db4 wavelet and verification based on PCA: (a) face detection results, (b) the saliency maps, and (c) the detected coarse eye regions using the proposed approach.

Table 2.1 tabulates the normalized errors for the different algorithms, with $e \leq 0.25$ and $e \leq 0.1$ to assess the detection accuracy. We see that our method can achieve a superior accuracy to that of the state-of-the-art methods. We also compare our method with the other detection and localization methods in terms of the run-times required, as tabulated in Table 2.2. It can be seen that the runtime required by our method is similar to that of the other methods.

Table 2.1. Detection accuracy of different eye-detection and -localization methods using the BioID Face Database.

| Algorithms: | Accuracy ($e \leq 0.1$) | Accuracy ($e \leq 0.25$) |
|---|---|---|
| Jesorsky [94] | 78.07% | 90.23% |
| Asteriadis [97] | 79.60% | 94.46% |
| Bai [98] | 65.08% | 95.54% |
| Campadelli [99] | 83.12% | 95.80% |
| Hamouz [101] | 82.33% | 95.40% |
| Valenti [102] | 87.45% | 96.69% |
| Our method | **90.56%** | **98.62%** |

Table 2.2. The run-times of different eye-detection and -localization methods using the BioID Face Database.

| Algorithms: | Run-times (second per image) |
|---|---|
| Jesorsky [94] | 0.45 |
| Asteriadis [97] | 0.47 |
| Bai [98] | 0.56 |
| Campadelli [99] | 0.52 |
| Hamouz [101] | 0.54 |
| Valenti [102] | 0.67 |
| Our method | **0.62** |

We notice that failed detection mainly occurs in those face images with eyeglasses, especially with specularities around them, as shown in Fig. 2.9(a). The saliency values around the eyeglasses' frame increase due to its edges, corners, highlights and specularities. This also leads to incorrect detection in some cases, as shown in Fig. 2.9(b). Nevertheless, an automatic eyeglasses-removal algorithm can be employed to solve this problem [122]. In this experiment, face images with eyeglasses are pre-processed using the eyeglasses-removal algorithm. Fig. 2.9(c) shows the results for our method using the eyeglasses-removal algorithm.

For the issue whether the eyeglasses removal algorithm is decided to apply automatically, as shown in Fig. 2.8(b), the saliency maps of faces with eyeglasses are more "complex" than those saliency maps of faces without eyeglasses. The mean value of the saliency map of face with eyeglasses is significantly large than

the face without eyeglasses, therefore, if the mean value of the saliency map is significantly large than a threshold $T_h$, which is set emprically, it will apply the eyeglasses removal algorithm automatically, vice versa.



(a)　　　　　　　(b)　　　　　　　(c)
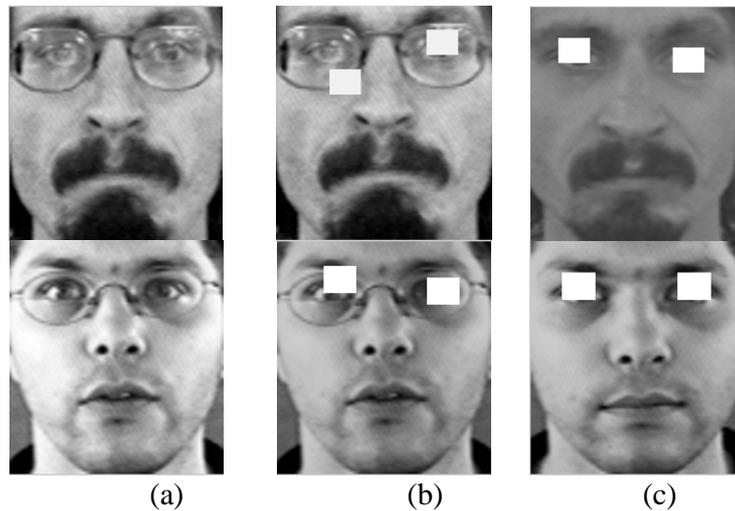
Fig. 2.9. Results of eye detection and localization using our proposed scheme: (a) original image with eyeglasses, highlights and specularities, (b) failed detection without performing eyeglasses removal, and (c) detection results with eyeglasses-removal performed.



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

Fig. 2.10. Face images with out-of-plane rotations [130].

(a), (b), and (c) face image with left out-of-plane rotation of 10, 20, and 30 degree, respectively.

(d) and (e) face image with left out-of-plane rotation of 45 degree, and with right out-of-plane rotation of 45 degree, respectively.

(f) and (g) face image with left out-of-plane rotation of 90 degree, and with right out-of-plane rotation of 90 degree, respectively.

In addition, for out-of-plane rotation cases, as shown in Fig. 2.10, our proposed method work well on the face images with rotation degrees within 30 degree, i.e. Figs. 2.10(a), 2.10(b) and 2.10(c). However, for face images with rotation degrees larger than 30 degree, such as shown in Figs. 2.10(d), 2.10(e) and Figs. 2.10(f), 2.10(g), it is hard to accurate localization for both existing approaches and our proposed method. The issue on out-of-plane rotation with significant degree (e.g. larger than 45 degree) is very challenging, which will be our future work.

**2.3.2. Experiments on mouth and nose detection and localization**

Experiments were also performed to evaluate the effectiveness of the proposed method for mouth and nose detection and localization. All of the experiments empirically set the size of the rectangle for nose-region detection at $nlen = 0.25 \times L$ and $nwid = 0.12 \times W$, and the size of the rectangle for mouth-region detection at $mlen = 0.30 \times L$ and $mwid = 0.12 \times W$, respectively. Figs. 2.7(b) and 2.7(d) show the nose and mouth facial-feature detection results using the proposed method. We see that the proposed scheme can effectively localize the mouth and nose regions.

There are some variations between the mouths: some face images have a wide-open mouth, while others have an almost-closed mouth; and males usually have a larger mouth than females. However, the distance between the different people's eyes is usually similar. In other words, the variation in the size of the mouth (mouth size and shape vary between people) is larger than that of the distance between the two eyes. Therefore, we still use the distance between the two eyes $d_{lr}$ of the ground-truth to normalize the errors so as to form the measure for the detection accuracy of the mouth and nose. In order to objectively assess the detection performance, we define the detection accuracy for the mouth and nose as follows:

$$e_{nose} = d_n / d_{lr} \text{ and} \tag{2.21}$$

$$e_{mou} = d_m / d_{lr} , \tag{2.22}$$

where $d_n$ and $d_m$ are the Euclidean distances between the located positions and the corresponding ground-true positions of the nose and mouth, respectively. In the

experiments, $e_{mou} \leq 0.25$ and $e_{mou} \leq 0.1$ are used as the criteria to assess the detection accuracy.

We have compared the mouth-detection accuracy of our proposed algorithm with five state-of-the-art mouth-detection approaches. The five methods are:

(1) the method in [104] based on morphological operations and geometrical constraints;

(2) the method in [105] using the image edge map and projections for the detection of nose and mouth regions;

(3) a multi-stage approach to locate interesting points around mouths and noses [106];

(4) the binarization method in [107] to detect mouth regions; and

(5) the approach in [108] using a mean template for mouth-corner localization.

In addition, the methods in [105] and [106] were also compared to our proposed method in terms of nose-detection accuracy. Table 2.3 tabulates the mouth-detection accuracy of all the methods. Our detection rate with $e_{mou} \leq 0.25$ is 92.19%. By comparison, the detection rates of [104] and [108] are 83.70% and 85.33%, respectively. Experiments show that our proposed method can achieve a superior performance for mouth detection.

Table 2.4 tabulates the nose-detection accuracy. Our saliency-region-based scheme produces satisfactory results, with a detection rate of 95.65% when $e_{nose} \leq 0.25$, while the detection rates of [104] and [106] are 83.38% and 87.40%, respectively. Extensive experiments have verified the effectiveness of our proposed scheme for locating noses.

Table 2.3. Mouth-detection rates of the different methods based on the BioID Face Database.

| **Algorithms**: | **Accuracy** ($e \leq 0.1$) | **Accuracy** ($e \leq 0.25$) |
|---|---|---|
| Perlibakas [104] | 81.66% | 83.70% |
| Shih [105] | 75.25% | 81.47% |
| Cristinacce [106] | 81.50% | 83.11% |
| Travieso[107] | 66.20% | 69.80% |
| Asteriadis [108] | 83.37% | 85.33% |
| Our method | **90.24%** | **92.19%** |

Table 2.4. Nose-detection rates of the different methods based on the BioID Face Database.

| Algorithms: | Accuracy ($e \leq 0.1$) | Accuracy ($e \leq 0.25$) |
|---|---|---|
| Shih [105] | 80.25% | 83.38% |
| Cristinacce [106] | 84.01% | 87.40% |
| Our method | **93.78%** | **95.65%** |

## 2.3.3 Experiments on eye detection and localization based on the Yale Face Databases

In order to test the effectiveness of the proposed method under illumination variations, experiments were also carried out using the Yale Face Database B [72] and the extended Yale Face Database B [73], which have been commonly used to evaluate the performances of illumination-invariant face recognition methods. The Yale Face Database B consists of 10 classes, named from yaleB01 to yaleB10. The extended Yale Face Database B contains 28 human subjects, named from yaleB11 to yaleB13, and from yaleB15 to yaleB39. The total number of distinct subjects in the two databases is 38. The face images named **_P00A-035E-20 and **_P00A+035E-20 ("**" represents yaleB01 − yaleB13 and yaleB15 − yaleB39) were selected in the experiment, i.e. 38×2=76 images in total.

Table 2.5. Detection accuracy of different eye-detection and -localization methods using the Yale Face Database.

| Algorithms: | Accuracy ($e \leq 0.25$) |
|---|---|
| Jesorsky [94] | 80.25% |
| Asteriadis [97] | 82.07% |
| Bai [98] | 81.41% |
| Campadelli [99] | 84.50% |
| Hamouz [101] | 81.83% |
| Valenti [102] | 83.75% |
| Our method | **93.42%** |

Table 2.5 tabulates the normalized errors for the different algorithms, with $e \leq$ 0.25. Although the face images used in this experiment have their illumination conditions changed dramatically, experiment results show that 71 out of the 76 face images (i.e. 93.42%) have the eye positions detected accurately, with $e \leq 0.25$. We can see that the detection accuracy of our method is significantly higher than that of other methods.

# 2.4 Conclusion

In this chapter, we have proposed an efficient algorithm for the accurate detection of facial features in face images. Our method has the advantages of being non-iterative and computationally simple. The proposed algorithm can locate the eye features precisely. This is due to the fact that the computation of saliency values considers saliency from three different directional wavelet subbands, so the proposed scheme is robust to face orientation and to variations caused by different poses, facial expressions, and resolutions. With the use of saliency detection and the verification of eye candidates using PCA, the required computation can be greatly reduced because the eye template is only applied to the selected coarse eye regions. In addition, the proposed eye template can handle eye regions with a large rotation and pose variation. The position of the nose and mouth can be determined by considering both the saliency values in the salient map and the detected eye positions as geometric references. Experiment results show that our method can achieve a higher detection-accuracy level than existing state-of-the-art algorithms.

# Chapter 3 Illumination-Invariant Face Recognition Using Illumination Compensation and Enhancement

Face images exhibit large variations in appearance under different illuminations. As shown in Figure 3.1, the different face images of the same person with variations in illuminations are not discernible as the same man. From Figure 3.1 we can see that human faces share a similar shape and structure, but illumination variations and different lighting directions always make images of the same person look dissimilar.



Figure 3.1. Images under different illumination directions: face images of the same person from the YaleB face database.

With the rapid development of the devices of digital image capturing and researches of complicated recognition algorithms, face recognition can achieve a highly accurate performance under controlled conditions, such as unchanged light sources, frontal-view images, no occlusion, neutral facial expression, etc. However, face recognition with different illuminations is a difficult problem, in particular in outdoor circumstances. Illumination variations remain an unsolved problem in face recognition, despite a lot of research having been devoted to solving it [1].

In the past decades, the illumination problem has received considerable attention in both the face-recognition-related industries and academic circles. However, it is still one of the prominent issues for appearance- or image-based face recognition approaches. The development of illumination-compensation techniques for face recognition is important, and modeling face variations in realistic settings is still a heuristic issue, especially in uncontrolled environments such as outdoor and natural

settings. Without solving this problem, accurate and robust face recognition cannot be achieved [1, 9].

In this chapter, we will focus on the problems of illumination variations for face images, and proposed a novel scheme for Illumination-Invariant Face Recognition Using Illumination Compensation and Enhancement. The rest of this chapter is organized as follows. First, we give a brief overview of related state-of-the-art work. Then, we describe our proposed methods for illumination compensation and enhancement, and experiment results are also presented. At the end, a conclusion and discussion for this chapter is given.

# 3.1 Related Work

Face recognition with different illumination conditions is a difficult issue, and currently illumination variations remain an unsolved problem, despite a lot of research having been devoted to solving it [1]. In the past several decades, the illumination problem has received considerable attention, yet it is still one of the most prominent issues for appearance- or image-based recognition approaches. The first such method used a number of images with different illuminations to extract the three-dimensional shape for illumination-invariant representation based on Lambertian reflectance [42][43][44][45]. In [46], the authors found that the ratio of two images from the same object is simpler than the ratio of images from different objects with Lambertian reflectance, and the ratio provides two of the three distinct values in the Hessian matrix of the object's surface. A method based on quotient images [47] was introduced, which assumes − based on the Lambertian model − that faces of the same class have the same shape but different textures. Recently, Chen et al. [39] improved self quotient image (SQI) by using Logarithmic Total Variation (LTV) smoothing and an efficient illumination-normalization method was proposed using an illumination model with a Lambertian surface for face recognition in [48]. Later, photometric stereo has been used to obtain a fast and non-contact surface reconstruction of Lambertian surfaces in [49], while in [50], the 3D face-reconstruction methods assume that human faces can be modeled as

Lambertian, and show that human skin exhibits nearly-Lambertian reflectance properties.

The illumination problem in face recognition has drawn many researchers' attention, besides the approaches based on Lambertian model, and many other methods have been also proposed. According to the Retinex theory [51], the intensities of an image $I(x, y)$ can be represented as the product of illumination $L(x, y)$ and surface reflectance $R(x, y)$. Based on this theory, an automatic image-processing algorithm for compensating illumination-induced variations was proposed in [52], which estimates the illumination field and then compensates for it. However, this method is subject to artifacts. Shan and Ward [53] proposed a wavelet-based normalization method, which enhances the contrast as well as the edges of face images for illumination normalization in order to facilitate face-recognition tasks. In [54], a facial-image illumination-invariant algorithm, based on the fusion of wavelet analysis and the local binary pattern, was introduced. In the same year, a simple algorithm which can alleviate illumination effects by setting the coefficients in the wavelet approximation sub-band to zero was proposed in [55]. Moreover, In [9], Adini et al. presented an empirical and systematic study, and evaluated the sensitivity of some representations to changes in illumination. Three different categories of approaches were discussed. The first method used gray-level information to extract the three-dimensional shape of an object, using the shape-from-shading approach [56]. This is an ill-posed problem, and the assumptions used make it difficult to apply to general object recognition. Therefore, this approach is not effective for face recognition. The second approach used image representations that are relatively insensitive to illumination changes, such as the edge maps of images [57][58][59] and a basic image-representation model for face recognition [60][61]. The third approach to solve the illumination-variation problem was to model several images of the same face taken under different illumination conditions [62][63]. More recently, a 3D morphable face model was employed to produce synthetic images under varying poses and illuminations. Frontal, semi-profile, and profile face images of the same person are used to generate 3D face models in [64]. Zhao and Chellappa [65] proposed a method using

symmetric shape-from-shading for Illumination-insensitive face recognition. In [66], Zhao et al. used illumination ratio images to produce new training images for face recognition with a single frontal-view image. Xie and Lam [67] proposed a 2D face-shape model to eliminate the effect of difference in the face shape of different individuals for face recognition. Tan and Triggs proposed an illumination normalization method, which incorporates three main stages designed to counter the effects of illumination variations, local shadowing, and highlights while preserving the essential elements of visual appearance [79]. The three main stages contain a nonlinear Gamma correction, Difference of Gaussian (DoG) filtering and a final stage of contrast equalization for face recognition under difficult lighting conditions.

### 3.1.1 Discussion of the disadvantages of the existing techniques

However, those existing techniques methods [42][43][44][45][49][50] based on the Lambertian model share the same drawbacks; a single-point light source placed at infinity is assumed. The Lambertian model is suitable for some applications, but it has proven difficult to build accurate 3-D models using only images taken in uncontrolled circumstances, and the assumptions make it difficult to apply to general object recognition in outdoors environments [46]. Furthermore, although traditional illumination-invariant face-recognition methods, which based on the Lambertian model, can construct a three-dimensional face representation by combining linearly a number of the images under different illuminations, there are two obvious drawbacks with the Lambertian model: a single-point light source placed at infinity is assumed, and multiple images need to be captured under a variety of illumination conditions for each class to obtain a three-dimensional representation of the images. Furthermore, most existing approaches to the illumination problem rely primarily on universal representations, which are in general insufficient to model the variations caused by illumination changes [9]. It has been shown theoretically that an illumination-invariant image representation or function does not exist [9][68]. Solving image variations caused by changes in illumination direction can be achieved by utilizing more domain-specific knowledge. In [69], a simple scheme based on the wavelet transform was proposed

for illumination compensation and enhancement for face images; however, it has the disadvantage that a face image within a class under even and frontal illuminations is needed in order to calculate the average representations of face images under changing illuminations.

In contrast to previous work and instead of deriving universal representations, illumination compensation and enhancement utilizing specific individual information can possibly provide an effective and useful way to achieve a better appearance and a higher recognition rate. In this chapter, we present a computationally efficient method for generating illumination-invariant texture and face images using the scheme of illumination compensation and enhancement. We use an illumination model which is universal and does not require the assumption of a single-point light source, thereby overcoming the limitation of the Lambertian model. Our proposed approach captures the mean illumination-effect matrix representations of images under a variety of different illumination conditions for each class, so as to compensate or enhance the images, and consequently, to achieve better classification performance. In particular, we will aim at devising a simple and effective scheme to compensate/enhance illumination, and to remove the shadow caused by uneven illumination, rather than obtaining a sophisticated and accurate representation of the texture surface reflection.

# 3.2 Illumination Compensation and Enhancement for Illumination-Invariant Face Recognition

### 3.2.1 An illumination model

Some methods have been proposed to handle varied illuminations based on the Lambertian model, with the assumption that a single illuminant is placed at infinity, and the utilization of a number of images, to construct 3D models that are invariant to illumination. In real situations, images are usually captured in outdoor, uncontrolled environments, with various illumination sources from different

directions. To overcome the limitations of the Lambertian model, the illumination model should be universal (can be used in multi-lighting circumstances), without requiring the assumption of a single-point light source.

According to the Retinex theory [70], the intensity of an image $I(x, y)$ can be represented as the product of illumination $s$ and surface reflectance $R(x, y)$. Based on this theory, in contrast to the previous work, a novel and effective scheme is proposed in this chapter for illumination compensation and enhancement, which is efficient and does not require an image under even and frontal illumination to learn or to be the reference image. Thus, our proposed algorithm is easy to implement. The intensity of an image $I(x, y)$ is expressed as follows:

$$I(x, y) = R(x, y)L(x, y), \tag{3.1}$$

where $R(x, y)$ is the surface-reflectance-representation matrix and $L(x, y)$ is the illumination-effect matrix.

Such a illumination model-decomposition could be advantageous for many computer vision algorithms. However, estimating this decomposition is a fundamentally ill-posed problem because every observed value there are multiple unknowns [70]. In this chapter, we employ a mathematical framework that solve the ill-posed problem and can be used to extract image representations for relighting. The framework is essentially based on the Singular Value Decomposition (SVD) representation of images, which are images under multiple and different illumination directions. The illumination model in (3.1) is nonlinear. Hence, the logarithmic transformation is applied so as to convert (3.1) into a linear model, as follows:

$$\begin{aligned} I_l(x, y) &= \log(I(x, y) + \beta) \\ &= \log(R(x, y)L(x, y) + \beta) \approx \log(R(x, y)) + \log(L(x, y)) \\ &= R'(x, y) + L'(x, y) \end{aligned} \tag{3.2}$$

where $\beta$ is a small positive integer.

First, after the logarithmic transformation an image $I(x, y)$ into $I_l(x, y)$, our proposed framework decomposes the image $I_l(x, y)$ of size $m \times n$ into the

eigenspace using SVD. SVD is commonly used in matrix analysis and can be applied to analyze an image matrix based on the following theorem of linear algebra:

The image $I_l(x, y)$ can be viewed as a matrix with $m$ rows and $n$ columns, and any $I_l(x, y)$ matrix, whose number of rows $m$ is greater than or equal to its number of columns $n$, can be written as the product of an $m \times m$ column-orthogonal matrix $U$, a $m \times n$ diagonal matrix $W$ with positive or zero elements, and the transpose of an $n \times n$ orthogonal matrix $V$. That is,

$$I_l(x, y) = UWV^T, \tag{3.3}$$

where $U^T U = V^T V = E$ and $E$ is the unit matrix. The matrix $U$ is a $m \times n$ column-orthogonal matrix, while $V$ is a $n \times n$ orthogonal matrix. The elements $w_i$ on the diagonal of $W$ are called singular values (the square root of the eigenvalues), i.e.

$$W = diag(w_1, \quad w_2, \quad ...., \quad w_i, \quad ...., \quad w_n). \tag{3.4}$$

The singular value vector $s$ of the image $I_l(x, y)$ is defined as follows:

$$s = [w_1, w_2, ...., w_i .... , w_n]^T, \tag{3.5}$$

where $1 \leq i \leq n$, and $w_i$ is the $i^{\text{th}}$ singular value of the image $I_l(x, y)$ in the singular value vector $s$ such that $w_i \geq w_{i+1}$. It can be observed that the singular values decrease dramatically and the mathematical framework of SVD can be used to represent texture images effectively [44, 45]. Particularly, Pentland et al. [71] have shown that the first three eigenvectors represent illuminations on face images, and have also empirically shown that a superior face recognition performance can be achieved if the first three eigenvectors are excluded. In general, the first $k$ major eigenvectors mainly reflect variation of illuminations.

Let:

$$W = diag(w_1, \quad w_2, \quad ...., \quad w_i, \quad ...., \quad w_n)$$

$$= \begin{pmatrix} w_1 & & & & & \\ & w_2 & & & & \\ & & .... & & & \\ & & & w_i & & \\ & & & & .... & \\ & & & & & w_n \end{pmatrix}$$

$$= diag(w_1, \quad ...., \quad w_k, \quad 0, \quad ...., \quad 0) + diag(0, \quad ...., \quad 0, \quad w_{k+1}, \quad ...., \quad w_n) \tag{3.6}$$

$$= \begin{pmatrix} w_1 & & & & \\ & .... & & & \\ & & w_k & & \\ & & & 0 & \\ & & & & .... \\ & & & & & 0 \end{pmatrix} + \begin{pmatrix} 0 & & & & \\ & .... & & & \\ & & 0 & & \\ & & & w_{k+1} & \\ & & & & .... \\ & & & & & w_n \end{pmatrix}$$

$$= W_1 + W_2$$

Then, (3.2) can be written as follows:

$$I_l(x, y) = UWV^T = U(W_1 + W_2)V^T = UW_1V^T + UW_2V^T$$

$$= L'(x, y) + R'(x, y) \tag{3.7}$$

where, suppose $UW_1V^T = L'(x, y)$ for the first $k$ major eigenvectors mainly reflect variation of illuminations, and the other residual component $UW_2V^T = R'(x, y)$ is the illumination-effect matrix.

This expression is similar to the formulation of the illumination model in (3.2). Specifically, $UW_2V^T$ can be treated as the component of the surface-reflectance-representation matrix $R'(x, y)$, while $UW_1V^T$ can be seen to be the component of the illumination-effect matrix $L'(x, y)$ in the illumination model. Now, we can see that an image represented in matrix form can be described using the illumination model in (3.1) and SVD in (3.7).

Now, we can see that an image represented in matrix form can be described using the illumination model in (3.1) and SVD in (3.2). Figure 3.2 shows an example of image decomposition based on using the illumination model and SVD with different $k$. The next step is how to select the value $k$ to obtain an optimal image decomposition representation, and so as to solve the ill-posed problem.

### 3.2.2 Surface-reflectance-representation matrix in the illumination model

What make the face images of one person with the same surface reflectance structure look dissimilar, as illustrated in Figure 3.1? Face images of the same person have identical facial structures and patterns, sharing a similar shape surface reflectance structure, so it is reasonable to assume that the surface-reflectance-representation matrix $R(x, y)$ is a slowly-changing matrix, which reflects the intrinsic property of a face surface. Consequently, the dissimilarity between images of the same person under different illumination conditions is mainly caused by the differences in the illumination-effect matrix $L(x, y)$. That is to say, assume that there are $M$ face images of the same person; the differences between the components of the surface-reflectance-representation matrix $R'(x, y)$ of the $M$ face images are small. The following root mean squared value (RMS) can be used to measure the differences between the components of the surface-reflectance-representation matrix $R'(x, y)$ of the $M$ face images with different $k$.

$$RMS^k = \frac{\dfrac{1}{mn}\sqrt{\displaystyle\sum_{\substack{1\le a\le M \\ 1\le b\le M \\ a\ne b}}\sum_{\substack{1\le x\le m \\ 1\le y\le n}}(R'_a(x,y)-R'_b(x,y))^2}}{\sigma_k} \; , \tag{3.8}$$

where $\sigma_k$ is the standard deviation of the component of the surface-reflectance-representation matrix $R'_a(x,y)$ $(1\le a\le M)$, $m$ and $n$ are the numbers of rows and columns of the images $I(x, y)$. $R'_a(x,y)$ and $R'_b(x,y)$ represent the component of the surface-reflectance-representation matrix of the original images $I_a(x, y)$ and $I_b(x, y)$ $(1\le a,b\le M)$, respectively, i. e.
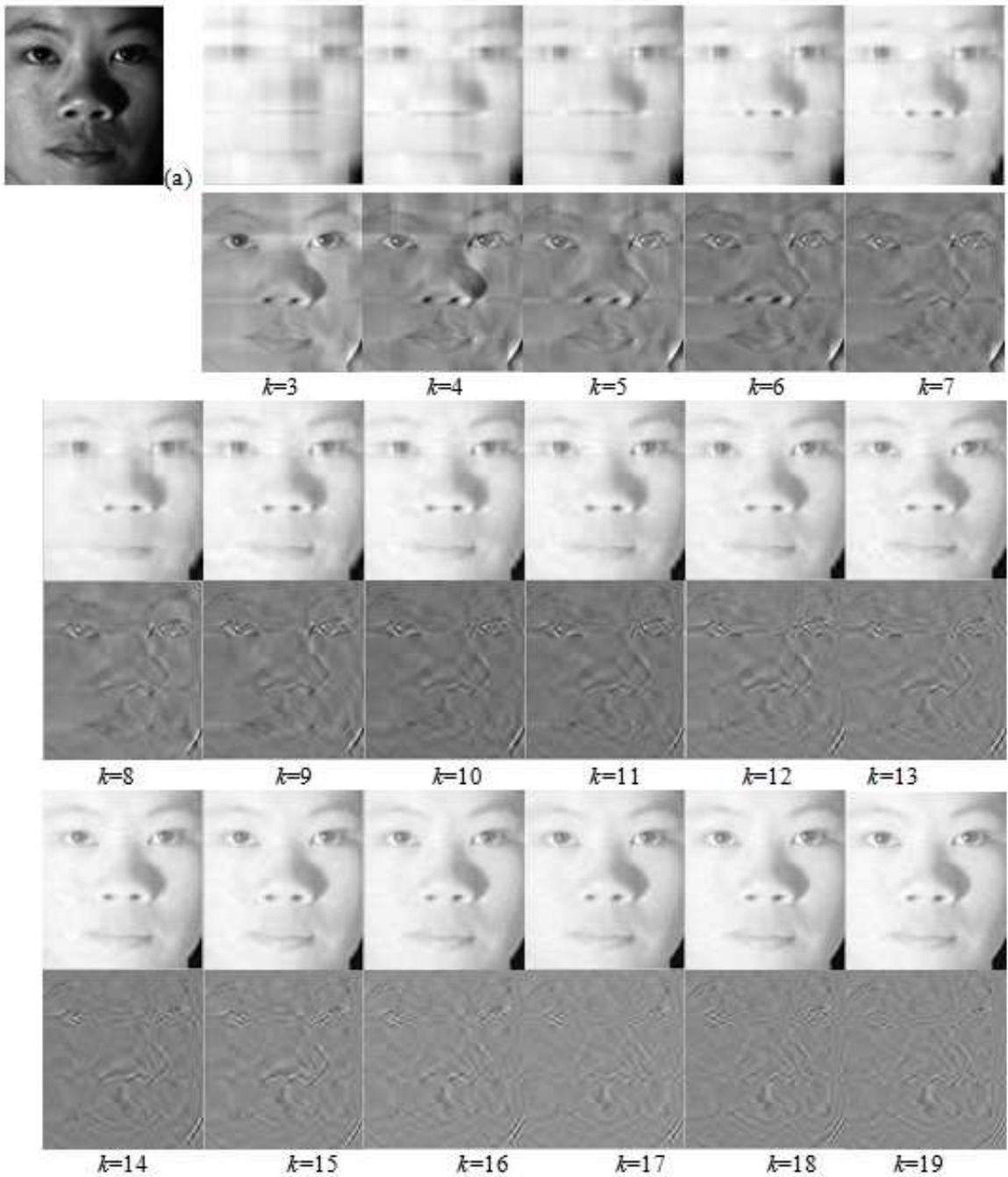
Figure 3.2. An example of image decomposition based on the illumination model and SVD with different $k$. (a) Input image. Odd rows are the component of the illumination-effect matrix $L'(x, y)$ in the illumination model, and even rows are the component of the surface-reflectance-representation matrix $R'(x, y)$.

$$R_a'(x,y) = U_a W_{a,2} V_a^T = U_a \begin{pmatrix} 0 & & & & & \\ & .... & & & & \\ & & 0 & & & \\ & & & w_{a,k+1} & & \\ & & & & .... & \\ & & & & & w_{a,n} \end{pmatrix} V_a^T ,$$ and

$$R_b'(x,y) = U_b W_{b,2} V_b^T = U_b \begin{pmatrix} 0 & & & & & \\ & .... & & & & \\ & & 0 & & & \\ & & & w_{b,k+1} & & \\ & & & & .... & \\ & & & & & w_{b,n} \end{pmatrix} V_b^T .$$

Suppose that there are $N$ persons with $M$ face images of the same person in the training set. Every person with $M$ face images in the training set are transformed in the same way, using equation (3.8). The average root mean squared value (ARMS) can be used to measure the overall differences between the $N$ persons in the training set to determine the value $k$:

$$ARMS^k = \frac{\sum_{1 \le j \le N} RMS_j^k}{N} \tag{3.9}$$

The scope of the value $k$ is set empirically in the range $2 \le k \le 19$, because, as shown in Figure 3.2, when the value $k$ becomes larger than 19, the energy of the component of the surface-reflectance-representation matrix $R'(x,y)$ becomes nearer to zero. The optimal $k$ can be selected such that the difference between the components of the surface-reflectance-representation matrix $R'(x,y)$ is the smallest:

$$k = min\{ARMS^k, 2 \le k \le 19\} \tag{3.10}$$

The global optimal $k$ can be determined using the training set (i.e. $k = 9$ in our experiment. An optimal $k$ can produce the best results, but the performance is not very sensitive to $k$'s value. Actually, our experiments have shown that satisfactory illumination-compensation and -enhancement performances can still be achieved, even though the value of $k$ is ±2 of the optimal value.). Figure 3.3 illustrates the decomposition of the intensity of an image $I(x,y)$ into the component of the

surface-reflectance-representation matrix $R'(x, y)$ and the component of the illumination-effect matrix $L'(x, y)$ in the illumination model.



(a)           (b)           (c)
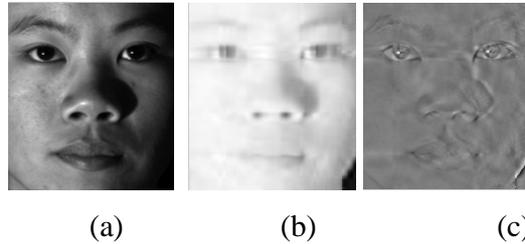
Figure 3.3. The decomposition of the image into the two component of the illumination model. (a) The input intensity image. (b) the component of the illumination-effect matrix $L'(x, y)$ in the illumination model, and (c) the component of the surface-reflectance-representation matrix $R'(x, y)$ in the illumination model.

### 3.2.3 Illumination compensation algorithm (ICA)

Humans have similar face structures and shapes, but the face images of the same person do not look similar under different lighting conditions. Thus, it is reasonable to infer that the component of the reflectance-representation matrix $R'(x, y)$ of faces with a similar shape and structure has a slight difference, while the component of the illuminations-effect matrix $L'(x, y)$ can vary significantly, depending on the illumination conditions. This is due to the fact that images under uneven illumination conditions produce shadows, and look different in those regions with insufficient illumination. If we can learn a mean illumination-effect matrix $\overline{L}(x, y)$ so as to compensate the component of the illumination-effect matrix $L'(x, y)$ of the images with uneven lighting and shadows, it will make images lighter and shadowless. Figure 3.4 illustrates the process of our algorithm for illumination compensation and enhancement.

(a)

(b)

| (a) An image under uneven illumination. | (b) $Q$ face images similar to (a) are searched. |

Compute the illumination-effect matrix $L_c$ of (a) ⟺ Compute the learned mean illumination-effect matrix $\overline{L}$

Illumination-invariant process

Illumination-compensation algorithm (ICA): $L_c + \overline{L}$

Illumination-enhancement algorithm (IEA) : $L_c + \lambda\overline{L}$ ($\lambda > 1$)

(c)

(d)
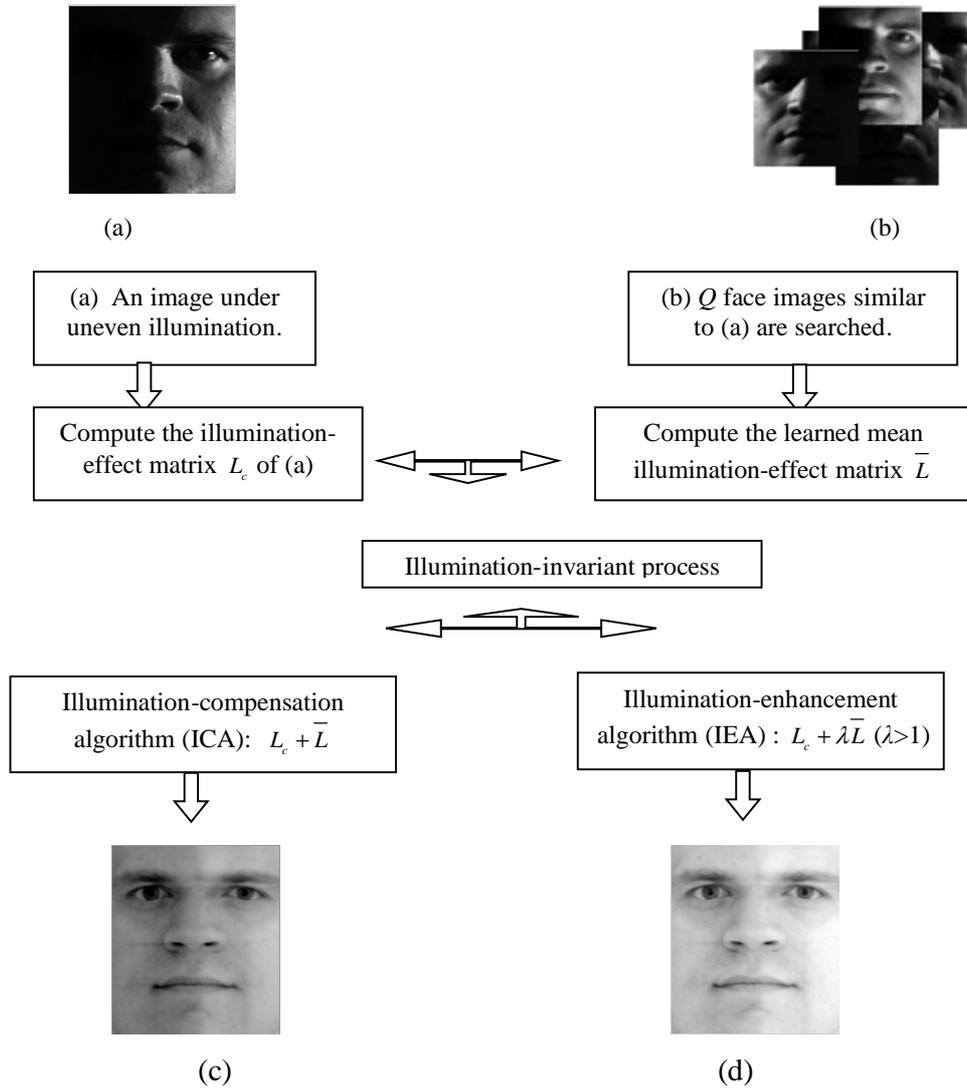
Figure 3.4. The illumination-invariant scheme using illumination compensation and enhancement. (c) Result with illumination compensation for (a). (d) Result with illumination enhancement for (a).

Suppose $I_c(x, y)$ is a face image under uneven lighting and with shadows, $Q$ face images are searched with a similar shape and structure to $I_c(x, y)$ , then the searched $Q$ face images are transformed in the same way. We can learn the mean illumination-effect matrix $\overline{L}(x, y)$ to compensate the images with uneven illuminations and shadows.

Moreover, in [80], Jenkins and Burton indicated that averaging different face images dilutes some transients (e.g., lighting condition and age) while preserving appearances of the face image, which improves face recognition accuracy. The $Q$ face images can express the image-data matrix as:

$$I_t(x, y) = R_t(x, y)L_t(x, y), \quad (1 \le t \le Q) \tag{3.11}$$

where $log(L_t(x, y)) = L_t'(x, y) = U_t W_{t,1} V_t^T$ according to equ. (3.7), and $1 \le t \le Q$. The mean illumination-effect matrices $\overline{L}(x, y)$ can be computed as follows:

$$\overline{L}(x, y) = \frac{1}{Q} \sum_{t=1}^{Q} L_t(x, y). \tag{3.12}$$

Decompose $I_c(x, y)$, which is under uneven and non-frontal illumination, and also write $I_c(x, y)$ using the illumination model accordingly:

$$I_c(x, y) = R_c(x, y)L_c(x, y). \tag{13.3}$$

The mean illumination-effect matrices $\overline{L}(x, y)$ can then be used for illumination compensation as follows:

$$I_c^{ICA} = R_c(L_c + \overline{L}). \tag{3.14}$$

When an image $I_c(x, y)$ is under an uneven illumination condition, shadows may appear, and the image may look different in those regions with insufficient illumination. Therefore, the formulation of $\overline{L}(x, y)$ in (3.12) takes the value of different illumination conditions to generate a mean illumination-effect matrix $\overline{L}(x, y)$ for compensating the face images for uneven lighting and shadows. This can make images lighter and shadowless. We call it the illumination-compensation algorithm (ICA).

### 3.2.4 Illumination enhancement algorithm (IEA)

Inspired by the shadowless lamp used in surgical operations to compensate illumination and remove shadows, we propose an efficient method for image illumination enhancement. We call it the illumination-enhancement algorithm (IEA).

The mean lighting matrix $\overline{L}(x, y)$ can be utilized for image illumination enhancement, not only to compensate for uneven lighting but also to enhance the image by removing any shadows in the image $I_j$ under uneven illumination conditions, as follows:

$$I_c^{IEA} = R_c(L_c + \lambda \overline{L}),\qquad\qquad(3.15)$$

where $\lambda \geq 1$, and is called the "illumination-enhancement factor". When $\lambda=1$, the illumination-enhancement algorithm (IEA) will become the illumination-compensation algorithm (ICA) described in Section 3.2.2.

The effect of IEA can be seen in Fig. 3.5; experimental results will show the performance of this algorithm in the next section of this chapter. The processed face or texture images $I_c^{IEA}$ will have their lighting smoothed, will look similar under even and frontal light sources, and will have any shadow effects greatly reduced (similar to the function of a shadowless lamp in surgical operations, which enhances the uneven illumination conditions so as to remove any shadows).

# 3.3 Experimental results

In this section, we will evaluate our proposed illumination-compensation and -enhancement algorithms. We first show the visual quality of the face images processed by our algorithms. Then, our algorithms are evaluated in terms of a benchmark face-recognition algorithm.

### 3.3.1 Comparison based on visual quality

We carried out a large number of experiments to verify the effectiveness of our proposed methods. To evaluate the performances of our proposed algorithms, we employed the Yale Face Database B [72] and the extended Yale Face Database B [73], which have been commonly used to evaluate the performance of illumination-invariant face recognition methods. The Yale Face Database B consists of 10 classes, named from yaleB01 to yaleB10. The extended Yale Face Database B contains 28 human subjects, named from yaleB11 to yaleB13, and from yaleB15 to yaleB39. Each subject in these two databases has 9 poses and is under 64

illumination conditions. The total number of distinct subjects in the two databases is 38. All the face images are cropped with a size of 168×196.

Experiments on face textures were performed to verify the effectiveness of our proposed methods in terms of visual quality. For comparisons with other works, the following methods have been implemented: − a widely used method in image processing for modifying the dynamic range and contrast of images − was first employed to improve the visual appearance of all images used in experiments [74]; the wavelet-based normalization method proposed in [53], which enhances the contrast as well as the edges of face images; the algorithm [55] which sets the coefficients in the wavelet approximation sub-band to zero. In our experiment, the "db4" wavelet is used with 1-level decomposition for the cropped face images; and the illumination normalization method [79] including a series of stages (Gamma correction, DoG filtering, Masking (optional) and contrast equalization) designed to deal with the effects of illumination variations.

Figure 3.5 shows the results based on two distinct subjects from the database, which can illustrate the superior performance of our algorithms in terms of the illumination-compensation and -enhancement capabilities. Figure 3.5(b) shows the images processed by histogram equalization [74]. Although the visual appearance and the contrast of the textures are enhanced, the results are not illumination-invariant. Figure 3.5(c) shows the results using [53], which employs histogram equalization to enhance the contrast of the approximation coefficients and multiply each element in the detail coefficient matrix with a scale factor. In our experiments, we set the scale factor to 2. We see that our method can enhance the details in the face texture better than the histogram-equalization method, e.g. the eye regions shown in the first row of Figure 3.5(b) and 3.5(c). Figure 3.5(d) illustrates the results using the method in [55], which alleviates the illumination influence by setting the coefficients in the wavelet approximation subband to zero. However, this scheme also discards some detailed information about a face, which is important for recognition. As shown in Figure 3.5(e), the illumination normalization method in [79] can deal with the effects of illumination remove shadows effectively, however, face mages of the same class are look dissimilar

|   (a)   |   (b)   |   (c)   |   (d)   |   (e)   |   (f)   |   (g)   |   (h)   |
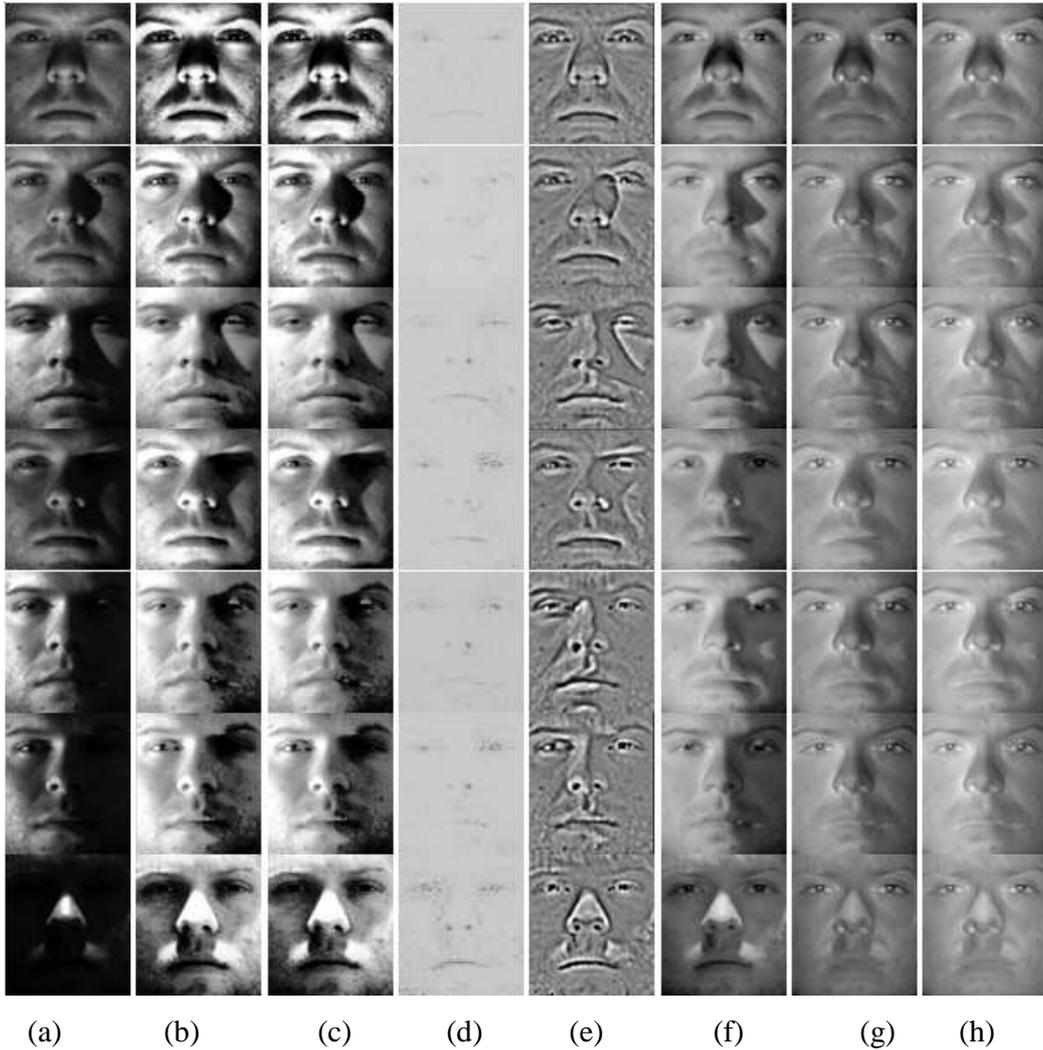
Figure 3.5. Face images of the same subject from the Yale Face Database B using different illumination-compensation and illumination-enhancement methods: (a) the original face images, (b) results using the histogram-equalization method [74], (c) results using the method in [53], (d) results using the method in [55], (e) results using the method in [79], (f) results using our illumination-compensation algorithm, (g) results using our illumination-enhancement algorithm with $\lambda = 3$, and (h) results using our illumination-enhancement algorithm with $\lambda = 5$.

under different illumination variations and some artifacts are appeared. Figure 3.6(f) show the results using our proposed illumination-compensation scheme, which can effectively remove uneven illumination and shadows on face images. However, this

compensation scheme may produce small artifacts over those face textures where the 3-D face heightmap changes dramatically. This results in a large variation of the illumination-effect matrix $L(x, y)$ under different illumination conditions, and therefore $L(x, y)$ has a greater difference to the learned mean illumination-effect matrix $\bar{L}$. Figure 3.5(g) and 3.5(h) show the results using our illumination-enhancement methods with the illumination-enhancement factor $\lambda$ set at 3 and 5, respectively. We can observe that uneven lighting is compensated and the shadows are smoothed. The results show that our proposed method can effectively alleviate the illumination effect, and can produce a better visual quality than the results using the methods in [53] and [55]. In conclusion, the illumination-enhancement method is more suitable for illumination-invariant face textures. With the illumination-enhancement factor set at different values, the effect of uneven illumination can be alleviated effectively, or even completely, without producing any artifacts.

Experimental results using all the face images in the database show that our simple, non-iterative illumination-compensation and illumination-enhancement algorithms can achieve a good performance level, and can effectively reduce the illumination effects while retaining the symmetrical structures and patterns of faces.

### 3.3.2 Performances in terms of recognition accuracy

In this section, we will evaluate the effectiveness of the proposed approaches for illumination-invariant face recognition. In this chapter, we concentrate on the issue of illumination compensation and illumination enhancement for illumination-invariant face images, rather than face recognition. Nevertheless, a good illumination-compensation and -enhancement method should also help to improve the face recognition rate.

The PCA-based algorithm [3] (also known as eigenfaces) is a benchmark for appearance-based and image-based face recognition approaches [1]. Therefore, it is used in these experiments to illustrate the effectiveness of our algorithm for face recognition. Pentland et al. [71] have shown that the first three eigenvectors represent illuminations on face images, and have also empirically shown that a superior face recognition performance can be achieved if the first three

eigenvectors are excluded. Therefore, we will evaluate the following illumination-compensation and -enhancement algorithms using PCA-based algorithms for comparison:

- PCA-based algorithm [3];
- PCA-based algorithm with the first three eigenvectors excluded [71];
- Histogram equalization and PCA-based algorithm [74];
- Illumination normalization [53] and PCA-based algorithm;
- Illumination invariant [55] and PCA-based algorithm;
- Illumination normalization [79] and PCA-based algorithm;
- Illumination compensation and PCA-based algorithm;
- Illumination enhancement with $\lambda = 3$ and PCA-based algorithm; and
- Illumination enhancement with $\lambda = 5$ and PCA-based algorithm.

These eight PCA-based algorithms are denoted as Algorithm1, Algorithm2, Algorithm3, Algorithm4, Algorithm5, Algorithm6, Algorithm7, Algorithm8, and Algorithm9, respectively. In the experiments, all 38 distinct subjects from the Yale Face Database B were used. A randomly selected face of each subject is used as a training sample, while the remaining faces are used for testing. As in [76, 77], the $L_1$ norm distance metric is used, which is a more suitable distance measure than the Euclidean distance metric ($L_2$) for PCA-based algorithms.

Figure 3.6 shows the recognition rates for each of the 38 subjects, based on the eight PCA-based algorithms. Table 3.1 tabulates the average recognition rates of the eight PCA-based algorithms. As is shown in Table 3.1, if no compensation/normalization scheme is employed, the average recognition rate is only 26.58%. The average recognition rate increases to 31.97% if the first three eigenvectors are not used. The performance of the algorithm using histogram equalization [74] in the spatial domain can further improve the rate slightly, to 34.45%. These experimental results are consistent with those in [71], and prove that this simple scheme is effective. Because the illumination normalization method in [79] can remove shadows and normalize illumination variations, the average recognition rate increases to 67.46%. Our proposed algorithms − Algorithm7, Algorithm8, and Algorithm9 −significantly outperform Algorithm4 and Algorithm5.

This is because, as proposed in [53], although Algorithm4 can enhance the detailed information on face images by multiplying the detail coefficient matrix with a scaling factor, it cannot eliminate the illumination effect completely. Algorithm5 [55] attempts to smooth out the illumination influence by setting the wavelet approximation coefficients to zero. This results in some detailed information, which is useful and important for recognition, being missed. It is obvious that Algorithm7, Algorithm8, and Algorithm9 can achieve better performances than the other five methods. The average recognition rates are 73.65%, 82.33%, and 84.39% for the illumination-compensation algorithm, and the illumination-enhancement algorithm with $\lambda=3$ and $\lambda=5$, respectively. As discussed in section 3.1, the illumination-compensation scheme may produce artifacts under some extreme illumination conditions. Nevertheless, the effect of these artifacts will be reduced when the face images are projected into the PCA sub-space. Consequently, the average recognition rate is still high, and only slightly lower than that of Algorithm8 and Algorithm9. Our illumination-enhancement algorithm can achieve the best performance. This is due to the fact that our illumination-enhancement scheme can make face images having more even illumination; in particular, shadows can be removed effectively. As a result, face images of the same class will resemble each other more when compared to the illumination-compensation scheme. Moreover, the recognition rate of the illumination-enhancement approach with a larger illumination-enhancement factor ($\lambda=5$) is higher than with the smaller one ($\lambda=3$). This is because a larger illumination-enhancement factor ($\lambda=5$) can alleviate uneven illumination more completely than a smaller factor ($\lambda=3$) can.
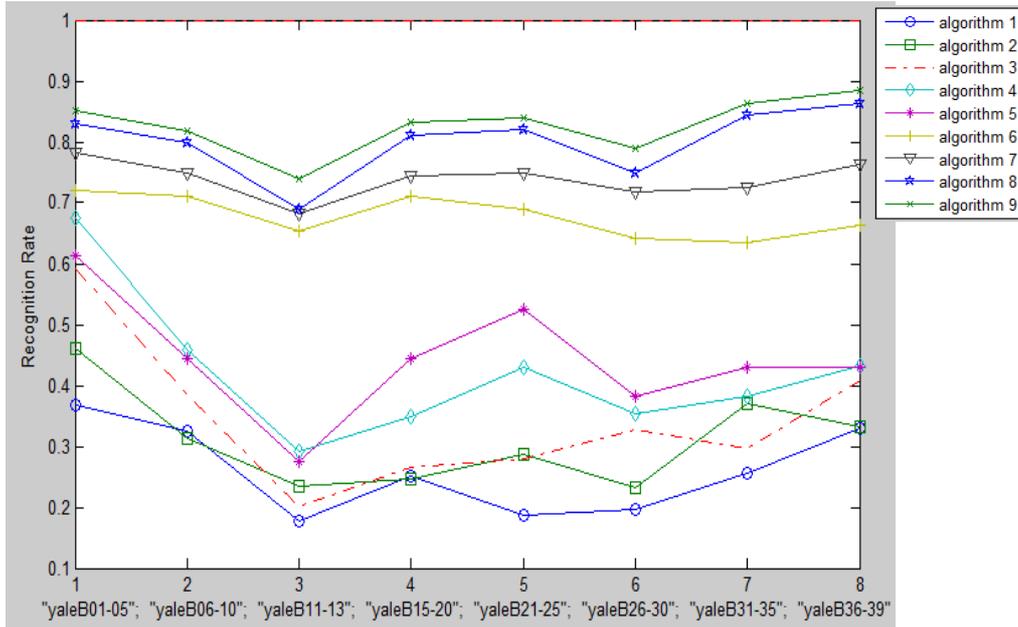
Figure 3.6. Face recognition rates for the 38 distinct subjects in the Yale Face Database B and the extended Yale Face Database B.

Table 3.1. The average recognition rates (ARR) of the eight face recognition schemes for the Yale Face Database B and the extended Yale Face Database B.

| | Alg.1 | Alg.2 | Alg.3 | Alg.4 | Alg.5 | Alg.6 | Alg.7 | Alg.8 | Alg.9 |
|---|---|---|---|---|---|---|---|---|---|
| ARR | 26.6% | 32.0% | 34.5% | 42.2% | 44.2% | 67.5% | 73.7% | 82.3% | 84.4% |

Figure 3.7 demonstrates the $L_1$ norm distance measures of those images in the class "yaleB10" (indexed as 10 in the figure) using our illumination-compensation and -enhancement methods. It can be seen that the respective measured distances for the subject indexed as 10 are the lowest for our three algorithms, i.e. Algorithm7, Algorithm8, and Algorithm9. In addition, the distances based on Algorithm8 and Algorithm9 are close, and are much lower than that of Algorithm. In summary, using the illumination-compensation and -enhancement methods, the recognition rate can be improved. This demonstrates that the proposed schemes form a useful pre-processing step for practical face recognition.

|       |       |       |
|-------|-------|-------|
| (a)   | (b)   | (c)   |
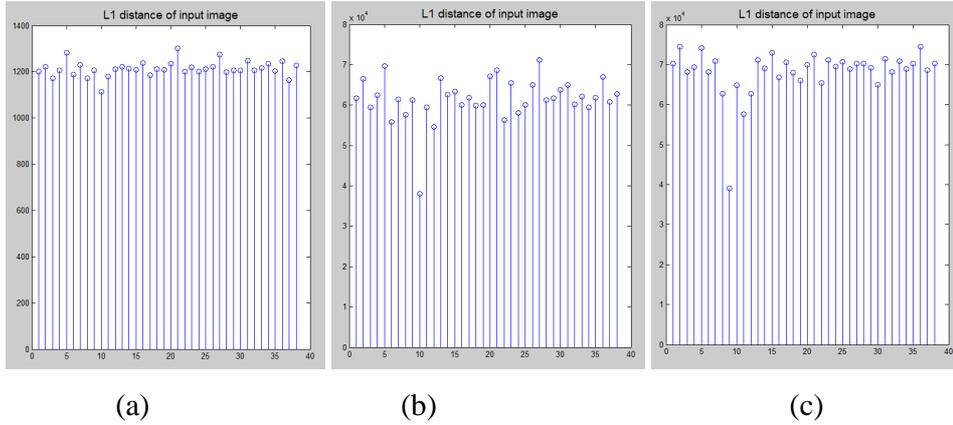
Figure 3.7. $L_1$ norm distance measures for an image of the class "yaleB10" using our illumination-compensation and -enhancement methods: (a) illumination-compensation, (b) illumination-enhancement with λ=3, and (c) illumination-enhancement with λ=5.

# 3.4 Summary

In this chapter, we have proposed an effective scheme for illumination compensation and enhancement, which can generate illumination-invariant face images. Unlike the traditional Lambertian model, which requires a number of images to reconstruct 3D models for illumination-invariant representation, with the assumption of the existence of a single-point light source, our proposed methods can overcome all these limitations, and are suitable for outdoor environments without any postulation of the light sources. Experiment results show the superior performances of our proposed methods for face images in terms of both visual quality and recognition rate.

# Chapter 4 A Novel Face-Hallucination Scheme Based on SVD

Constructing a high-resolution (HR) image from its low-resolution (LR) inputs is called super-resolution in image-processing research field. For face images, this technology is also called face hallucination [8], which has become one of the most important fields for face recognition. This technique is particularly important for video surveillance, in which images are captured at a long distance by a camera. Face images in these videos are usually of low resolution and the videos are compressed with a high compression ratio, which pose an challenging issue in face recognition.

The rest of this chapter is organized as follows. First, a brief introduction of related work on face hallucination will be given. Then we will present a simple mapping model and introduce our proposed face-hallucination scheme in detail. Experiment results and a conclusion are presented at the end.

## 4.1 Related work on face hallucination

Face hallucination was firstly proposed by Baker et al. [8], and has drawn many researchers' attention since then. A pixel-wise super-resolution (SR) method was proposed, which uses the Laplacian pyramid and the Gaussian pyramid to decompose an image into a pyramid of features in order to generate a HR face image. Later, the limitations on super-resolution and how to break these limitations were introduced in [14]. Freeman et al. [15] proposed a nonparametric patch-based prior along with the Markov random field model to produce the desired HR images. In [16], temporal correspondence and a prior model are combined to hallucinate faces. Many researchers [17] [18] have further developed patch-based SR frameworks. A sparse-coding method [19] was proposed to represent a LR input

patch as a combination of its raw neighboring image patches, and the target HR patch is generated directly by using the same combination coefficients as the corresponding neighboring HR patches. The algorithms proposed in [17] [18] also used the same approach, in which a number of similar neighbors to the LR input patches are searched from a training dataset, and then a specific method is adopted to reconstruct the corresponding HR images. In [20], Wang et al. proposed a holistic face-hallucination method which employs Principal Component Analysis (PCA) to represent a LR input image as a linear combination of LR training samples. The corresponding HR image is then estimated using the same linear combination as the corresponding HR training samples. Park et al. [21] utilized the PCA-based SR framework [20] to develop an example-based face-hallucination method. As the PCA method considers global-structure information about facial images, it is less suitable for use in patch-based approaches [24]. In [22], a hybrid method was proposed based on global and local constraints to apply face hallucination to unregistered images. In [23], a novel example-based image SR method was proposed, in which a class-specific predictor is designed for each class of patches so as to improve the accuracy of estimating the high-frequency content. Another early learning-based SR technique was developed by Qiu [24, 25]. These methods are based on the observation that, for a LR version of an image that visually resembles its HR counterpart, the LR and corresponding HR image must be intrinsically correlated. In other words, the LR image can be used as an input to predict its HR counterpart. Ma et al. [26] proposed a method to hallucinate HR image patches using patches at the same position of each training image. Then, the optimal weights for the training position-patches are estimated and used to reconstruct the HR patches. Recently, a new face-hallucination framework − namely, from local-pixel structure to global image SR (LPS-GIS) − was developed in [27] and is based on the assumption that two similar face images should have similar local-pixel structures. This new framework uses an input LR face image to search a face database for similar example HR faces in order to learn the local-pixel structures for the target HR face. In [28], a synthesized approach which utilizes both the shape and texture components is proposed. These two components are

based on accurately aligned local image regions. To achieve sufficient accuracy in alignment, shape reconstruction is solved together with texture reconstruction in a coordinated manner.

In contrast to these previous works, we propose a novel and efficient face-hallucination scheme based on a SVD-based mapping model. We firstly verify that a simple mapping model in the image space $R^{m \times n}$ is inappropriate and unfeasible, as the mapping scheme is too coarse to generate satisfactory results. Then, we propose a more sophisticated mapping model in the eigenspace based on SVD. This proposed face-hallucination framework consists of three steps. In the first step, $M$ example faces that share the most similarity to the input LR face image are searched from a database, and the optical-flow method is then used to align these $M$ example image pairs. In the second step, we compute the leading eigenvectors, which account for most of the information contained in the image. We also observe and prove that, based on the Frobenius norm, the corresponding singular values of an image across different resolutions have approximately linear relationships. Furthermore, we can interpolate the other two data matrices generated by SVD to a higher resolution, as the interpolation of these matrices will not change the holistic structure or the pattern of the face image. The mapping scheme, which utilizes the interpolated SVD matrices multiplied by the learned corresponding mapping matrices to generate more details of face images, can be viewed as a "coarse-to-fine" manner. In the third step, a residual-error matrix, which represents the high-frequency information or the detailed local features missed in the previously predicted HR face image, is generated and added to the one produced in the second stage.

### 4.1.1 Analysis the drawbacks of the existing techniques

For the existing techniques, the main drawback is that when a new and novel face which is significantly different from the training samples, most of the existing learning-based face-hallucination methods will likely produce artifacts and discontinuities in the reconstruction results. Our proposed method has three advantages such that the abovementioned problem can be solved to a certain extent.

First, aligning and warping the retrieved *M* example images via optical flow makes our algorithm able to estimate fine details more accurately. As the *M* warped images should have similar holistic structures and patterns, they can help to prevent the algorithm from producing artifacts and discontinuities in the reconstruction results. Second, as it is proven that the major singular values of the associated eigenvectors of the same image at different resolutions have approximately linear relationships, the estimation of the corresponding singular values of the HR face images will become more reliable. This can effectively preserve and reconstruct the dominant information in the HR face image. The third reason is that the learned mapping matrices for the other two matrices of the SVD representation can be seen as holistic constraints, which do not change the global structure during the reconstruction of HR images. Experimental results show that our algorithm is effective and produces promising hallucination results.

# 4.2 The mapping model

### 4.2.1 The mapping model in the spatial domain

Denote $I_h$ as a gray-scale face image with a size of $m \times n$ pixels and the corresponding LR image as $I_l$, which is subsampled from the original HR face image as follows:

$$I_l = I_h \downarrow \alpha,\qquad(4.1)$$

where $\alpha$ is the down-sampling factor used.

The LR image $I_l$ can be interpolated by using the same factor $\alpha$, which is called the magnification factor, to generate a new image $I_l'$. Therefore, $I_l'$ and $I_h$ have the same resolution. If a mapping matrix *P* exists as follows:

$$I_h = I_l'P,\qquad(4.2)$$

then the HR image $I_h$ can be reconstructed from its LR subsampled images $I_l$ using the mapping matrix *P*. This mapping model in the $R^{m \times n}$ space expresses super-resolution as a product of two matrices, $I_l$ and *P*.

The key issue for this model is to learn the matrix $P$ from a set of pairs of LR and HR face images. The learned mapping matrix, denoted as $\tilde{P}$, can be computed using the pseudo-inverse technique as follows:

$$\tilde{P} = \left(I_l'^T I_l'\right)^{-1} I_l'^T I_h. \tag{4.3}$$

In practice, $I_l'^T I_l'$ is usually a singular matrix. The learned optimal mapping matrix, denoted as $\tilde{P}$, can be computed using an approximation of $P$ as follows:

$$\tilde{P} = \left(I_l'^T I_l' + \lambda E\right)^{-1} I_l'^T I_h, \tag{4.4}$$

where $\lambda$ is a small positive integer, and $E$ is a unit matrix with a size of $n \times n$. The estimated HR face image, denoted as $\hat{I}_h$, can be reconstructed as follows:

$$\hat{I}_h = I_l' \tilde{P}. \tag{4.5}$$

However, in real situations, the mapping model in the $R^{m \times n}$ space is too simple and coarse to achieve satisfactory results. Fig. 4.1 shows an example of the result using the mapping model in the $R^{m \times n}$ space. As shown in Fig. 4.1(d), the visual quality of the reconstructed image using the approximate mapping matrix $\tilde{P}$ is not satisfactory.



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

Figure 4.1. Reconstruction results based on the mapping model in the $R^{m \times n}$ space and in the eigenspace, respectively, with a magnification factor of 4: (a) the LR image obtained by downsampling the original image by a factor of 4, (b) the original HR image, (c) the result based on bicubic interpolation, (d) the reconstructed image based on the mapping model in the $R^{m \times n}$ space, and (e) the reconstructed image based on the SVD-based mapping model in the eigenspace.

## 4.2.2 The mapping model in the eigenspace based on singular value decomposition

In the previous section of this chapter, we showed that a simple mapping model in the $R^{m \times n}$ space is unable to provide satisfactory results in face hallucination. In this section, we use a mathematical framework to achieve a more effective image representation for face hallucination. Our proposed framework projects an image $I$ of size $m \times n$ to the eigenspace using SVD.

The image $I$ can be viewed as a matrix with $m$ rows and $n$ columns. Assume that $m \geq n$, by using SVD, $I$ can be written as the product of a left matrix $U$, a $m \times n$ diagonal matrix $W$ with positive or zero diagonal elements, and the transpose of a right matrix $V$, i.e.

$$I = UWV^T, \tag{4.6}$$

where $U^T U = V^T V = E$. The matrix $U$ is a $m \times m$ column-orthogonal matrix, while $V$ is a $n \times n$ orthogonal matrix. The elements $w_i$ on the diagonal of $W$ are called singular values (the square root of the eigenvalues), i.e.

$$W = diag(w_1, w_2, ...., w_i, ...., w_n). \tag{4.7}$$

The singular value vector $s$ of the image $I$ is defined as follows:

$$s = [w_1, w_2, ...., w_i ...., w_n]^T, \tag{4.8}$$

where $1 \leq i \leq n$, and $w_i$ is the $i^{\text{th}}$ singular value of $I$ in the singular value vector $s$ such that $w_i \geq w_{i+1}$. It can be observed that the singular values decrease dramatically. The following definitions can be used to measure the information accounted for by an eigenvector [29]:

$$f_{indi}(i) = \sqrt{w_i^2 \Big/ \sum_{j=1}^{n} w_j^2}, \tag{4.9}$$

and the accumulation of the first $k$ eigenvectors [30] can be measured as follows:

$$f_{cumu}(k) = \sqrt{\sum_{i=1}^{k} w_i^2 \Big/ \sum_{i=1}^{n} w_i^2}. \tag{4.10}$$

We found that the first several eigenvectors are sufficient to account for almost all of the information contained in an image. This observation is also true for

texture images [31][32]. For illustration, Fig. 4.2 shows the information accounted for by the eigenvectors for two different face images. Figs. 4.2(a) and 4.2(b) are two distinct face images used. Figs. 4.2 (c) and 4.2 (d) show the information accounted for by individual eigenvectors and by the accumulated eigenvectors, respectively. It can be seen that the first 20 eigenvectors can account for nearly all the information in a face image.



(a) Face Image 1     (b) Face Image 2
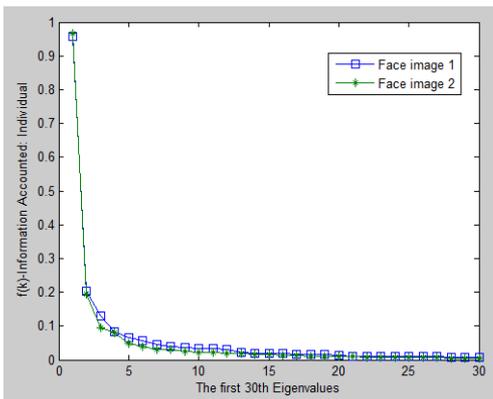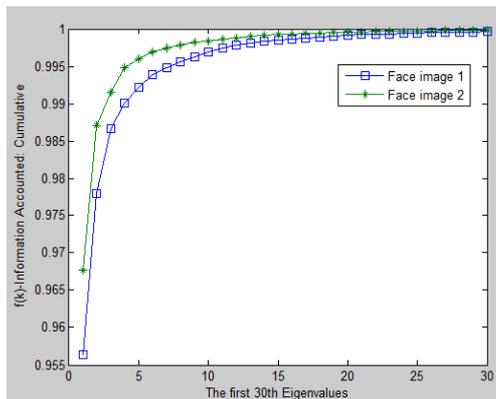


(c)                                    (d)

Figure 4.2: An example of the information accounted for by the first thirty eigenvectors of face images: (a) and (b) are two distinct face images of a man and a woman, respectively; (c) the information accounted for by individual eigenvectors, calculated using (4.9); and (d) the cumulative information accounted for, calculated using (4.10).

It also can be seen that the fast decay of the eigenvalues can be commonly observed across a wide variety of image types. Fig. 4.3 shows different types of images, and a similar rate of decay of the eigenvalues can be seen.
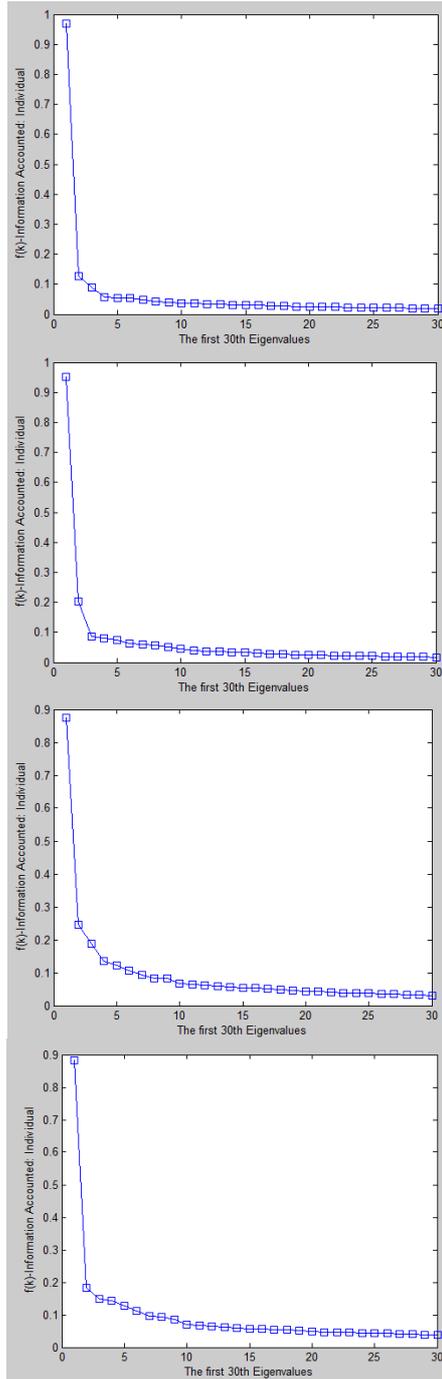
(a)                                    (b)

Figure 4.3: The information accounted for by the first thirty eigenvectors of different image types: (a) images of different types, (b) the information accounted for by the individual eigenvectors, calculated using (4.9).

Since the singular values decrease rapidly and the first few eigenvectors can account for most of the information, the original matrix $W$ can be approximated as $\hat{W}_I$ as follows:



(a) Original Image     (b) $k=3$     (c) $k=10$     (d) $k=20$

(e) $k=30$     (f) $k=40$     (g) $k=50$     (h) $k=60$
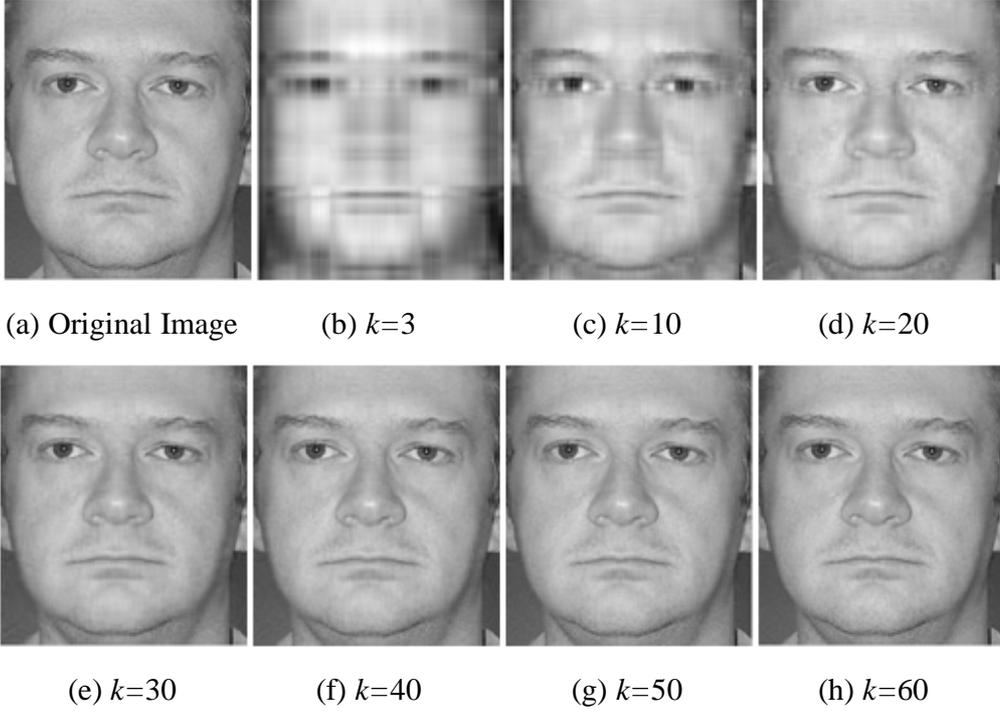
Figure 4.4. Examples of different approximated images with $k$ set at different values.

$$\hat{W}_I = diag(w_1, w_2, ...., w_k, 0, ...., 0),\tag{4.11}$$

and we have

$$\sum_{i=1}^{n} w_i^2 \cong \sum_{i=1}^{k} w_i^2,\tag{4.12}$$

where $k$ is the number of singular values or eigenvectors to be retained. We choose the first $k$ singular values, rather than all the $n$ non-zero singular values available, because those high-order singular values represent the high-frequency content, or noise, in the interpolated LR face image under consideration. The linear relationship for the low-order singular values does not hold for those high-order singular values. Based on the first $k$ singular values, we can obtain an approximated

image $\hat{I}$, which contains almost the same information as the original image *I*. The

image $\hat{I}$, which can be viewed as a matrix, can be expressed as follows:

$$\hat{I} = U\hat{W}_I V^T.$$
(4.13)

Fig. 4.4 shows different approximated images when the image has a resolution of $262 \times 300$ and *k* is set at different values. The number of non-zero singular values for the image is 262. We can observe that the image quality has a very slight difference when *k* is larger than 40.

**4.2.2.1 The diagonal matrix *W* at different resolutions**

According to [33], if a matrix *A* has singular values $w_i$, where $1 \le i \le n$, then

$$\| A \|_F^2 = \sum_{i=1}^{n} w_i^2 ,$$
(4.14)

where $\| A \|_F$ is the Frobenius norm of the matrix *A*, which is defined as the square root of the sum of the squares of all its entries. The following is a brief proof of (4.14):

**Proof:** If matrix *A* has singular values, $w_1, w_2, ...., w_i ...., w_n$, then $\| A \|_F^2 = \sum_{i=1}^{n} w_i^2$ .

Using SVD, a matrix *A* can be expressed as $A = UWV^T$ . First note that, for any matrix *C* whose $i^{\text{th}}$ column is denoted as $c_i$, i.e. $C = (c_1 / ... / c_n)$ , then $\| C \|_F^2 = \| c_1 \|_F^2 + ... + \| c_n \|_F^2$ . Now, we have

$$\| A \|_F^2 = \| UWV^T \|_F^2$$

$$\overset{[1]}{=} \| WV^T \|_F^2 = \| (WV^T)^T \|_F^2 = \| (V^T)^T W^T \|_F^2 = \| VW^T \|_F^2$$

$$\overset{[2]}{=} \| W^T \|_F^2 = \| W \|_F^2$$

$$\overset{[3]}{=} \sum_{i=1}^{n} w_i^2$$

[1] Let $(c_1 / ... / c_n)$ be the columns of $WV^T$. Since the *U* matrix simply rotates the columns of $(c_1 / ... / c_n)$ without changing their lengths, the two sides are equal.

[2] The *V* matrix simply rotates the columns of $W^T$ without changing their lengths.

76

[3] The diagonal matrix $W$ has positive or zero values on its diagonal and the other elements are zeros.

We observe that when a LR image $I_l$ is interpolated or super-resolved to produce a new HR image $I_h$ with a magnification factor of $\alpha$, the first $k$ main singular values in the singular-value vector $s_h = \begin{bmatrix} w_h^1, & w_h^2, & ..., & w_h^k, \end{bmatrix}^T$ of the image $I_h$ can be approximated as $\alpha$ times of the first $k$ main singular values in the singular value vector $s_l = \begin{bmatrix} w_l^1, & w_l^2, & ..., & w_l^k, \end{bmatrix}^T$ of the original image $I_l$. Hence, we have

$$s_h \cong \alpha s_l. \tag{4.15}$$

In other words, the main singular values of the same image under different resolutions are approximately proportional to each other, with the magnification factor $\alpha$ as the proportional constant.



(a)          (b)



(c)          (d)          (e)
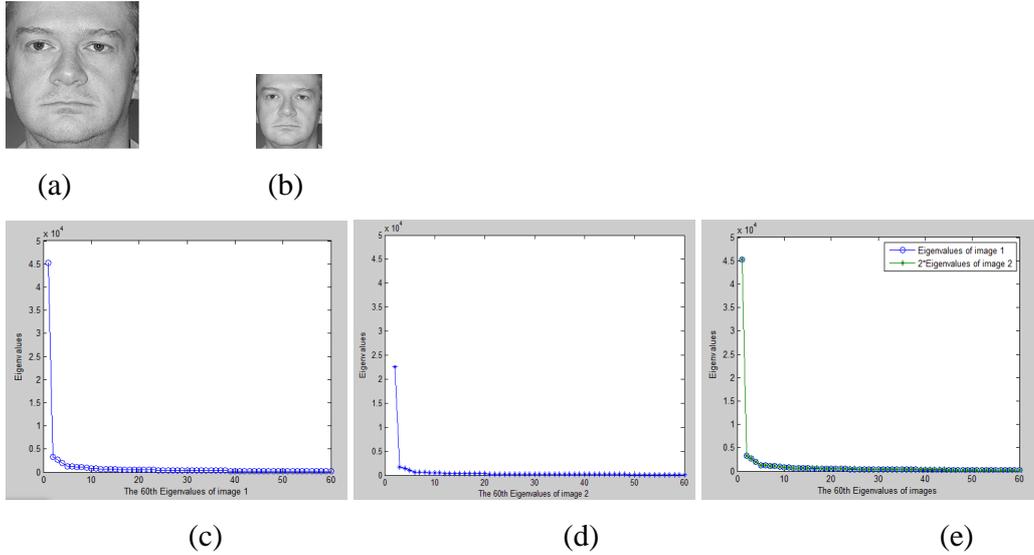
Figure 4.5. The singular values of two images of the same face with different resolutions: (a) An original image of size 262×300; (b) a downsampled image of size 131×150, obtained with a downsampling factor of 2 in both the horizontal and vertical directions; (c) the first 60 singular values, denoted as $s_h$, of the image in (a); (d) the first 60 singular values, denoted as $s_l$, of the image in (b); (e) the first 60

singular values of the image in (a) and twice the values of the first 60 singular values of the image in (b).

**Theorem 1**: If a LR image $I_l$ is interpolated to produce a HR image $I_h$ with a magnification factor of $\alpha$, then the first $k$ main singular values in the singular-value vector $s_h$ of the new image $I_h$ are $\alpha$ times the corresponding first $k$ main singular values of the singular-value vector $s_l$ of the original image $I_l$.

**Proof:** Suppose that $I_l$ is interpolated to produce a new HR image $I_h$ with a magnification factor of $\alpha$, then we have:

$$\left\| I_h \right\|_F^2 \overset{[4]}{\cong} \left\| \alpha I_l \right\|_F^2 = a^2 \left\| I_l \right\|_F^2$$

$$\overset{[5]}{=} \alpha^2 \sum_{i=1}^{n} \left( w_l^i \right)^2$$

$$\overset{[6]}{\cong} \alpha^2 \sum_{i=1}^{k} \left( w_l^i \right)^2 = \sum_{i=1}^{k} \left( \alpha w_l^i \right)^2 = \sum_{i=1}^{k} \left( w_h^i \right)^2$$

$$\therefore \quad s_h \cong \alpha s_l .$$

[4] An interpolation method, with a magnification factor of $\alpha$, produces $\alpha \times \alpha$ neighboring pixels in the interpolated image $I_h$ with similar values to the corresponding pixel in the original LR image $I_l$. For instance, the nearest-neighbor interpolation generates $\alpha \times \alpha$ neighbors of equal values, and the bilinear and bicubic-interpolation methods produce $\alpha \times \alpha$ similar pixels.

[5] According to (4.14), $\| A \|_F^2 = \sum_{i=1}^{n} w_i^2$ .

[6] According to (4.12), $\sum_{i=1}^{n} w_i^2 \cong \sum_{i=1}^{k} w_i^2$ .

Figs. 4.5(c), 4.5(d) and 4.5(e) show the linear relationship of the first $k$=60 main singular values for two images of the same person at different resolutions. The image in Fig. 4.5(a) has double the resolution of the image in Fig. 4.5(b), both horizontally and vertically. Figs. 4.5(c) and 4.5(d) show the first 60 eigenvectors of

the images in Figs. 4.5(a) and 4.5(b), respectively. To illustrate the relationship of the singular values of the two images, Fig. 4.5(e) shows the first 60 singular values of the image in Fig. 4.5(a), as well as the first 60 singular values of the image in Fig. 4.5(b), multiplied by two. We can see that $s_h \cong \alpha s_l$, with the magnification factor $\alpha=2$.

### 4.2.2.2 *Mapping U and V in the eigenspace*

Suppose that an original face image is denoted as $I_h$ and the corresponding sub-sampled LR image is denoted as $I_l$. These two images, $I_h$ and $I_l$, can be expressed in the eigenspace using SVD as follows:

$$I_h = U_h W_h V_h^T \text{ and} \tag{4.16}$$

$$I_l = U_l W_l V_l^T. \tag{4.17}$$

The two matrices $U_l$ and $V_l$ for $I_l$ can be interpolated to form two new matrices $U_l'$ and $V_l'$ that have the same size as $U_h$ and $V_h$, respectively. Define two mapping matrices $P_u$ and $P_v$, as follows:

$$U_h = U_l' P_u, \text{ and} \tag{4.18}$$

$$V_h = V_l' P_v. \tag{4.19}$$

In the eigenspace, these two matrices $P_u$ and $P_v$ can be calculated using pseudo-inverse as follows:

$$P_u = \left(U_l'^T U_l'\right)^{-1} U_l'^T U_h, \text{ and} \tag{4.20}$$

$$P_v = \left(V_l'^T V_l'\right)^{-1} V_l'^T V_h. \tag{4.21}$$

As $U_l'^T U_l'$ and $V_l'^T V_l'$ are always singular matrices, $P_u$ and $P_v$ can be computed approximately as follows:

$$\tilde{P}_u = \left(U_l'^T U_l' + \lambda E\right)^{-1} U_l'^T U_h, \text{ and} \tag{4.22}$$

$$\tilde{P}_v = \left(V_l'^T V_l' + \lambda E\right)^{-1} V_l'^T V_h, \tag{4.23}$$

where $\lambda$ is a small positive integer and $E$ is a unit matrix. An estimation of $U_h$ and $V_h$, denoted as $\hat{U}_h$ and $\hat{V}_h$, respectively, can be expressed as follows:

$$\hat{U}_h = U'_l \tilde{P}_u, \text{ and} \qquad\qquad (4.24)$$

$$\hat{V}_h = V'_l \tilde{P}_v. \qquad\qquad (4.25)$$

For the purpose of super-resolution, a HR image can be reconstructed from a LR image using the matrices $\hat{U}_h$ and $\hat{V}_h$, which can be learned from a pair of LR-HR training face images; and the diagonal matrix $\hat{W}_h$ is computed using the scheme described in Section 4.2.2.2.1, i.e.

$$I_h = \hat{U}_h \hat{W}_h \hat{V}_h^T. \qquad\qquad (4.26)$$

Fig. 4.1 also shows an example of using the mapping model in the eigenspace. Fig. 4.1(e) shows that the HR image reconstructed using the mapping model in the eigenspace is effective, and that our algorithm can produce a promising result.

# 4.3 SVD-based face-hallucination scheme

Suppose that an input LR face image $I_l$ is to be super-resolved with a magnification factor of $\alpha$ to form a HR face image. With this LR input face, *M* LR face images resembling $I_l$, as well as the corresponding reference HR images, are searched from a dataset which contains pairs of LR and HR face images. These *M* pairs of LR and HR training face images, denoted as $I_l^i$ and $I_h^i$ ($i$ = 1, ..., *M*), respectively, should have a high structural similarity to the LR input face after alignment. Each of the similar HR face images can be reconstructed from the corresponding LR faces using the mapping matrices, as described in Section 4.2.2.2, i.e.

$$\hat{U}_h^i = U'^i_l \tilde{P}_u^i, \text{ and} \qquad\qquad (4.27)$$

$$\hat{V}_h^i = V'^i_l \tilde{P}_v^i, \qquad\qquad (4.28)$$

where $1 \leq i \leq M$.

We can now learn the mapping matrices $\tilde{P}_u$ and $\tilde{P}_v$ for face hallucination using the individual mapping matrices $\tilde{P}_u^i$ and $\tilde{P}_v^i$, learned from the *M* pairs of HR and LR images. A linear combination of the mapping matrices can be written as follows:

$$\tilde{P}_u = \sum_{i=1}^{M} \beta_i \tilde{P}_u^i, \text{ and} \tag{4.29}$$

$$\tilde{P}_v = \sum_{i=1}^{M} \gamma_i \tilde{P}_v^i, \tag{4.30}$$

where $\beta_i$ and $\gamma_i$ are the embedding coefficients for $\tilde{P}_u^i$ and $\tilde{P}_v^i$, respectively.

After the LR input image $I_l$ is decomposed by SVD using (4.17), its $U_l$ and $V_l$ can be interpolated to generate two new matrices $U_l'$ and $V_l'$ with the same size as $U_h$ and $V_h$ for the desired HR image $I_h$, respectively. Then, the corresponding approximated matrices $\hat{U}_h$ and $\hat{V}_h$ can be computed as follows:

$$\hat{U}_h = U_l' \tilde{P}_u, \text{ and} \tag{4.31}$$

$$\hat{V}_h = V_l' \tilde{P}_v. \tag{4.32}$$

As described in Section 4.2.2.1, the leading singular values in the diagonal matrix $\hat{W}_h$ can be estimated using the linear relationship $s_h \cong \alpha s_l$. The number of leading singular values $k$ to be used can be determined by using (4.10), such that the first $k$ leading singular values can represent a sufficient amount of information about the face images. In our algorithm, we choose $f_{cumu}(k) \geq \eta$ (where $\eta = 0.99$ in our experiment). After estimating the first $k$ leading singular values, we can also estimate the remaining singular values, denoted as $s_h'(k) = \left[ w_h^{k+1}, w_h^{k+2}, \ldots, w_h^{\alpha n} \right]^T$, in the diagonal matrix $\hat{W}_h$ using a linear combination of the remaining singular values $s_h'^i(k) = \left[ w_h^{i,k+1}, w_h^{i,k+2}, \ldots, w_h^{i,\alpha n} \right]^T$ of the $M$ similar HR images, as follows:

$$s_h'(k) = \left[ w_h^{k+1}, w_h^{k+2}, \ldots, w_h^{\alpha n} \right]^T$$

$$= \sum_{i=1}^{M} \beta_i s_h'^i = \sum_{i=1}^{M} \beta_i \left[ w_h^{i,k+1}, w_h^{i,k+2}, \ldots, w_h^{i,\alpha n} \right]^T. \tag{4.33}$$

Now, the HR image $\hat{I}_h$ can be reconstructed using the estimated matrices $\hat{U}_h$ and $\hat{V}_h$, and the diagonal matrix $\hat{W}_h$, as follows:

$$\hat{I}_h = \hat{U}_h \hat{W}_h \hat{V}_h^T. \tag{4.34}$$

The reconstructed image $\hat{I}_h$ should be similar to the $M$ HR training face images. The following squared error $E(i)$ can be used to measure the reconstruction error:

$$E(i) = \sum_{i=1}^{M} \left\| I_h^i - \hat{I}_h \right\|^2$$

$$\text{s.t. } \hat{I}_h \downarrow \alpha = I_l. \qquad (4.35)$$

The optimal reconstruction weights $\beta_i$ and $\gamma_i$ can be derived by minimizing the following formulation:

$$\beta = \underset{\beta_i, \ \gamma_i}{arg\,min} \left\{ E(i) \right\}$$

$$= \underset{\beta_i, \ \gamma_i}{arg\,min} \left\{ \sum_{i=1}^{M} \left\| I_h^i - \hat{I}_h \right\|^2 \right\}$$

$$\text{s.t. } \hat{I}_h \downarrow \alpha = I_l. \qquad (4.36)$$

The global constrained least-square problem can be computed using the iterative method in [34], and the determined weights are normalized so that their sum is one.

To further improve the visual quality of the reconstructed HR images, our proposed algorithm estimates the residual-error matrix $\overline{C}$, which comprises the high-frequency information about a face image, and represents the detailed local features missing from the global HR image. This residual information is added as the missing high-frequency information to achieve high-quality face hallucination. Based on our proposed SVD-based mapping model, the matrix $\overline{C}$ can be estimated from the individual residual errors of the selected training samples $\overline{C}^i = I_h^i - \hat{I}_h$, where $\hat{I}_h = \hat{U}_h \hat{W}_h \hat{V}_h^T$, as defined in (4.34). We use a Gaussian function to measure the similarity of two images, and the weight $\delta_i$ reflects the global similarity as follows:

$$\delta_i = \exp\left( \frac{-\left\| I_h^i - \hat{I}_h \right\|^2}{\sigma^2} \right), \qquad (4.37)$$

where $i = 1, ..., M$, and $\sigma$ controls the effect of the gray-level difference between the HR training sample and the reconstructed HR image. The weights $\delta_i$ are then normalized so that their sum is equal to 1. Then, the matrix $\overline{C}$ can be computed as follows:

$$\overline{C} = \sum_{i=1}^{M} \delta_i \overline{C}^i$$

$$= \sum_{i=1}^{M} \delta_i \left( I_h^i - \hat{I}_h \right). \tag{4.38}$$

Having determined the residual-error matrix, the final reconstructed HR face image, denoted as $\overline{I}_h$, can be computed as follows:

$$\overline{I}_h = \hat{I}_h + \overline{C}. \tag{4.39}$$

# 4.4 Experimental results

In order to verify the effectiveness of the proposed scheme, the dataset used in [27] is used to evaluate the performance of our proposed framework. The facial images in the dataset were selected from the GT [35], AR [36], and FERET [37] databases, which contain 40, 70, and 120 images, respectively. All the facial images are aligned based on the position of the two eyes, using the method in [38]. The original HR facial images are cropped to a size of 124×108. The number of reference face examples is set at 3, and the parameters $\lambda$ in (22) and $\sigma$ in (35) are empirically set at 0.001 and 50, respectively. Experiments show that using all of the above settings can achieve a satisfactory overall performance. In the experiments, we evaluate all the methods by reconstructing the HR facial images with a magnification factor of 4. All the testing images are evaluated using the "leave-one-out" approach. Two objective quality measures, PSNR (peak signal-to-noise ratio) and SSIM (structural similarity) [39], are used to evaluate the performances of the different methods.

## 4.4.1 The three stages of our proposed SR scheme

| | | | |
|---|---|---|---|
| SSIM | 0.7147 | 0.7646 | 0.7887 |
| SSIM | 0.7025 | 0.7699 | 0.7797 |
| SSIM | 0.7057 | 0.7957 | 0.8135 |
| (a) | (b) | (c) | (d) |

Figure 4.6. Reconstruction results at different stages of our proposed SR scheme: (a) the original HR images, (b) bicubic interpolation, (c) Stage 2 of our proposed method, and (d) Stage 3 of our proposed method after adding high-frequency information.

In our experiment, three face images similar to an input LR face are first searched from a training set, using the same method as in [27]. Then, we warp the reference images to the input using the optical-flow method. In the second stage of

our proposed method, interpolation is employed, which does not change the global structure or the pattern of a face image. It can be seen as a "coarse-to-fine" process using the learned projection matrices $P_u$ and $P_v$ based on the mapping model. In the third phase, a residual matrix, which represents the high-frequency information about a face image, is added to the previously estimated HR face image to further improve the reconstruction accuracy.

Fig. 4.6 illustrates some example results produced at the different stages of our proposed SR scheme, with a magnification factor $\alpha=4$. As shown in Fig. 4.6(b), the bicubic-interpolation method [40] produces blurry results. Fig. 4.6(c) illustrates the results for the second stage of our proposed method, which utilizes the global structure of reference images and the learned projection matrices $P_u$ and $P_v$ based on the mapping model for face hallucination. It can be seen that the results in Fig. 4.6(c) have a better visual quality than those in Fig. 4.6(b). This is because the two learned matrices can be viewed as holistic constraints in the reconstruction of the HR image using the similar global structure and pattern of the searched reference face images. Fig. 4.6(d) shows the results produced by the third stage of our algorithm. By adding the residual matrix, which contains the missing high-frequency information about the face image, the proposed model can produce more photo-realistic images. In addition, we employ SSIM as an objective quality measure to assess the visual quality of the hallucinated results. As shown in Fig. 4.6, the SSIM values for the second stage of our proposed method are larger than those for the bicubic interpolation [40]. After the addition of the missing high-frequency information, the SSIM values for the third stage of our method are larger than those for the second stage. The SSIM values are consistent with human visual perception in terms of the visual quality of the images, and can objectively reflect the effectiveness of the different algorithms.

## 4.4.2 Comparison with the state-of-the-art methods

Two interpolation algorithms, the bicubic-interpolation algorithm [40] and the edge-directed interpolation (NEDI) method [41], are applied for face hallucination and are compared with the second stage of our proposed algorithm, which can also

be considered as an interpolation technique. Another four state-of-the-art face-hallucination algorithms are also compared to our proposed method: two PCA-based, holistic SR methods (the eigentransformation method [20] and a variant of Park's method [21]), one patch-based method (Freeman's method [15]), and Liu's method [22] based on a global parametric model and local nonparametric model. The version of Park's method used in this chapter is different from the original algorithm in [21] in that the training images are warped with reference to the LR input face rather than with reference to a predefined reference face.

Fig. 4.7 shows some samples of the reconstruction results generated using these respective methods, with a magnification factor α=4. It can be seen from Fig. 4.7(b) that the bicubic-interpolation algorithm produces the blurriest results. The results in Fig. 4.7(c) are generated using the NEDI method. However, if a face image has very low resolution, the NEDI method struggles to distinguish edges, and hence also produces blurry results as compared to other SR methods. Therefore, the results of Fig. 4.7(d) show a better visual quality than both the bicubic-interpolation and the NEDI algorithms. Figs. 4.7(d) and 4.7(e) are the results generated by using the eigentransformation method and the variant of Park's method, which are both holistic/global face-hallucination methods. Plausible face structures can be well inferred in the resulting HR images. Nevertheless, as the method is purely holistic, it cannot effectively reconstruct the fine individual facial details of novel testing faces. If a testing face image is very different from the faces in the database, infidelity will result in the reconstructed HR faces. Park's method here employs the morphable face model to capture the shape variations of novel testing faces, so it outperforms the eigentransformation method. However, the HR textures are still reconstructed in a holistic manner, like the eigentransformation method. The face-hallucination results using the patch-based SR methods are illustrated in Fig. 4.7(f). It can be seen that Freeman's method can provide plausible HR facial images with sharp edges and corners. However, as some of the patches are badly matched or conflict with adjacent ones, some structural errors and discontinuities appear in the reconstructed HR images; these errors are drawbacks with most patch-based SR methods. Furthermore, the patch-based SR usually requires thousands of image-

patch pairs to learn, and therefore is computationally expensive. In addition, there are artifacts in the reconstructed HR images, as shown in Fig. 4.7(f).



| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) |
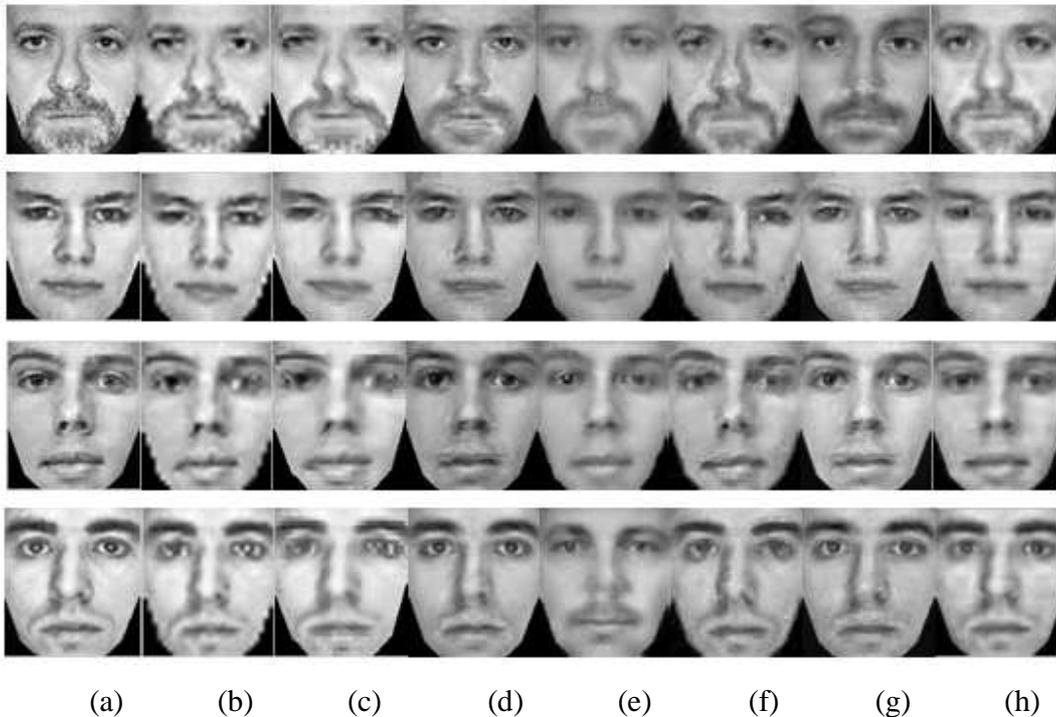
Figure 4.7: Face-hallucination results reconstructed using different methods with a magnification factor of 4 ($\alpha=4$): (a) the original images, (b) bicubic interpolation, (c) NEDI, (d) the eigentransformation method, (e) a variant of Park's method, (f) Freeman's method, (g) Liu's method, and (h) our proposed method.

Compared with the holistic-based and the patch-based approaches, Liu's method [10] utilizes both global and local prior information through a global parametric model and a local nonparametric model. Thus, as shown in Fig. 4.7(g), the method can produce not only visually plausible face structures, but also fine details or textures like those in the HR training images. However, some parts of the hallucinated face, such as the mouth, are somewhat different from the original face. This can be partially attributed to the properties of the PCA-based global model used in this approach. Unlike Liu's method, our framework does not change the global structure of face images, and the main energy can be retained, as explained in Section 4.2.2.1. Furthermore, the mapping scheme and the residual matrix can

produce high-resolution results with clear details. As can be seen from Fig. 4.7(h), compared to other typical state-of-the-art algorithms, plausible HR images with a holistic structure and more details with a better visual quality can be obtained.



| (a) | (b) | (c) | (d) | (e) | (f) | (g) |
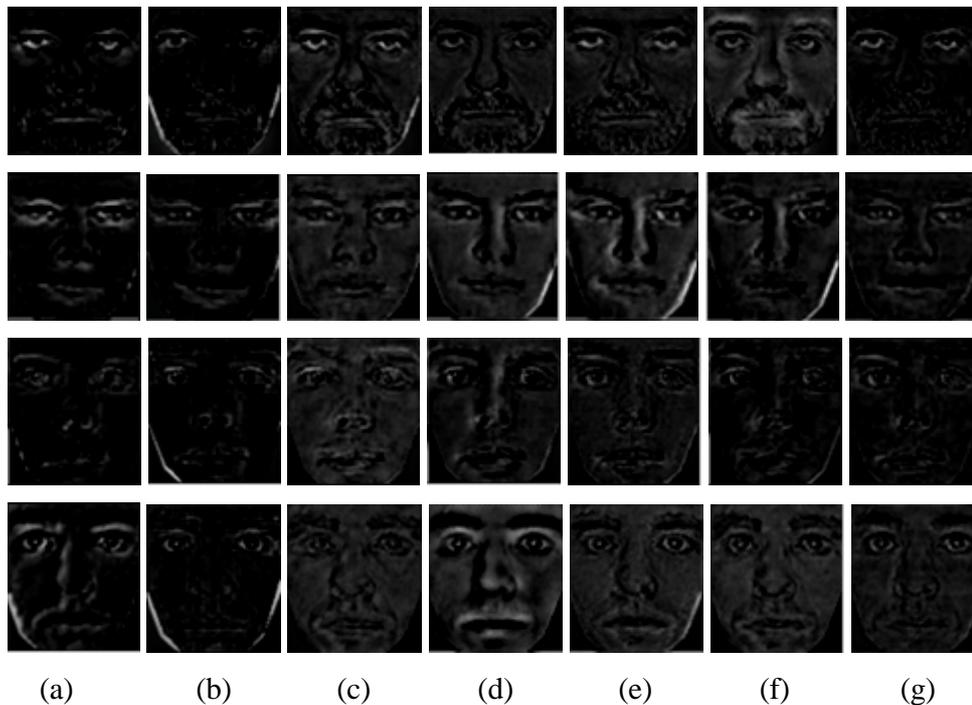
Fig. 4.8. The error maps displaying the reconstruction errors of different methods, with a magnification factor of 4 ($\alpha$=4): (a) bicubic interpolation, (b) NEDI, (c) the eigentransformation method, (d) a variant of Park's method, (e) Freeman's method, (f) Liu's method, and (g) our proposed method.

Fig. 4.8 shows some sample error maps displaying the reconstruction errors produced using these typical state-of-the-art algorithms, with a magnification factor $\alpha$=4. With the visual aid of the error maps in Fig. 4.8, it can be seen that our proposed method produces more accurate results, compared to other typical state-of-the-art methods.

| (a) | (b) | (c) | (d) | (e) | (f) | (g) |
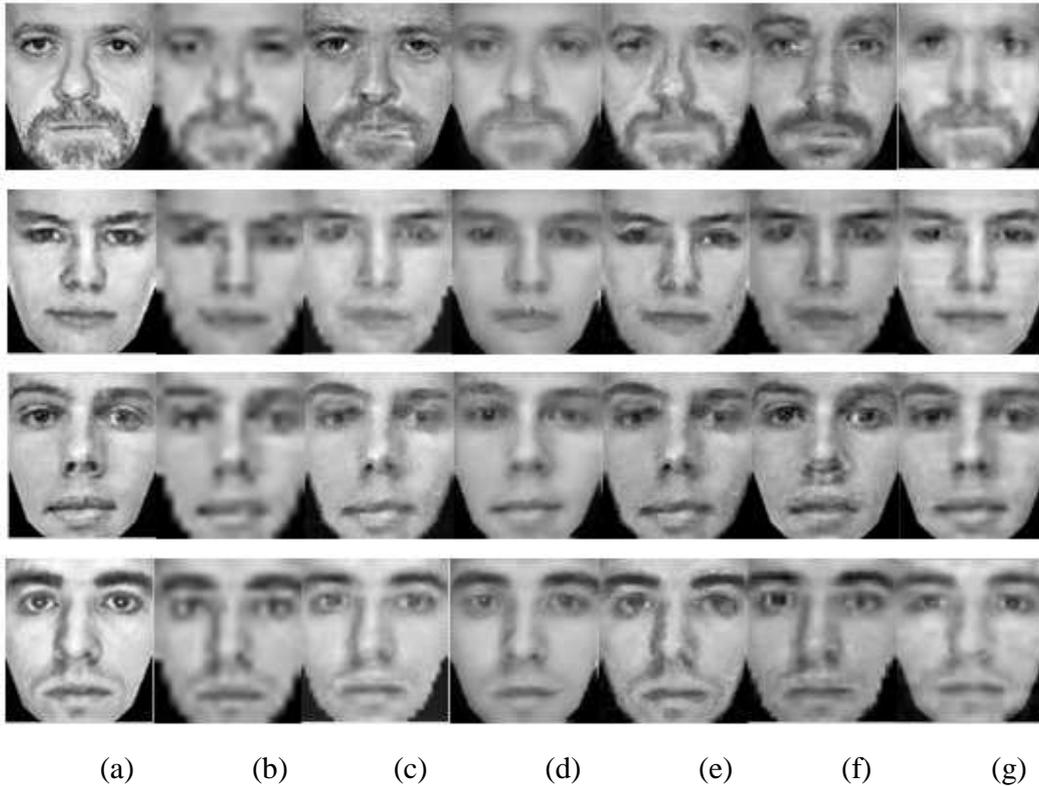
Figure 4.9: Face-hallucination results reconstructed using different methods with a magnification factor of 6: (a) the original images, (b) bicubic interpolation, (c) the eigentransformation method, (d) a variant of Park's method, (e) Freeman's method, (f) Liu's method, and (g) our proposed method.

We also evaluate our method by reconstructing the HR facial images with a magnification factor of 6 ($\alpha$=6). Fig. 4.9 shows the reconstruction results for the images used in the previous experiment. It can be seen that the generic, patch-based SR method is no longer able to produce promising HR facial structures effectively, while those methods that employ the holistic structure of facial images can provide a better performance than the patch-based method in terms of visual quality. This is because, when the LR observations have very low resolution, using only the local patch-based prior is not sufficient to infer the target HR image. In contrast, by using the holistic characteristics of the HR face examples, a stronger prior can be imposed on the reconstructed facial images, which can result in a better facial appearance. The images in Fig. 4.9(e) are produced by the patch-based method;

with very low resolution images as the input, it is hard to generate good results. Fig. 4.9(c) and (d) show the holistic SR methods, which can provide better HR results for every LR image. The hallucinated faces generated using Liu's method have a better holistic appearance than those generated using the patch-based method, as shown in Fig. 4.9(f). However, it is also observed that Liu's method results in some distortions at facial features different from those in the original images. Our method is illustrated in Fig. 4.9(g): the results look consistent with the original HR face even though these were very low-resolution images. Experimental results based on visual inspection show that the proposed method is effective and can fulfill the task of face hallucination. It achieves a satisfactory performance and produces promising results while preserving the details and the structure of the human face.

Two quantitatively objective quality measures, i.e. PSNR and SSIM, are also employed to evaluate different methods with different magnification factors. Table 4.1 tabulates the average PSNR and SSIM, and the corresponding standard deviations, of the different methods with a magnification factor of 4. Table 4.1 shows that our method is superior, in terms of the two measurements, to the other two classical interpolation approaches, and it can achieve a comparable performance to the other typical state-of-the-art algorithms.

Table 4.1. The average PSNR and SSIM, and the corresponding standard deviations shown in brackets, of the different face-hallucination algorithms with a magnification factor of 4.

|  | PSNR | SSIM |
| --- | --- | --- |
| Bicubic interpolation | 22.16 (1.84) | 0.7022 (0.1071) |
| NEDI | 21.64 (2.75) | 0.7034 (0.0943) |
| Eigentransformation method | 25.40 (2.38) | 0.7167 (0.1035) |
| A variant of Park's method | 25.97 (3.65) | 0.7541 (0.1130) |
| Freeman's method | 26.28 (3.23) | 0.7894 (0.1581) |
| Liu's method | 25.72 (3.61) | 0.7345 (0.1766) |
| Our proposed method | 26.70 (3.55) | 0.7887 (0.1224) |

Table 4.2. The average PSNR and SSIM, and the corresponding standard deviations shown in brackets, of the different face-hallucination algorithms with a magnification factor of 6.

|  | PSNR | SSIM |
|---|---|---|
| Bicubic interpolation | 19.40 (1.92) | 0.5743 (0.0905) |
| Eigentransformation method | 23.47 (2.68) | 0.6735 (0.0921) |
| A variant of Park's method | 24.49 (2.45) | 0.7044 (0.1248) |
| Freeman's method | 23.15 (2.50) | 0.6881 (0.1415) |
| Liu's method | 23.88 (2.65) | 0.6642 (0.1503) |
| Our proposed method | 24.52 (2.37) | 0.7109 (0.1167) |

Table 4.2 tabulates the average PSNR and SSIM, and the corresponding standard deviations, produced by using the very low-resolution images as inputs, with a magnification factor of 6. It can be seen from Table 4.2 that our method again outperforms other state-of-the-art algorithms. In addition, we show the statistical results in terms of the SSIM across all the faces in the database (230 images in total). Instead of showing the SSIM of each face, we group 20 faces and compute the corresponding average SSIM. And the last 30 faces of the 230 images are grouped, rather than having groups of 20 and then the final 10. Fig. 4.10 and Fig. 4.11 illustrate the average SSIM for the different face-hallucination algorithms with a magnification factor of 4 and 6, respectively. When the magnification factor is 4, the performance of Freeman's method is similar to our method, but our method can achieve a better performance in terms of PSNR. When the magnification factor is increased to 6, our method is significantly better than Freeman's method in terms of both PSNR and SSIM. Also our method has a slightly better performance than does the variant of Park's method. In conclusion, our method can achieve a better performance statistically.
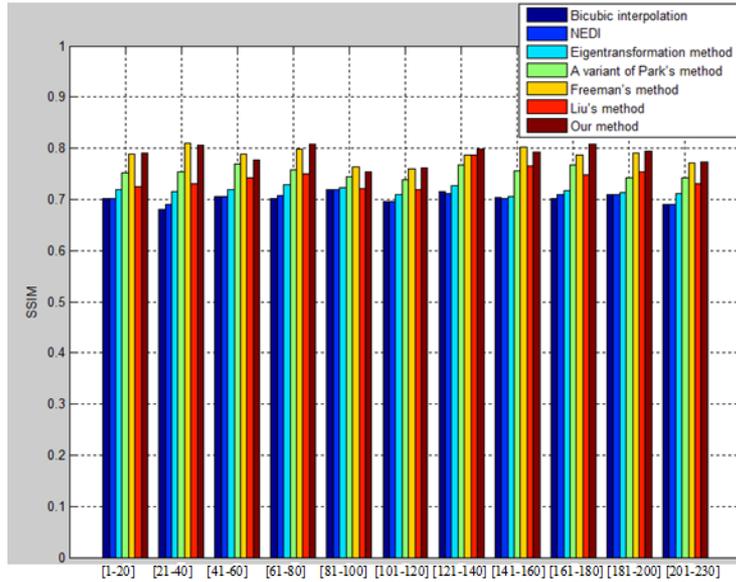
Figure 4.10: The average SSIM of the different face-hallucination algorithms on groups of faces (20 faces for each group except the last, which has 30 faces) in the database, with a magnification factor of 4.
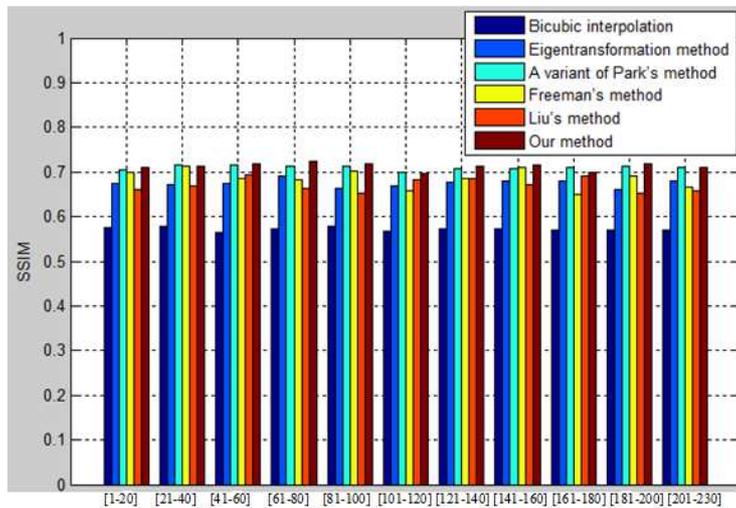


Figure 4.11: The average SSIM of the different face-hallucination algorithms on groups of faces (20 faces for each group except the last, which has 30 faces) in the database, with a magnification factor of 6.

# 4.5 Conclusion and discussion

In this chapter, a novel face-hallucination scheme based on a simple mapping model is proposed. The superior performance of our algorithm is due to the use of three different stages in estimating a HR face from a LR face. In the first stage, the optical-flow method is used to align the facial images. Therefore, the effects of warping errors can be reduced by using adaptive weighting in the local prior model, and the method can infer comparably more faithful individual facial structures of the target HR faces. We have also proved that the major singular values of the associated eigenvectors of the same image at different resolutions are proportional to the magnification factor. In the second stage, bicubic interpolation is applied to the matrices decomposed from the input images by SVD. This method retains the holistic structure of face images, and the learned mapping matrices, which are represented as embedding coefficients of the mapping matrices derived using an iterative method, can be seen as holistic constraints in the reconstruction of HR images. Finally, the residual matrix is added, which contains the missing high-frequency information and details required for face hallucination. Compared to typical state-of-the-art algorithms, experiments show that our proposed method is practicable and can produce plausible HR images with both a holistic structure and high-frequency details.

# Chapter 5 Simultaneous Hallucination and Recognition of Low-resolution Faces Based on SVD

Face recognition is an important task for video surveillance. The two primary tasks for face recognition are face identification and verification. In face identification, a query face is compared to the gallery faces in a dataset so as to identify its identity. For face verification, the claimed identity of a query face is verified. Face recognition can achieve a highly accurate performance under controlled conditions, such as under frontal light sources, frontal view, no occlusion, neutral facial expression, etc. However, low-resolution (LR) faces are a difficult problem in the face-recognition domain. Although current digital cameras can capture images at high resolution, face images captured in outdoor circumstances and at a distance, with a compressed video format, are usually of low resolution and low quality. To achieve effective video surveillance, both face hallucination and face recognition are needed simultaneously.

## 5.1 Related Work

In the image-processing research field, reconstruction of a high-resolution (HR) image from its LR inputs is called super-resolution (SR). For face images, this technology is also called face hallucination [8], which has become one of the most important fields of face recognition. A lot of face-hallucination methods has described in the Section 4.1.

Currently, most existing methods focus only on reconstruction, and seldom consider face recognition and hallucination simultaneously. In [123], an approach for simultaneous face SR and feature extraction for LR face verification was proposed. This approach simultaneously provides fitness measures of the SR results from both the reconstruction and recognition perspectives. In [125], Zou and Yuen proposed an approach to learn the relationship between the HR image space and the

very low-resolution (VLR) image space for face SR. The proposed discriminative SR (DSR) method, with a discriminative constraint, is used to learn the proper relationship, based on class information, for face-recognition applications.

In contrast to the previous works, we propose a novel and efficient scheme for the simultaneous hallucination and recognition of LR face images via singular value decomposition (SVD) and the LR-HR mapping model for the SVD representation. In our approach, face images are represented using SVD, and the hallucination and recognition of LR faces are taken into account simultaneously. We have proved [124] that, based on the Frobenius norm, the corresponding singular values of an image across different resolutions have approximately a linear relationship. This makes the estimation of the singular values of HR face images more reliable. Furthermore, the left and right matrices in the SVD representation can be interpolated to a higher resolution using bicubic interpolation; this interpolation method applied to these two matrices will not change the holistic structure or the pattern of the face image. Our proposed approach can be applied to both face verification and identification.

Our simultaneous face-hallucination and -verification algorithm is denoted as SHV, As the claimed identity of the query is known, the claim will simply be rejected if the difference between the singular values of the query and those of the claimed faces in the database is larger than a certain threshold. If the difference is smaller than the threshold, super-resolution will be performed based on the mapping models learned from the claimed LR-HR face pairs. Then, hallucination is performed based on the LR-HR face pairs of the claimed identity, and the LR-to-HR mapping matrices of the respective claimed face pairs are learned for estimating the high-frequency information or any detailed local features missed in the estimated HR faces generated by interpolating the two SVD matrices. The hallucinated faces are then used for verification again On the other hand, the algorithm for simultaneous face hallucination and identification is denoted as SHI. In this algorithm, $Q$ faces that are the most similar to the input LR face image are first searched from a gallery database of LR-HR pairs based on its singular values. Suppose that these $Q$ faces belong to $M$ distinct subjects, where $M < Q$. For each of

these *M* identities, the corresponding mapping models are learned and used to super-resolve the query input. Therefore, *M* HR face images for the LR query are generated. Then, the differences between each of the *M* HR face images and the corresponding HR face images in the database are computed based on PCA. The input LR face is assigned to the class of the face with the smallest difference. Fig. 5.1 illustrates the proposed framework for simultaneous hallucination and identification of LR faces based on SVD and the LR-to-HR mapping models.

As is well known, for a novel face which is significantly different from the training samples, most of the existing learning-based face-hallucination methods will likely produce artifacts and discontinuities in the reconstruction results. The face-recognition steps in the proposed approaches have the advantage that the abovementioned problem can be solved, to a certain extent. If the referred faces do not have similar holistic structures and patterns to the LR input, these faces will be rejected during face recognition. Concurrently, with the aid of face recognition, the estimation of the high-frequency details of the HR face images will become more reliable and effective. Experimental results show that our algorithm is effective and can produce promising hallucination results.

The rest of the chapter is organized as follows. First, we will present a simple mapping model for hallucinating SVD matrices. Then we introduce our proposed face-hallucination and -recognition scheme. An the end, experimental results are presented and a conclusion of this part is given, respectively.
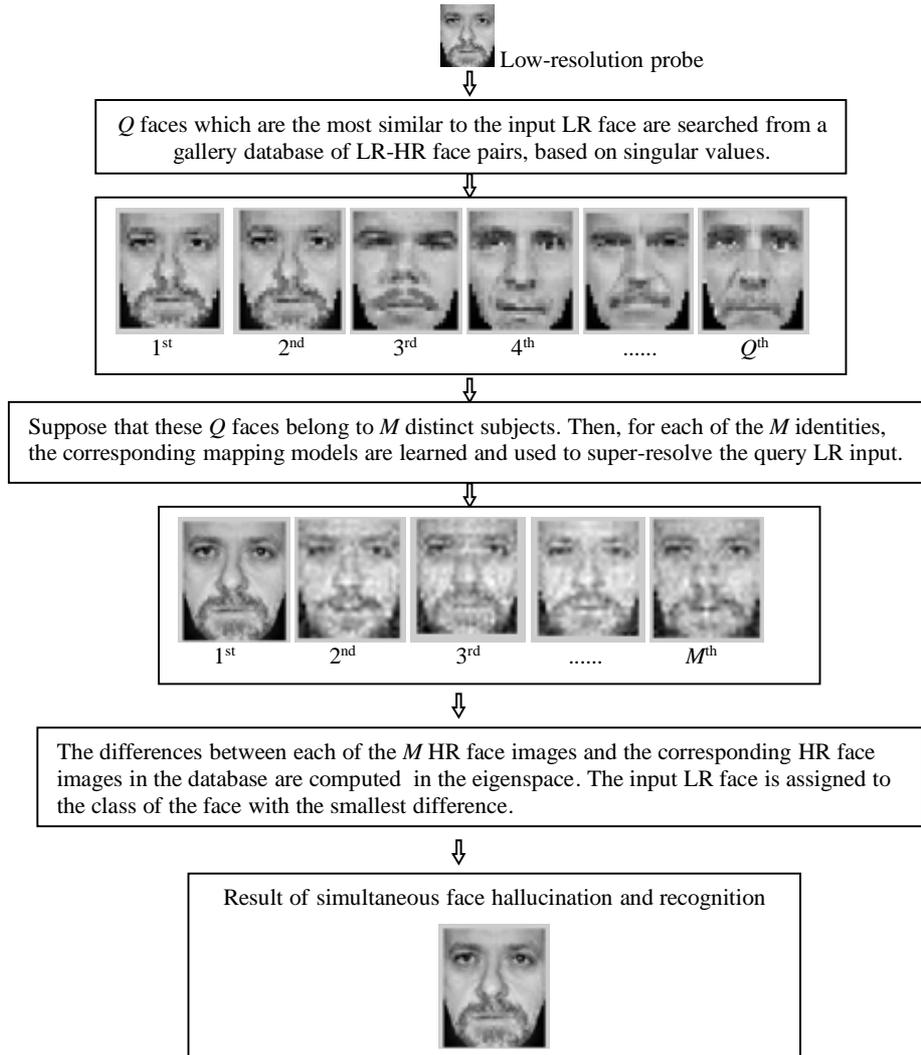
Fig. 5.1. The proposed framework for the simultaneous hallucination and identification of LR faces, based on SVD and a mapping-model method.

## 5.2 SVD of Face Images

In the section 4.2, a mathematical framework using SVD has described to achieve an effective image representation for face hallucination. In simple words, the main singular values of the same image under different resolutions are approximately proportional to each other, with the magnification factor $\alpha$ as the proportional constant [124, 127].

By linear algebra, the spectral norm (i.e. the Euclidean norm) of the matrix $A$ is the largest singular value of $A$, i.e.

$$\| A \|_2 = w_1. \tag{5.1}$$

For image $I_h$ and the original LR image $I_l$, we have:

$$\| I_h \|_2 = w_h^1, \text{ and} \tag{5.2}$$

$$\| I_l \|_2 = w_l^1. \tag{5.3}$$

Since the image $I_h$, with a magnification factor of $\alpha$, has similar values to the corresponding pixel in the original LR image $I_l$, we have $\|I_h\|_2 \cong \|\alpha I_l\|_2 = a\|I_l\|_2$. This is equivalent to:

$$w_h^1 \cong \alpha w_l^1, \tag{5.4}$$

Then, the linear relationship becomes:

$$\sum_{i=2}^{k} \left(w_h^i\right)^2 \cong \alpha^2 \sum_{i=2}^{k} \left(w_l^i\right)^2. \tag{5.5}$$

Now, we can set $w_h^1 = 0$ and $w_l^1 = 0$, the new image $I_h^{'}$ and $I_l^{'}$ can be expressed as follows:

$$I_h^{'} = U_h \begin{pmatrix} 0 & & & & \\ & w_h^2 & & & \\ & & \ddots & & \\ & & & w_h^k & \\ & & & & 0 \\ & & & & & 0 \end{pmatrix} V_h^T, \text{ and} \tag{5.6}$$

$$I_l^{'} = U_l \begin{pmatrix} 0 & & & & \\ & w_l^2 & & & \\ & & \ddots & & \\ & & & w_l^k & \\ & & & & 0 \\ & & & & & 0 \end{pmatrix} V_l^T. \tag{5.7}$$

The Euclidean norms of the matrix $I_h^{'}$ and $I_l^{'}$ are their corresponding largest singular values, as follows:

$$\| I_h^{'} \|_2 = w_h^2, \text{ and} \tag{5.8}$$

$$\| I_l^{'} \|_2 = w_l^2 . \tag{5.9}$$

Similarly, since $\| I_h^{'} \|_2 \cong \| \alpha I_l^{'} \|_2 = \alpha \| I_l^{'} \|_2$, we have:

$$w_h^2 \cong \alpha w_l^2 . \tag{5.10}$$

Therefore, the linear relationship becomes:

$$\sum_{i=3}^{k} \left( w_h^i \right)^2 \cong \alpha^2 \sum_{i=3}^{k} \left( w_l^i \right)^2 . \tag{5.11}$$

Using the principle of mathematical induction, we can also set $w_h^1 = 0$, $w_h^2 = 0$ and $w_l^1 = 0$, $w_l^2 = 0$. Thus, we have:

$$w_h^3 \cong \alpha w_l^3, \text{ until} \tag{5.12}$$

$$w_h^k \cong \alpha w_l^k . \tag{5.13}$$

This proves that the leading singular values of the same image under different resolutions are approximately proportional to each other with the magnification factor $\alpha$. In addition, since the Euclidean norm of the matrix $A$ is the largest singular value of $A$ [124, 127], and we can utilize the largest singular value $w_h^1$ of $I_h$ and the largest singular value $w_l^1$ of the original LR image $I_l$, respectively, to normalize the global feature to form scale-invariant feature vectors as follows:

$$s_h' = s_h / w_h^1, \text{ and } s_l' = s_l / w_l^1. \tag{5.14}$$

We can see that $s_h' = s_l'$, which means that the singular values are normalized so that face images at different resolutions can be compared directly [127].

# 5.3 Simultaneous Hallucination and Verification/Recognition of Low-resolution Faces

For the scheme on simultaneous face hallucination and face recognition, we consider the algorithms SHV and SHI separately. The major difference between

these two algorithms is in the selection of the LR-HR training pairs when performing face super-resolution for face verification and identification.

### 5.3.1 Simultaneous face hallucination and verification (SHV)

Given a pair of face images, the task of face verification is to verify whether they belong to the same person identity or not. In our framework for simultaneous face hallucination and verification, we will first use the singular values of face images as a global scale-invariant feature vector for their representation according to Equation (5.14). We define a similarity function, $SIM$, to measure the similarity between the input query and the claimed identity in the gallery database, based on singular values only, as follows:

$$SIM(I_1, I_2) = 1 / \left\| s'_{I_1} - s'_{I_2} \right\|_2 , \qquad (5.15)$$

where $\| \ \|_2$ is the $L_2$ norm, and $s'_{I_1}$ and $s'_{I_2}$ are the normalized leading singular values utilizing the largest singular value, i.e. using Equation (5.14), of the input LR probe and the claimed identity in the gallery database, respectively. If the difference is larger than the threshold, i.e. $SIM \leq T_1$, the claimed face will be rejected. If the difference between the singular values of the query and the claimed identity is smaller than a certain threshold, i.e. $SIM > T_1$, super-resolution will be performed based on the mapping models learned from the claimed LR-HR face pairs. And the threshold $T_1$ is set empirically by experiments. Then, the face hallucination of the input LR face, only using the claimed face pairs as references for estimation of the mapping functions for interpolating the two matrices in the SVD representation, is described in the following.

Suppose that $N$ face images of the claimed identity in the gallery set can pass the face-verification process based on singular values, and face images in the gallery are $\alpha \times \alpha$ times of the input LR face image $I_l$. These $N$ face images are down-sampled to form $N$ pairs of LR and HR training face images, denoted as $I_l^i$ and $I_h^i$ ($i = 1, ..., N$), respectively. The LR training images should have a high structural similarity to the LR input face after alignment. Based on the $N$ face pairs, $N$

corresponding HR face images can be reconstructed from the LR face using the mapping model scheme, as described in Section 2.2 in Chapter 2.

With the aid of face verification, this will make the reconstructed HR image more reliable and accurate. In addition, another merit of our proposed algorithm is that face verification can help prevent from producing artifacts and discontinuities during the stage of face hallucination. Based on the reconstructed HR face images, combine with Potential-Field Representation of HR face images [127], face verification can be conducted based on PCA. If the query input is the claimed identity, the mapping models learned should be correct and effective for the reconstruction. High verification accuracy can be achieved. Otherwise, i.e. the query is not the claimed identity, the accuracy should be degraded.

**5.3.2 Simultaneous face hallucination and identification (SHI)**

For the scheme of simultaneous face hallucination and identification, $Q$ faces that are the most similar to the input LR face image are first searched from a gallery database of LR-HR pairs based on its singular values [124], as shown in Fig. 5.1. Suppose that these $Q$ faces belong to $M$ distinct subjects, where $M < Q$, and for each of the $M$ identities, the corresponding mapping models are learned and used to super-resolve the query input. Therefore, $M$ hallucinated HR face images $\overline{I}_h^j$, $j = 1, 2, ..., M,$ for the LR query are generated using the face hallucination algorithm introduced in Section 2.2. Then, the differences between each of the $M$ hallucinated HR face images and the corresponding HR face images of the $j^{\text{th}}$ distinct subject, $j = 1, 2, ..., M,$ in the gallery database are computed using the eigenface method [3].

Fig. 5.1 illustrates the proposed framework for simultaneous hallucination and identification of LR faces. In our framework, the targeted HR $U$ and $V$ matrices of the query are computed using the mapping models [124] learned from the LR-HR pairs of the respective $M$ distinct subjects. If the query and the identities of the subject under consideration are of the same person, the mapping models should be correct, and hence the HR query face image generated should resemble the corresponding HR face images in the database. Otherwise, the reconstructed HR

face is unlikely to be similar to the HR face images of different subjects in the gallery database. By considering face hallucination and identification simultaneously, both the face-hallucination and -identification performances can be improved.

# 5.4 Experimental results

In order to verify the effectiveness of the proposed schemes, the combined dataset used in [27, 124] is used to evaluate the performance of our proposed framework. The facial images in the dataset were selected from the GT [35], AR [36], and FERET [37] databases, which contain 40, 70, and 500 persons. Five images from each class with a near-frontal view, neutral expressions, and different illuminations are randomly chosen for the experiments. Thus, the total number of images in the database is 610×5 = 3,050 images. All the facial images are well aligned based on the position of the two eyes, using the method in [128]. The parameters $\lambda$ and $\sigma$ are empirically set at 0.001 and 50, respectively. Experiments show that using all of the above settings can achieve a satisfactory overall performance. A number of experiments were conducted to verify the effectiveness of our schemes. Our proposed SHV and SHI schemes will be evaluated in Section 5.4.1 and Section 5.4.2, respectively.

**5.4.1 Experiments on Simultaneous Face Hallucination and Verification (SHV)**

The task of face verification is to determine whether a pair of face images belong to the same person. In the experiments, the database consists of 3,050 face images of 610 distinct subjects, the original HR facial images are cropped to a size of 72×64 pixels, and the LR faces are of size 18×16 pixels. We followed the standard 10-fold cross-validation over the combined database in the experiments for face verification. The similarity function *SIM* defined in Equation (5.15) is used to measure the similarity between an input query and the claimed identity in the gallery database for face verification.

In this section, we evaluate the effectiveness of the proposed face-verification approach. The PCA-based algorithm [3] (also known as eigenfaces) is a benchmark

for appearance-based and image-based face recognition/verification approaches [1]. Therefore, it is used in these experiments to illustrate the effectiveness of our algorithm. As in [33, 34], the $L_1$ norm distance metric is used, which is a more suitable distance measure than the Euclidean distance metric ( $L_2$ ) for PCA-based algorithms. The state-of-the-art algorithm proposed in [123] for recognition of low-resolution faces, namely $S^2R^2$, is employed for comparison. Furthermore, two more state-of-the-art methods, namely the discriminative constraint-based DSR method [125], which employs the class-label information, and the coupled kernel embedding (CKE) feature-extraction method [126], are also compared for LR face recognition. For the DSR method [125], images from 610 persons (one per person) are randomly selected to form the training set, and the rest of the images form the testing (probe) set. As in [125], the training pairs are clustered using linearity clustering, so that the relationship between the data pairs in each cluster can be linearly approximated. Following [126], 610 images are also randomly selected from the database, and projection directions are train using the Gaussian-kernel-based CKE algorithm. The kernel parameter is set at 3, and 40 features in the embedding space are extracted for matching.

Figure 5.2 shows the identification accuracy (IDA), which is the percentage of the probes that are correctly identified by an algorithm. As shown in Fig. 5.2, the PCA-based algorithm achieves an IDA of 32.36% only. The performance of the $S^2R^2$ algorithm can further improve the IDA significantly, to 46.05%. Both the DSR and CKE methods produce good recognition results compared to that of the PCA-based algorithm, which are 53.22% and 52.35%, respectively. Our proposed SHV method can increase the IDA to 54.11%, i.e. achieving a better performance than both the DSR-based algorithm and the CKE algorithm. These experimental results prove that our proposed scheme is effective.
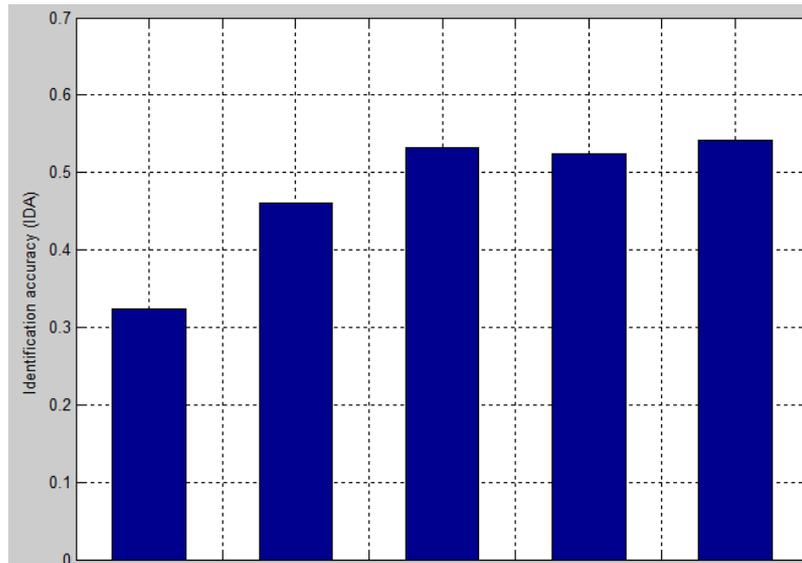
Eigenfaces[3]   S$^2$R$^2$ [123]  DSR[125]   CKE [126]   Our SHV method

Figure 5.2. Identification accuracy (IDA) for different algorithms.

The receiver operating characteristics (ROC) curve, which is constructed from the true-positive rate (TPR) and the false-positive rate (FPR) by changing the threshold, is shown in Figure 5.3. The ROC characteristics show that the proposed SHV method outperforms the other, state-of-the-art methods. In addition, each ROC curve can be summarized by the area under the ROC curve (AUC). Therefore, AUC is also used as a quantitative measure for the evaluation of different algorithms. Table 5.1 shows that our proposed method is superior in terms of AUC, compared to other algorithms.
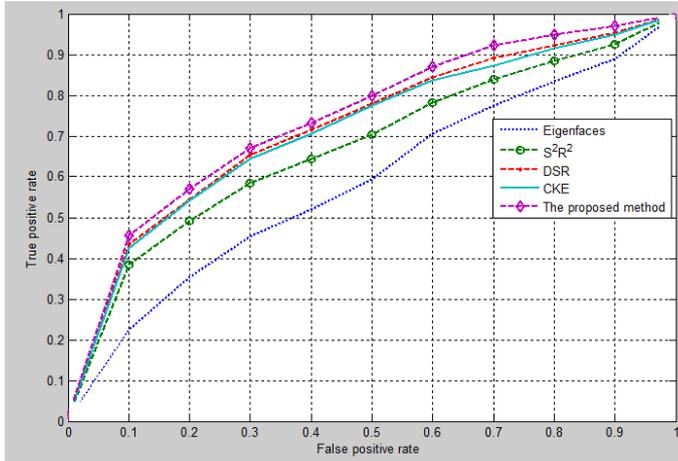
Figure 5.3. Receiver operating characteristics (ROC) curve for the different algorithms.

Table 5.1. AUC of five different scheme.

|  | Eigenfaces | $S^2R^2$ | DSR | CKE | Our SHI method |
|---|---|---|---|---|---|
| AUC | 0.6595 | 0.7471 | 0.7763 | 0.7649 | 0.7954 |

The input LR face can be hallucinated by using the respective mapping models for the matrices $U$ and $V$ based on the LR-HR face pairs of the claimed identity. For comparison, two interpolation algorithms, namely the bicubic-interpolation algorithm [40] and the edge-directed interpolation (NEDI) method [41], are applied for face hallucination. Five more state-of-the-art face-hallucination algorithms are also compared to our proposed method: two PCA-based, holistic SR methods (the eigentransformation method [20] and a variant of Park's method [21]); one patch-based method (Freeman's method [15]); Liu's method [22] based on a global parametric model and a local nonparametric model; and a SVD mapping-based method [124]. The version of Park's method used in this chapter is different from the original algorithm in [21] in that the training images are warped with reference to the LR input face rather than with reference to a predefined reference face.

In the experiments, we evaluate all the methods by reconstructing the HR facial images with a magnification factor of 4. All the testing images of resolution 18×16 pixels are evaluated using the "leave-one-out" approach. Two objective quality

measures, PSNR (peak signal-to-noise ratio) and SSIM [39], are used to evaluate the performances of the different methods.

Fig. 5.4 shows some samples of the reconstruction results generated using the different state-of-the-art face-hallucination algorithms, with a magnification factor $\alpha=4$. It can be seen from Fig. 5.4(b) that the bicubic-interpolation algorithm produces the blurriest results. The results in Fig. 5.4(c) are generated using the NEDI method. However, if a face image has a very low resolution (i.e. below $18\times16$ pixels), the NEDI method struggles to distinguish edges, and hence also produces blurry results as compared to the other SR methods. Figs. 5.4(d) and 5.4(e) are the results generated by using the eigentransformation method and the variant of Park's method, which are both holistic/global face-hallucination methods. Plausible face structures can be well inferred in the resulting HR images, and show a better visual quality than both the bicubic-interpolation and the NEDI algorithms. Nevertheless, as the method is purely holistic, it cannot effectively reconstruct the fine individual facial details of those novel testing faces. If a testing face image is very different from the faces in the database, infidelity will result in the reconstructed HR faces. Park's method employs the morphable face model to capture the shape variations of novel testing faces, so it outperforms the eigentransformation method. However, the HR textures are still reconstructed in a holistic manner, like the eigentransformation method. The face-hallucination results using the patch-based SR methods are illustrated in Fig. 5.4(f). It can be seen that Freeman's method can provide plausible HR facial images with sharp edges and corners. However, as some of the patches are badly matched or conflict with adjacent ones, some structural errors and discontinuities appear in the reconstructed HR images; these errors are the drawbacks of most patch-based SR methods. Furthermore, patch-based SR usually requires a large number of image-patch pairs for learning, and therefore is computationally expensive. In addition, there are artifacts in the reconstructed HR images, as shown in Fig. 5.4(f). Compared with the holistic-based and the patch-based approaches, Liu's method [22] utilizes both global and local prior information through a global parametric model and a local nonparametric model. Thus, as shown in Fig. 5.4(g), the method can produce not

only visually plausible face structures, but also fine details or textures like those in the HR training images. However, some parts of the hallucinated faces, such as the mouth, are somewhat different from the original face. This can be partially attributed to the properties of the PCA-based global model used in this approach.



    (a)        (b)        (c)        (d)        (e)        (f)        (g)        (h)        (i)
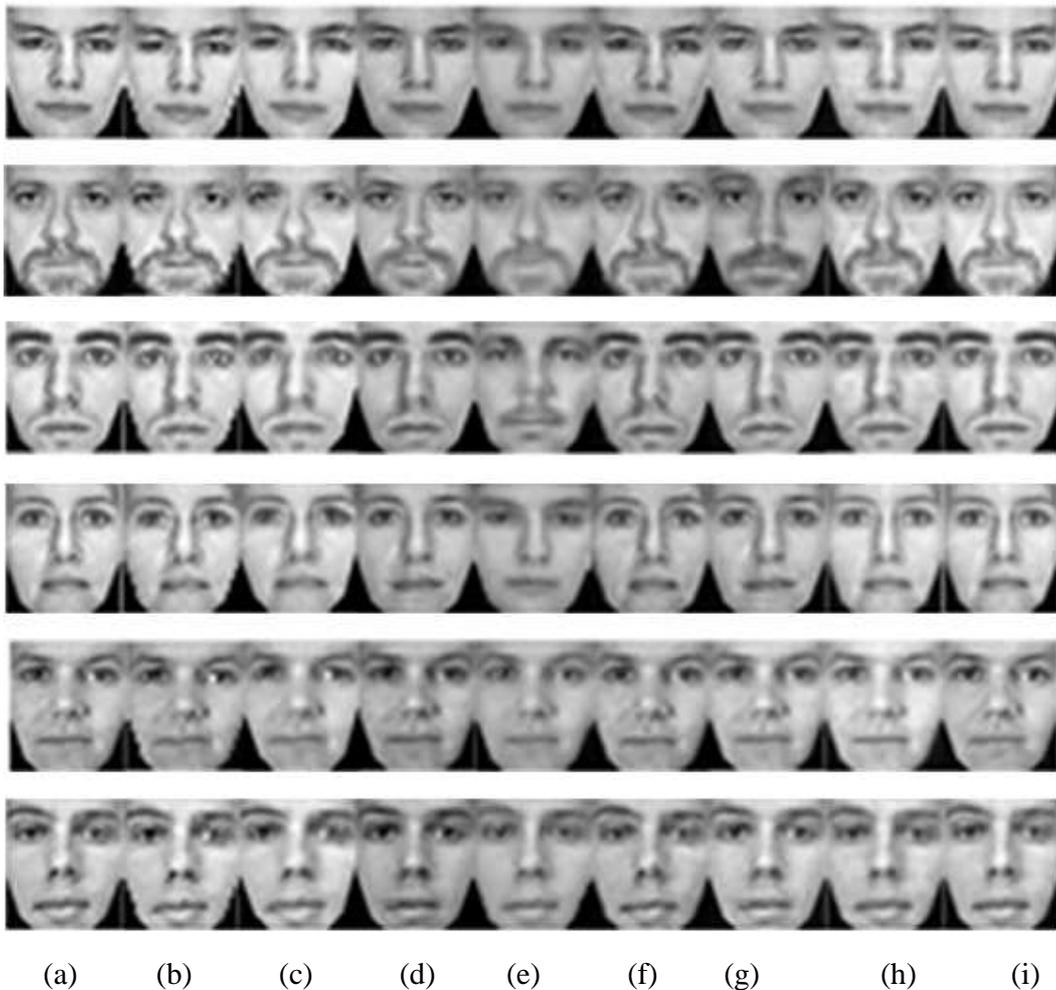
Figure 5.4: Face-hallucination results reconstructed using different methods with a magnification factor of 4: (a) the original images, (b) bicubic interpolation, (c) NEDI, (d) the eigentransformation method, (e) a variant of Park's method, (f) Freeman's method, (g) Liu's method, (h) the SVD mapping-based method, and (i) our proposed SHV method.

Unlike Liu's method, our proposed framework can retain the main energy of LR faces, meanwhile the mapping scheme and the residual matrix can produce high-resolution results affectively with clear details, as shown in Fig. 5.4(h). In this chapter, our proposed framework for simultaneous verification and hallucination of LR faces can select similar holistic structures and patterns with the LR input during the stage of face verification. It has the advantage of preventing the algorithm from producing artifacts and discontinuities in the reconstruction results, and the estimation of the high-frequency details of the HR face images will become more reliable and effective. As can be seen from Fig. 5.4(i), plausible HR images with a holistic structure and more details with a better visual quality can be obtained.

Table 5.2 tabulates the average PSNR and SSIM of the different methods with a magnification factor of 4. The results show that our method is superior, in terms of both the two measurements, to the other, state-of-the-art algorithms.

Table 5.2. The average PSNR and SSIM of the different face-hallucination algorithms with a magnification factor of 4.

| Face-hallucination algorithms | PSNR(dB) | SSIM |
|---|---|---|
| Bicubic interpolation | 19.32 | 0.5507 |
| NEDI | 19.58 | 0.5585 |
| Eigentransformation method | 21.58 | 0.6392 |
| A variant of Park's method | 22.07 | 0.6410 |
| Freeman's method | 22.34 | 0.6515 |
| Liu's method | 21.86 | 0.6443 |
| Mapping method | 22.69 | 0.6532 |
| Our proposed SHV method. | 22.72 | 0.6627 |
| Our proposed SHI method. | 22.83 | 0.6685 |

**5.4.2 Experiments on Simultaneous Face Hallucination and Identification (SHI)**

In this section, extensive experiments were performed to evaluate the effectiveness of the proposed simultaneous face hallucination and identification

(SHI) algorithm. In the experiments, LR faces of three different resolutions, 18×16, 16×14, and 14×12 pixels are considered. The number of searched faces $Q$ is set at 40, i.e. $Q = 40$ faces that are the most similar to the input LR face are searched from a gallery database of LR-HR face pairs based on singular values using Equation (11). A face of each subject is randomly selected in the database to form a gallery dataset, while the remaining faces are used for testing in the experiments. Eigenfaces are used to measure the differences between each of the $M$ hallucinated HR face images and the corresponding HR face images in the gallery dataset.

For comparison, the state-of-the-art methods used in Section 5.4.1 are also performed for LR face recognition. Table 5.2 tabulates the average recognition rates of the different state-of-the-art algorithms for the LR faces of three different resolutions. We can see that, for LR faces with 18×16 pixels, the average recognition rate of the PCA-based algorithm [3] is 39.44% only. The average recognition rate of the $S^2R^2$ algorithm [123] achieves 55.70%. Both the DSR[125] and CKE[126] can further improve the average recognition rate significantly to 71.66% and 71.24%%, respectively. Our proposed SHI method can achieve the best average recognition rate of 72.15%.

When the LR faces are reduced to the resolution of 16×14 and 14×12 pixels, the recognition performances of all the methods degrade. For the resolution of 16×14 pixels, the DSR [125] and CKE [126] methods achieve recognition rates of 70.85% and 70.32%, respectively; this is much better than that of the PCA-based algorithm [3], which is 37.11%. The average recognition rate of our proposed SHI method is 71.33%. When the resolution is 14×12 pixels, the PCA-based algorithm [3], the DSR [125] method, and CKE [126] method achieves recognition rates of 34.06%, 69.19%, and 68.85%, respectively. Our proposed SHI method outperforms all these methods, and achieve a recognition rates of 69.50%; this shows the effectiveness of our proposed method.

Table 5.3. The average recognition rates of five different face recognition schemes with the LR faces of sizes 18×16 pixels and 16×14 pixels, respectively.

| Average recognition rates | Eigenfaces | $S^2R^2$ | DSR | CKE | Our SHI method |
|---|---|---|---|---|---|
| 18×16 | 39.44% | 55.70% | 71.66% | 71.24% | 72.15% |
| 16×14 | 37.11% | 53.05% | 70.85% | 70.32% | 71.33% |
| 14×12 | 34. 06% | 51.88% | 69.19% | 68.85% | 69.50% |

Fig. 5.5, Fig. 5.6, and Fig. 5.7 show the recognition performance of the different methods in terms of the cumulative matching characteristic (CMC) curve, which evaluates the ranking capability of an identification algorithm, when the resolutions of the LR faces are 18×16, 16×14, and 14×12 pixels, respectively. From these results, we can see that our proposed SHI method outperforms the other state-of-the-art algorithms.
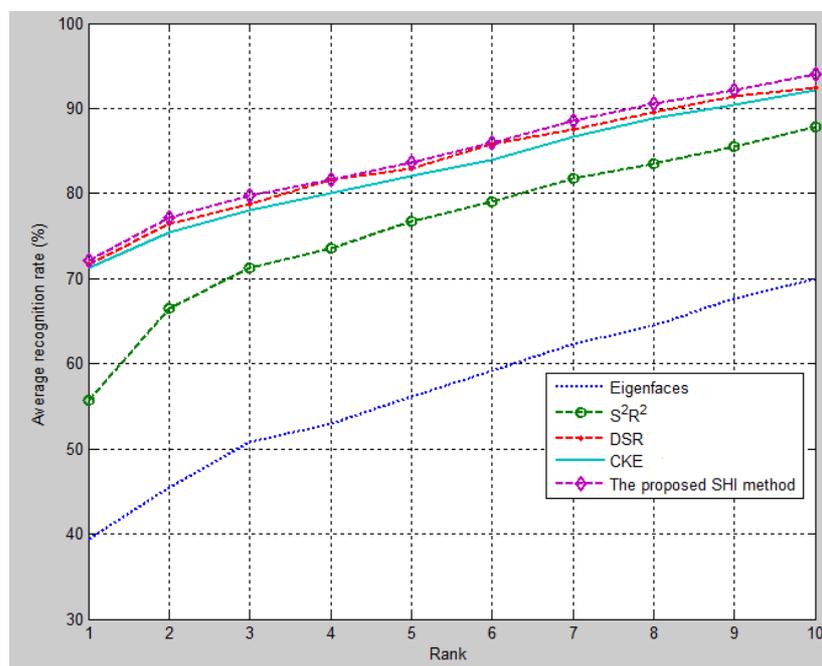


Figure 5.5. The cumulative matching characteristic (CMC) curves of the different methods with the LR faces of size 18×16 pixels.
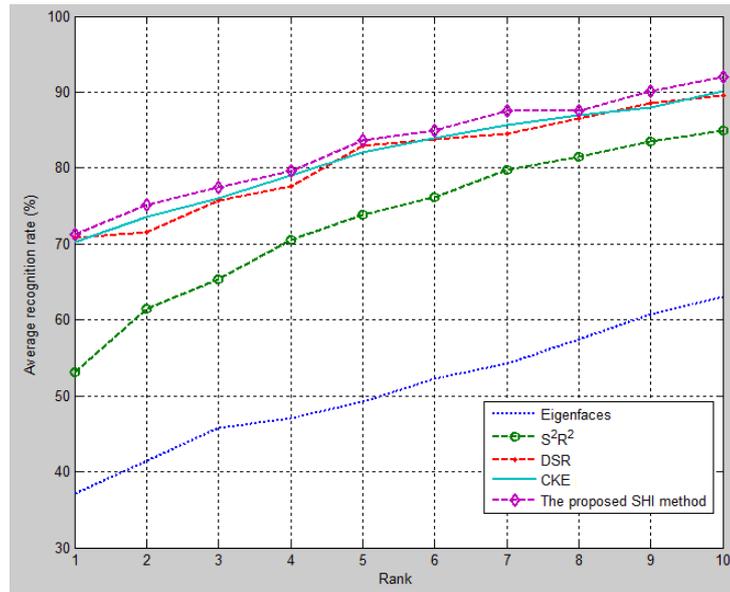
Figure 5.6. The cumulative matching characteristic (CMC) curves of different methods with the LR faces of size 16×14 pixels.
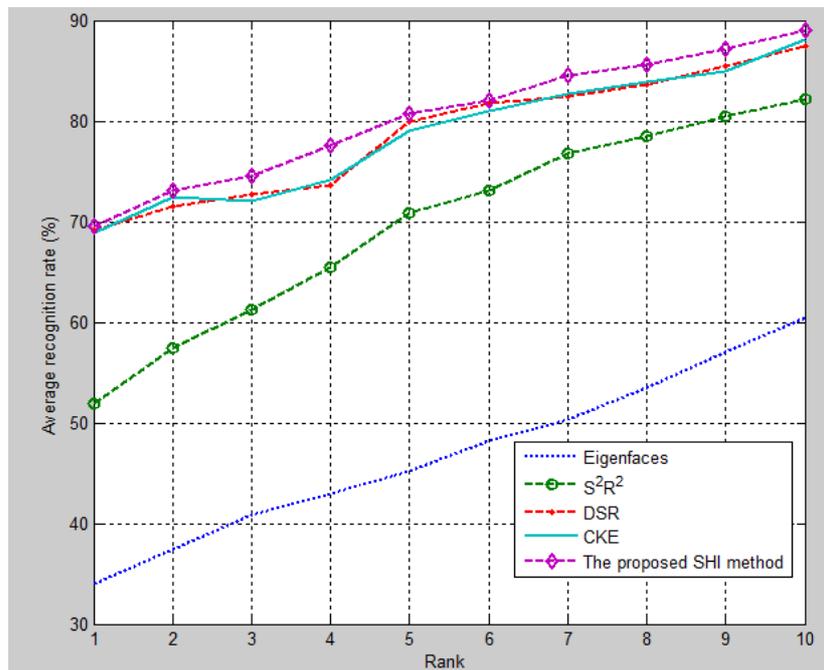


Figure 5.7. The cumulative matching characteristic (CMC) curves of the different methods with the LR faces of size 14×12 pixels.

For simultaneous face hallucination, Table 5.2 also tabulates the average PSNR and SSIM of the proposed simultaneous face-hallucination and -identification (SHI) method with a magnification factor of 4. From Table 5.2, it can be seen that our SHI method is superior to the different state-of-the-art methods, in term of the average PSNR and SSIM. Table 5.3 and Figs. 5.5, 5.6, and 5.7 show that our proposed SHI method is superior in terms of the recognition rate and the CMC curve compared to other, state-of-the-art face-identification algorithms. These experimental results prove that our proposed scheme of simultaneous face hallucination and identification is effective, and can achieve excellent performance.

# 5.5 Conclusion and discussion

A novel approach for simultaneous hallucination and recognition of LR faces has been proposed. In our framework, hallucination and recognition of LR faces are taken into account simultaneously. Our proposed scheme can retain the holistic structure and the high-frequency details of face images, and outperforms other, state-of-the-art algorithms in terms of PSNR and SSIM. Experiments have shown that our proposed method is practicable and can produce plausible HR images with both a holistic structure and high-frequency details. Meanwhile, experiment results also demonstrate that our proposed simultaneous framework can achieve superior results for both face verification and identification.

# Chapter 6 Conclusion and future work

In this thesis, we have firstly introduced face recognition and some of its current typical challenging problems. Although these typical challenging problems have attracted increasing attention in recent years, they are still ongoing and tough issues for robust face recognition.

Concerning Chapter 2 to Chapter 5: we have presented the accurate eye-detection problem in Chapter 2, the illumination-variation issue in Chapter 3, the low-resolution problem in Chapter 4, and the simultaneous hallucination and recognition of LR faces in Chapter 5. In this final chapter, we will summarize the main contributions of this research, and discuss some possible future research directions based on the work presented in the previous chapters.

## 6.1 Summary and conclusions

This thesis concentrates on some typical challenging problems for current face recognition. We have surveyed three fields and the work described in this thesis involves the following research:

(1) the accurate eye-detection problem in Chapter 2. Since accurate eye detection is an important problem for robust face recognition, in this chapter, an efficient hierarchical scheme is proposed for accurate facial-feature detection and localization. The proposed algorithm, which is non-iterative and computationally simple, achieves a superior performance compared to other state-of-the-art methods.

(2) the illumination-variation issue in Chapter 3. In this chapter, we introduces a facial-image lighting-compensation and -enhancement algorithm for face recognition. The advantage of the proposed method is that the assumption of a single-point light source is not required, so it circumvents and overcomes the limitations of the Lambertian model and is also suitable for outdoor circumstances.

(3) the low-resolution problem in Chapter 4. We presented a novel face-hallucination scheme based on singular value decomposition, which is effective in terms of producing plausible HR images with both a holistic structure and high-frequency details.

(4) the issue of simultaneous hallucination and recognition of LR faces in Chapter 5. In this chapter, a novel approach for simultaneous hallucination and recognition of LR faces has been proposed. In our framework, hallucination and recognition of LR faces are taken into account simultaneously. Our proposed scheme can retain the holistic structure and the high-frequency details of face images, and outperforms other, state-of-the-art algorithms in terms of PSNR and SSIM. Experiments have shown that our proposed method is practicable and can produce plausible HR images with both a holistic structure and high-frequency details. Meanwhile, experiment results also demonstrate that our proposed simultaneous framework can achieve superior results for both face verification and identification.

# 6.2 The new contributions in this thesis

In this thesis, we have made a number of new contributions to the research community of current face recognition. It is believed that this thesis contains the following original work and new contributions:

- Chapter 2 - The accurate eye-detection problem for robust face recognition: an efficient hierarchical scheme, which is robust to illumination and pose variations in face images, is proposed for accurate facial-feature detection and localization. In our algorithm, having detected a face region using a face detector, a wavelet-based saliency map − which can reflect the most visually meaningful regions − is computed on the detected face region. As the eye region always has the most variations in a face image, the coarse eye region can be reliably located based on the saliency map, and verified by means of principal component analysis. This step in the proposed hierarchical scheme narrows down the search space, thereby reducing the computational cost in the further precise localization of the two eye

positions based on a pose-adapted eye template. Moreover, among the facial features, the eyes play the most important role, and their positions can be used as an approximate geometric reference to localize the other facial features. Therefore, localization of the nose and mouth can be determined by using the saliency values in the saliency map and the detected eye positions as geometric references. Our proposed algorithm is non-iterative and computationally simple. Experimental results show that our algorithm can achieve a superior performance compared to other state-of-the-art methods.

- Chapter 3 - The illumination-variation issue: For this issue, we propose a novel scheme for generating illumination-invariant face images using illumination compensation and enhancement, which is applied to face recognition. It is reasonable to assume that the variations of the face surface-reflectance representation matrix, which reflects the intrinsic property of a face surface of the same person under different illumination conditions are small, while the dissimilarity between images of the same person under different illumination conditions is mainly caused by the differences in the illumination-effect matrix. In this chapter, the proposed scheme learns the average illumination-effect matrix for face image representation under changing illumination, which can be used to compensate or enhance images, and to eliminate the effect of different and uneven illuminations while retaining the intrinsic properties of the face surfaces. The advantage of our method is that the assumption of a single-point light source is not required, so it circumvents and overcomes the limitations of the Lambertian model and is also suitable for outdoor circumstances. Experimental results have produced promising results, which demonstrate the effectiveness of our proposed method.

- Chapter 4 - The low-resolution problem: In this chapter, an efficient mapping model based on singular value decomposition (SVD) is proposed for face hallucination. We observe and prove that the main singular values of an image at one resolution have approximately linear relationships with

their counterparts at other resolutions. This makes the estimation of the singular values of the corresponding high-resolution (HR) face images from a low-resolution (LR) face image more reliable. From the signal-processing point of view, this can effectively preserve and reconstruct the dominant information in the HR face images. Interpolating the other two matrices obtained from the SVD of the LR image does not change either the primary facial structure or the pattern of the face image. The corresponding two matrices for the HR face images can be constructed in a "coarse-to-fine" manner using global reconstruction. Our proposed method retains the holistic structure of face images, while the learned mapping matrices, which are represented as embedding coefficients of the individual mapping matrices learned from LR-HR training pairs, can be seen as holistic constraints in the reconstruction of HR images. Compared to state-of-the-art algorithms, experiments show that our proposed face-hallucination scheme is effective in terms of producing plausible HR images with both a holistic structure and high-frequency details.

- Chapter 5 - The simultaneous hallucination and recognition of LR faces: A framework based on singular value decomposition (SVD) for performing both face hallucination and recognition simultaneously is proposed in this thesis. Conventionally, low-resolution (LR) face recognition is carried out by super-resolving the LR input face first, and then performing face recognition to identify the input face. By considering face hallucination and recognition simultaneously, the accuracy of both the hallucination and the recognition can be improved. In our algorithm, each face image is represented by using SVD. For each LR input face, the corresponding LR and high-resolution (HR) face-image pairs can then be selected from the face gallery. With the aid of face recognition, using the selected LR-HR pairs, the estimation of the mapping functions for interpolating the two matrices in the SVD representation of the corresponding HR face image can be more accurate. Therefore, the final estimation of the high-frequency details of the HR face images will become more reliable and effective.

Experimental results demonstrate that our proposed framework can achieve promising results for both face hallucination and recognition.

## 6.3 Future work

This thesis has presented a number of new ideas and techniques, which are just a snapshot of our on-going research undertaken in the field of robust face recognition. In this section, some directions for possible future research will be discussed. Future research may be carried out in the following fields:

(1) The pose problem: This is another typical challenging problem for face recognition at present. We will explore an accurate algorithm to reconstruct the 3D face for pose-invariant face recognition. As one of the effective technologies for capturing 3D surface information, Photometric Stereo, has attracted widespread attention. A face image can be seen as a special texture, and Photometric Stereo can capture more 3D information than traditional shape-from-shading techniques, therefore an accurate algorithm to reconstruct the 3D surface is useful and important for robust face recognition and face representation. In the future, we will focus on accurate algorithm to reconstruct the 3D face for pose-invariant face recognition.

(2) The occlusion issue: Currently, occlusion is another tough issue for robust face recognition. To continue our work, we will conduct research on effective local features for occluded face representation. We will also analyze and compare the effectiveness and efficiency of different local features to represent occluded faces; this can help to either determine or devise optimal local features to represent an occluded face for robust occlusion-insensitive face recognition.

(3) In the future, all of these techniques will then be integrated and combined with both existing and new face-recognition algorithms. We will investigate these related techniques to devise a compound framework suitable for face verification/recognition, develop efficient algorithms to alleviate the effect of uneven lighting on faces, study how to best combine multi-view face hallucination and pose-invariant face recognition, evaluate the relative performances of applying

SR and face recognition, design an accurate algorithm to reconstruct 3D faces, and investigate robust occlusion-insensitive face recognition, in order to achieve a robust and good performance level.

# Bibliography

[1]     W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face Recognition: A Literature Survey, ACM Computing Surveys, 2003, pp. 399-458.

[2]     M. Turk, A. Pentland, Face Recognition Using Eigenfaces, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1991, pp. 586-591.

[3]     M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neurosicence 3 (1) (1991) 71-86.

[4]     P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, IEEE Trans. on Pattern Analysis and Machine Intelligence 19 (1997) 711–720.

[5]     J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Face Recognition Using LDA-Based Algorithms, IEEE Trans. on Neural Networks 14 (1) (2003) 195-200.

[6]     A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Trans. on Pattern Analysis and Machine Intelligence 23 (2) (2001) 228-233.

[7]     L. Wiskott, J. M. Fellous, N. Kruger, and C.v.d. Malsburg, Face Recognition by Elastic Bunch Graph Matching, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp. 775-779, 1997.

[8]     S. Baker and T. Kanade. Hallucinating faces. In IEEE International Conference on Automatic Face and Gesture Recognition. pp. 83-88, 2000.

[9]     Y. Adini, Y. Moses, S. Ullman, Face Recognition: The Problem of Compensating for Changes in Illumination Direction, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 721-732, July 1997.

[10]    S. Shan, Y. Chang, W. Gao, et. al, Curse of mis-alignment in face recognition: problem and a novel mis-alignment learning solution, in:

international conference on Automatic Face and Gesture Recognition, 2004, pp.314-320.

[11]   P. Wang, M. Green, Q. Ji, J. Wayman, Automatic eye detection and its validation, in: IEEE conference on computer vision and pattern recognition, 2005, pp. 167-171.

[12]   Y. Dong, L. Zhen, S. Z. Li, A robust eye location method for low quality face image, in: International Joint Conference on Biometrics, 2011, pp. 1-6.

[13]   Y. Ma, X. Ding, Z. Wang, N. Wang, Robust precise eye location under probabilistic framework, in: international conference on Automatic Face and Gesture Recognition, 2004, pp. 339-344.

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

[15]   W. T. Freeman, T. R. Jones, E. C. Pasztor, Example-based superresolution, IEEE Comput. Graph. Appl. 22 (2) (2000) 56–65.

[16]   G. Dedeoglu, T. Kanade, High-zoom video hallucination by exploiting spatio-temporal regularities, in: Proceedings of the IEEE CVPR, vol. 2, 2004, pp. 151–158.

[17]   W. Fan, D. Y. Yeung, Image hallucination using neighbor embedding over visual primitive manifolds, in: Proceedings of the IEEE CVPR, 2007, pp. 1–7.

[18]   K. I. Kim, Y. Kwon, Example-based learning for single-image super-resolution, in: Proceedings of the 30th DAGM Symp. Pattern Recognition., Munich, Germany, 2008, pp. 456–465.

[19]   J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Processing 19 (11) (2010) 2861-2873.

[20]   X. Wang, X. Tang, Hallucinating face by eigentransformation, IEEE Trans. Syst., Man, Cybern. C, Appl. Rev, 35 (3) (2005) 425–434.

[21] J. S. Park, S. W. Lee, An example-based face hallucination method for single-frame, low-resolution facial images, IEEE Trans. Image Process. 17 (10) (2008) 1806–1816.

[22] C. Liu, H. Y. Shum, C. S. Zhang, Face Hallucination: Theory and Practice, IJCV 75 (1) (2007) 115–134.

[23] X. Li, K. M. Lam, et al, Example-based image super-resolution with class-specific predictors, J. Vis. Commun. Image Represent. (2009) 312–322.

[24] G. Qiu, A progressively predictive image pyramid for efficient lossless coding, IEEE Trans. Image Process. 8 (1) (1999) 109–115.

[25] G. Qiu, Interresolution look-up table for improved spatial magnification of image, J. Vis. Commun. Image Represent. 11 (4) (2000) 360–373.

[26] X. Ma, J. Zhang, C. Qi, Hallucinating face by position-patch, Pattern Recognition 43 (2010) 2224-2236.

[27] Y. Hu, K. M. Lam, G. Qiu, T. Shen, From Local Pixel Structure to Global Image Super-Resolution: A New Face Hallucination Framework, IEEE Trans. Image Process. (2011) 433-445.

[28] A. Akyol, M. Gokmen, Super-resolution reconstruction of faces by enhanced global models of shape and texture original, Pattern Recognition 45 (2012) 4103-4116.

[29] R. Epstein, P.W. Hallinan, A.L. Yuille, 5±2 eigenimages suffice: an empirical investigation of low-dimensional lighting models. in: Proceedings of the Workshop on Physics-Based Modeling in Computer Vision, 1995, pp. 108-116.

[30] Z. Zhang, Modeling Geometric Structure and Illumination Variation of a Scene from Real Images, in: Proceedings of the IEEE ICCV, 1998, pp. 1041–1046.

[31] J.Y. Dong , M. Chantler, Capture and synthesis of 3D surface texture, IJCV 62 (2005) 177–194.

[32] M. W. Jian, J. Y. Dong, Capture and fusion of 3d surface texture, Multimedia Tools and Applications 53 (1) (2011) 237-251.

[33] O. Bretsche, Linear Algebra with Applications (3rd ed.), Prentice Hall, June 28, 2004.

[34] P. Wesseling, An Introduction to Multigrid Methods. Hoboken, NJ: Wiley, 1992.

[35] Georgia Institute of Technology, GT Face Database Atlanta. Available: http://www.anefian.com/research/face_reco.htm

[36] A. M. Martinez, R. Benavente, The AR face database CVC Tech. Rep., #24, 1998.

[37] P. J. Phillips, H. Wechsler, J. Huang, P. Rauss, The FERET database and evaluation procedure for face-recognition algorithms, Image Vision Computing 16 (1998) 295–306.

[38] K.W. Wong, K. M. Lam, W. C. Siu, An efficient algorithm for human face detection and facial feature extraction under different conditions, Pattern Recognition. 34 (10) (2001) 1993–2004.

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error measurement to structural similarity, IEEE Trans. Image Process. (2004) 600-612.

[40] H. S. Hou, H. C. Andrews, Cubic spline for image interpolation and digital filtering, IEEE Transactions on Signal Processing (26) (1978) 508–517.

[41] X. Li, M. T. Orchard, New edge-directed interpolation, IEEE Trans. Image Processing (2001) 1521–1527.

[42] J. Fan and L. Wolff, Surface Curvature and Shape Reconstruction from Unknown Multiple Illumination and Integrability, Computer Vision and Image Understanding 65(2):347–359, 1997.

[43] H. Murase and S. Nayar, Visual learning and recognition of 3D objects from appearance. Int. Journal of Computer Vision, 14(1):5–25, 1995.

[44]    J. Y. Dong, M. Chantler. Capture and synthesis of 3D surface texture. International Journal of Computer Vision (IJCV), 62(1–2):177–194, 2005.

[45]    M. W. Jian and J. Y. Dong. Capture and fusion of 3d surface texture. Multimedia Tools and Applications, Volume 53, Number 1, 237-251, 2011.

[46]    D. Jacobs, P. Belhumeur, and R. Basri. Comparing images under variable illumination. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 610 – 617. IEEE, June 1998.

[47]    A. Shashua, T. Riklin-Raviv. The quotient image: Class based rerendering and recognition with varying illuminations. IEEE Trans. Pattern Anal. Machine Intell. 23 (2), 129–139, 2001

[48]    X. Xie, K.M. Lam, An efficient illumination normalization method for face recognition. Pattern Recognition Letters 27(6): 609-617, 2006.

[49]    T. Kuparinen and V. Kyrki, Optimal Reconstruction of Approximate Planar Surfaces Using Photometric Stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 12, pp. 2282–2289, 2009.

[50]    I. Kemelmacher-Shlizerman, R. Basri, 3D Face Reconstruction from a Single Image Using a Single Reference Face Shape, IEEE Trans. on Pattern Analysis and Machine Intelligence,  Volume: 33, pp. 394–405, Feb. 2011.

[51]    B. K. P. Horn, Determining lightness from an image, Comput. Graphics Image Process. 3, 277-299, 1974.

[52]    R. Gross and V. Brajovic, An Image Preprocessing Algorithms for Illumination Invariant Face Recognition, Audio- and Video-Based Biometric Person Authentication, pp.10 - 18.2003

[53]    D. Shan and R. Ward. Wavelet-based Illumination Normalization for Face Recognition. Image Processing, ICIP 2005. IEEE International Conference on, 2005, 2, pp. 954-947

[54]    A.B.J. Teoh, Y.Z. Goh, M.K.O. Goh, Illuminated Face Normalization Technique by Using Wavelet Fusion and Local Binary Patterns. 10th Intl.

Conf. on Control, Automation, Robotics and Vision Hanoi, Vietnam, 17–20 December 2008. Pp. 422-427.

[55] Y.Z. Goh, A.B.J. Teoh, M.K.O. Goh, Wavelet-based illumination invariant preprocessing in face recognition, in: Proceedings of the Congress on Image and Signal Processing, vol.3, 2008, pp.421–425.

[56] B.K.P. Horn and M. Brooks, Seeing Shape from Shading. Cambridge,Mass.: MIT Press, 1989.

[57] J.F. Canny, A Computational Approach to Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 679-698, 1986.

[58] R.M. Haralick, Digital Step Edges from Zero Crossing of Second Directional Derivatives, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 6, pp. 58-68, 1984.

[59] V. Torre and T. Poggio, On Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 147-163, 1986.

[60] K. Wong, H. Law, and P. Tsang, A System for Recognizing Human Faces, Proc. ICASSP, pp. 1,638-1,642, 1989.

[61] V. Govindaraju, D.B. Sher, R. Srihari, and S.N. Srihari, Locating Human Faces in Newspaper Photographs, Proc. CVPR 89, pp. 549-554, 1989.

[62] S. Edelman, D. Reisfeld, and Y. Yeshurun, A System for Face Recognition that Learns from Examples, Proc. European Conf. Computer Vision, S. Sandini, ed., pp. 787-791. Springer-Verlag, 1992.

[63] P. Hallinan, A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 995-999, 1994.

[64] V. BLANZ AND T. VETTER. A Morphable model for the synthesis of 3D faces. In Proceedings, SIGGRAPH 99, 187–194, 1999.

[65] W. Zhao, R. Chellappa, 2000. Illumination-insensitive face recognition using symmetric shape-from-shading. In: Proc. IEEE Conf. CVPR, Hilton Head.

[66] J. Zhao, Y. Su, D. Wang, S. Luo, 2003. Illumination ratio image: Synthesizing and recognition with varying illuminations. Pattern Recognition Lett. 24 (15), 2703–2710.

[67] X. Xie, , K.M. Lam, Face recognition under varying illumination based on a 2D face shape model. Pattern Recognition, 38(2), 221–230, 2005.

[68] Y. Moses and S. Ullman, Limitation of Non-Model-Based Recognition Schemes, Proc. ECCV-92, Springer-Verlag, pp. 820-828, 1992.

[69] M. W. Jian, K.M. Lam, J. Y. dong. Illumination Compensation and Enhancement for Face Recognition, APSIPA Annual Summit and Conference 2011, Xi'an, 2011.

[70] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, Ground truth dataset and baseline evaluations for intrinsic image algorithms, in Proceedings of the International Conference on Computer Vision (ICCV), 2009.

[71] A. Pentland, T. Starner, N. Etcoff, N. Masoiu, O. Oliyide, and M. Turk, Experiments with Eigenfaces, Proc. Looking at People Workshop Int'l Joint Conf. Artifical Intelligence, Aug. 1993.

[72] K.C. Lee and J. Ho and D. Kriegman, Acquiring Linear Subspaces for Face Recognition under Variable Lighting , IEEE Trans. Pattern Anal. Mach. Intelligence, 2005, vol. 27, no. 5, pages 684-698.

[73] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose, IEEE Trans. Pattern Anal. Mach. Intelligence, 2001, vol. 23, no. 6, pages 643-660.

[74] R. Gonzalez and R. Woods Digital Image Processing, Addison-Wesley Publishing Company, , Chap. 3, 3rd, 2007.

[75] M. Turk, A. Pentland, 1991. Eigenfaces for recognition. J. Cognitive Neurosci. 3, 71–86.

[76] H. Moon, P.J. Phillips, Computational and Performance aspects of PCA-based Face Recognition Algorithms, Perception, Vol. 30, pp. 303-321, 2001.

[77] S. Barsky, M., Petrou. Surface texture using photometric stereo data: Classification and direction of illumination detection, Journal of Mathematical Imaging and Vision 29 (2-3) 185–204, 2007.

[78] T. Chen, W. Yin, X. Zhou, D. Comaniciu, and T. Huang, Total variation models for variable lighting face recognition, IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 9, pp. 1519–1524, Sep. 2006.

[79] X. Tan and B. Triggs. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. IEEE Trans. on Image Processing, pp. 1635-1650, 2010.

[80] R. Jenkins, and A.M. Burton, 100% accuracy in automatic face recognition. Science. v319. 435, 2008.

[81] Y. Xu, Q. Zhu, Z. Fan, D. Zhang, J. Mi, Z. Lai, Using the idea of the sparse representation to perform coarse-to-fine face recognition, Information Sciences 238 (2013) 138-148.

[82] G. Yang, Y. Lin, P. Bhattacharya, A driver fatigue recognition model based on information fusion and dynamic Bayesian network, Information Sciences 180 (10) (2010) 1942-1954.

[83] W. Chen, J. Lee, W. Han, C. Shih, K. Chang, Iris recognition based on bidimensional empirical mode decomposition and fractal dimension, Information Sciences 221 (2013) 439-451.

[84] C. Feher, Y. Elovici, R. Moskovitch, L. Rokach, A. Schclar, User identity verification via mouse dynamics, Information Sciences 201 (2012) 19-36.

[85] B. Zhang, X. Wang, F. Karray, Z. Yang, D. Zhang, Computerized facial diagnosis using both color and texture features, Information Sciences 221 (2013) 49-59.

[86] M. Song, C. Chen, J. Bu, T. Sha, Image-based facial sketch-to-photo synthesis via online coupled dictionary learning, Information Sciences 193 (2012) 233-246.

[87] Y. Ko, E. Lee, K. Park, A robust gaze detection method by compensating for facial movements based on corneal specularities, Pattern Recognition Letters 29 (10) (2008) 1474-1485.

[88] S. Zhao, R. Grigat, Robust Eye Detection Under Active Infrared Illumination, in: International conference on Pattern Recognition, 2006, pp. 481-484.

[89] P.W. Hallinan, Recognizing human eyes, in: SPIE Proceedings-Geometric Methods In Computer Vision, 1991, pp. 214–226.

[90] A.L. Yuille, D.S. Cohen, P.W. Hallinan, Feature extracting from faces using deformable templates, International Journal of Computer Vision 8 (1992) 99–112.

[91] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. on Pattern Analysis and Machine Intelligence 23 (2001) 681–685.

[92] S. Milborrow, F. Nicolls, Locating facial features with an extended active shape model, in: ECCV, 2008, pp. 504–513.

[93] K.M. Lam, H. Yan, Locating and extracting the eye in human face images, Pattern Recognition 29 (5) (1996) 771-779.

[94] O. Jesorsky, K.J. Kirchbergand, R. Frischholz, Robust face detection using the Hausdorff distance, in: Audio and Video Biom. Pers. Auth., 1992, pp. 90–95.

[95] Z. Zhou, X. Geng, Projection functions for eye detection, Pattern Recognition 37 (5) (2004) 1049-1056.

[96]   K.W. Wong, K.M. Lam, W.C. Siu, A novel approach for human face detection from color images under complex background, Pattern Recognition 34 (10) (2001) 1993-2004.

[97]   S. Asteriadis, N. Nikolaidis, A. Hajdu, I. Pitas, An eye detection algorithm using pixel to edge information, in: Int. Symp. on Control, Commun. and Sign. Proc., Marrakech, Morocco, vol. 1, 2006.

[98]   L. Bai, L. Shen, Y. Wang, A novel eye location algorithm based on radial symmetry transform, in: ICPR, 2006, pp. 511–514.

[99]   P. Campadelli, R. Lanzarotti, G. Lipori, Precise eye localization through a general-to-specific model definition, in: BMVC, 2006, pp. 187-196.

[100]  J. M. Guo, C. C. Lin, M. F. Wu, C. H. Chang, H. Lee, Complexity Reduced Face Detection Using Probability-Based Face Mask Prefiltering and Pixel-Based Hierarchical-Feature Adaboosting, IEEE Signal Processing Letters 18 (8) (2011) 447-450.

[101]  M. Hamouz, J. Kittlerand, J. K, Kamarainen, P. Paalanen, H. Kalviainen, J. Matas, Feature-based affine-invariant localization of faces, IEEE Trans. on Pattern Analysis and Machine Intelligence 27 (9)  (2005) 1490–1495.

[102]  R. Valenti, Th. Gevers, Accurate Eye Center Location and Tracking Using Isophote Curvature, in: IEEE CVPR, Alaska, USA, 2008, pp. 1-8.

[103]  V.H. Nguyen, T.H.B. Nguyen, H. Kim, Reliable detection of eye features and eyes in color facial images using ternary eye-verifier, Pattern Recognition 45(11) (2012) 3853-3865.

[104]   V. Perlibakas, Automatical detection of face features and exact face contour, Pattern Recognition Letters 24 (16) (2003) 2977-2985.

[105]  F.Y. Shih, C.F. Chuang, Automatic extraction of head and face boundaries and facial features, An International Journal Information Sciences Informatics and Computer Science 158 (1) (2004) 117-130.

[106] D. Cristinacce, T. Cootes, I. Scott, A multi-stage approach to facial feature detection, in: British Machine Vision Conference, 2004, pp. 231-240.

[107] C. Travieso, J. Zhang, et. al, Bimodal biometric verification based on face and lips, Neurocomputing 74 (14-15) (2011) 2407-2410.

[108] S. Asteriadis, N. Nikolaidis, I. Pitas, Facial Feature Detection using Distance Vector Fields, Pattern Recognition 42 (7) (2009) 1388-1398.

[109] M. W. Jian, K.M. Lam, Fast Eye Detection and Localization Using a Salient Map, in: Pacific-Rim Conference on Multimedia (PCM), Sydney, 2011, pp. 74-83.

[110] P. Viola, M.J. Jones, Robust real-time face detection. International Journal of Computer Vision 57 (2) (2004) 137–154.

[111] I. Daubechies, Orthonormal Bases of Compactly Supported Wavelets, Communications on Pure and Applied Mathematics, 1988, pp. 909-996.

[112] R.L. Devalois, D.G. Albrecht, L.G. Thorell, Spatial-frequency selectivity of cells in macaque visual cortex, Vision Research, 22 (1982) 545-559.

[113] N. Sebe, M.S. Lew, Salient points for content-based retrieval, in: British Machine Vision Conference, 2001, pp. 401–410.

[114] N. Sebe, M.S. Lew, Comparing salient point detectors, Pattern Recognition Letters 24 (1–3) (2003) 89–96.

[115] N. Sebe, Q. Tian, E. Loupias, M.S. Lew, T.S. Huang, Color indexing using wavelet-based salient points, in: IEEE Workshop on Content-based Access of Image and Video Libraries, 2000, pp. 15–19.

[116] N. Sebe, Q. Tian, E. Loupias, M.S. Lew, T.S. Huang, Evaluation of salient point techniques, in: International Conference on Image and Video Retrieval, 2002, pp. 367–377.

[117] N. Sebe, Q. Tian, E. Loupias, M.S. Lew, T.S. Huang, Evaluation of salient point techniques, Image and Vision Computing 21 (13-14) (2003) 1087-1095.

[118] Q. Tian, N. Sebe, E. Loupias, M.S. Lew, T.S. Huang, Image retrieval using wavelet-based salient points. Journal of Electronic Imaging 10 (4) (2001) 835–849.

[119] M. Jian, Y. Dong, Wavelet-Based Salient Regions and their Spatial Distribution for Image Retrieval, in: Int. Conf. on Multimedia & Expo (ICME), 2007. pp. 2194-2197.

[120] M.W. Jian, Y. J. Dong, J. Ma, Image retrieval using wavelet-based salient regions, The Imaging Science Journal 59 (4) (2011) 219-231.

[121] S. Solomon, Dictionary of Eye Terminology. Gainesville, Florida: Triad Publishing Company, 1990.

[122] C.Y. Wu, C. Liu, H. Y. Shum, Y. Q. Xu, Z. Y. Zhang, Automatic eyeglasses removal from face images, IEEE Trans. on Pattern Analysis and Machine Intelligence 26 (3) (2004) pp. 322-336.

[123] P. H. Hennings-Yeomans, S. Baker, and B. V. K. V. Kumar, Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, in IEEE Conf. Computer Vision and Pattern Recognition, 2008, pp. 1-8.

[124] M. W. Jian, K. M. Lam, J. Y. Dong,, "A Novel Face-Hallucination Scheme Based on Singular Value Decomposition," Pattern Recognition 46 (11): 3091-3102 (2013).

[125] W. W. Zou, and P. C. Yuen. Very Low Resolution Face Recognition Problem, IEEE Trans. Image Process., Vol. 21, N0. 1, Jan. 2012 327-340.

[126] C. X. Ren, D. Q. Dai, H. Yan. Coupled Kernel Embedding for Low-Resolution Face Image Recognition, IEEE Trans. Image Process., Vol. 21, No. 8, Aug. 2012.

[127] M. W. Jian and K. M. Lam, "Face-Image Retrieval Based on Singular Values and Potential-Field Representation." Signal Processing. Vol. 100, July 2014, Pages 9-15.

[128] M. W. Jian, K. M. Lam, J. Y. Dong, Facial-Feature Detection and Localization Based on a Hierarchical Scheme, Information Sciences, Vol. 262, pp. 1-14, 2014

[129] J. Kim, J. Choi, J. Yi, et al. Effective representation using ICA for face recognition robust to local distortion and partial occlusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 1977-1981, 2005.

[130] The Bosphorus Database: http://bosphorus.ee.boun.edu.tr/